

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/147752>

**Copyright and reuse:**

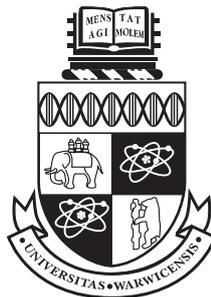
This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)



---

# A mathematical approach to identifying antibody affinity with modern biosensors, for pharmaceutical and clinical settings

by

**Harold Arthur James Moyse**

---

**Thesis**

Submitted to the University of Warwick  
for the degree of  
**Doctor of Philosophy**

---

*Supervisors:* Neil Evans, Daniel Zehnder, Daniel Mitchel

MOAC Doctoral Training Centre

July 2018



THE UNIVERSITY OF  
**WARWICK**



---

# A mathematical approach to identifying antibody affinity with modern biosensors, for pharmaceutical and clinical settings

by

Harold Arthur James Moyse

---

## Thesis

Submitted to the University of Warwick  
for the degree of  
**Doctor of Philosophy**

---

*Supervisors:* Neil Evans, Daniel Zehnder, Daniel Mitchel

MOAC Doctoral Training Centre

July 2018



THE UNIVERSITY OF  
**WARWICK**

# Contents

<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>xiii</b>
<b>Abbreviations</b>	<b>xv</b>
<b>Acknowledgements</b>	<b>xvi</b>
<b>Abstract</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Transplantation . . . . .	1
1.2 Antibody, antigen and the immune system . . . . .	5
1.3 Affinity and Avidity . . . . .	14
1.4 Surface plasmon resonance technology . . . . .	17
1.4.1 SPR studies of monoclonal antibody affinity . . . . .	21
1.4.2 SPR studies of patient antibody affinity . . . . .	25
1.5 Model development and analysis . . . . .	28
1.6 Bio layer interferometry technology . . . . .	32
1.7 Aims and Objectives . . . . .	34
1.8 Thesis Overview . . . . .	34
<b>2 Theory - Basic Concepts</b>	<b>38</b>
2.1 Concepts from statistics . . . . .	38
2.2 State-space models . . . . .	40
2.3 Dynamical systems with inputs, and concepts from calculus . . . . .	57
2.4 Gröbner bases, concepts from abstract algebra and the Wronskian . . . . .	62
2.5 Lie derivatives, fibres and continuity of solutions . . . . .	67
<b>3 Theory - Identifiability and Indistinguishability</b>	<b>69</b>
3.1 Structural Identifiability . . . . .	69
3.1.1 The Taylor Series Approach . . . . .	72
3.1.2 Observability Rank Condition . . . . .	84
3.1.3 The Output Equation Method . . . . .	91
3.2 Novel tools for analysing Identifiability . . . . .	99
3.2.1 Simple identifiability analysis . . . . .	105
3.2.2 Identifiability for functions of parameters and parallel experiments . . . . .	107
3.2.3 Output limits at the discontinuity . . . . .	112
3.2.4 Combining methods . . . . .	118

3.3	Structural Indistinguishability . . . . .	119
3.3.1	Limits of models . . . . .	122
3.4	Algorithms and software for model fitting . . . . .	123
3.4.1	Differential Evolution . . . . .	124
3.4.2	FACSIMILE . . . . .	125
3.5	Conclusion . . . . .	126
<b>4</b>	<b>Experimental Methods</b>	<b>129</b>
4.1	Goals . . . . .	129
4.2	Selecting experimental methods . . . . .	130
4.3	Original Monoclonal Experiments . . . . .	133
4.3.1	Reacting SN607D8 antibody with HLA-A2 antigen . . . . .	138
4.3.2	Reacting SN230G6 antibody with HLA-A2 antigen . . . . .	140
4.3.3	Reacting WK1D12 antibody with HLA-B7 and HLA-B27 antigens . . . . .	142
4.4	Supplementary Monoclonal Experiments . . . . .	144
4.5	Purified IgG/IgA protein samples . . . . .	148
4.6	Plasma effluent and serum samples . . . . .	149
4.7	High affinity plasma effluent sensorgrams . . . . .	152
4.8	Mixed monoclonal experiments . . . . .	154
4.9	Sensitised patient study . . . . .	158
4.10	Summary . . . . .	165
<b>5</b>	<b>Models for monoclonal antibody</b>	<b>167</b>
5.1	Models . . . . .	167
5.1.1	Langmuir model . . . . .	169
5.1.2	The Langmuir with transport model . . . . .	171
5.1.3	Effective rate constant approximation model . . . . .	178
5.1.4	Bivalent Langmuir model . . . . .	184
5.1.5	Bivalent ERC model . . . . .	190
5.1.6	Bivalent ERC model with spatial effects . . . . .	194
5.2	Equilibrium analysis and comparisons between the behaviour of models as parameters vary . . . . .	203
5.2.1	Extending models for parallel experiments with multiple antibody concentrations and antigen densities . . . . .	205
5.3	Identifiability and Indistinguishability . . . . .	208
5.3.1	Langmuir model . . . . .	209
5.3.2	Langmuir with transport model . . . . .	210
5.3.3	Effective rate constant model . . . . .	214
5.3.4	Bivalent Langmuir model . . . . .	216
5.3.5	Bivalent ERC model . . . . .	217
5.3.6	Bivalent ERC model with spatial effects . . . . .	217
5.4	Discussion . . . . .	218
<b>6</b>	<b>Affinity estimation for monoclonal antibody</b>	<b>221</b>

---

6.1	Error in SPR experiments . . . . .	223
6.2	Fitting methods . . . . .	225
6.3	Preparing the data for fitting . . . . .	225
6.4	Developing and comparing models . . . . .	231
6.4.1	Testing the ERC2 model against the L2 model . . . . .	231
6.4.2	Testing the BERC model against the BL model . . . . .	242
6.4.3	The bivalent ERC model with spatial effects . . . . .	250
6.4.4	Variations of the BERCs model for the SN230G6 vs A2 dataset . . . . .	258
6.4.5	Initial fitting the simplified BERCs model . . . . .	265
6.5	Processing expanded datasets with the simplified BERCs model . . . . .	269
6.6	Processing new data with the BERCs model . . . . .	275
6.6.1	Correlation of parameters . . . . .	281
6.7	Discussion . . . . .	283
<b>7</b>	<b>Models for polyclonal antibody</b>	<b>291</b>
7.1	Previously Published Models . . . . .	293
7.1.1	Langmuir model . . . . .	293
7.1.2	Effective rate constant approximation model . . . . .	299
7.2	General model of polyclonal antibody binding . . . . .	301
7.2.1	Paratope-epitope matrices and related functions . . . . .	308
7.2.2	Rates of change . . . . .	316
7.2.3	Two antibodies competing to bind for a single epitope . . . . .	327
7.2.4	Two antibodies binding to two separate epitopes . . . . .	332
7.2.5	Adapting the GPM for interactions between antibodies of different size . . . . .	334
7.3	Identifiability and indistinguishability . . . . .	337
7.3.1	Langmuir model . . . . .	337
7.3.2	Effective rate constant approximation model . . . . .	339
7.3.3	General model of polyclonal antibody binding . . . . .	340
7.4	Model selection and discussion . . . . .	341
<b>8</b>	<b>Fitting Models to polyclonal antibody experiments</b>	<b>344</b>
8.1	Fitting models to sensorgrams from purified samples of IgG, IgA and the remaining proteins . . . . .	346
8.1.1	Comparison . . . . .	350
8.2	Plasma effluent and serum samples . . . . .	352
8.3	Mixtures of monoclonal antibody . . . . .	354
8.3.1	Using the GPM2X2 to estimate concentration of antibody in a mixture . . . . .	360
8.3.2	Fitting the GPM to sensorgrams with mixtures of monoclonal antibody reacted against a single antigen . . . . .	363
8.4	Sensitized patient study . . . . .	369
8.5	Modelling sensitized patient sensorgrams with variations of the GPM model . . . . .	371
8.5.1	Correlation of parameters . . . . .	385
8.6	Discussion . . . . .	389

<b>9</b>	<b>Estimating the affinity of antibody with an unknown concentration</b>	<b>393</b>
9.1	ELISA . . . . .	394
9.2	Comparing BLI and SPR models . . . . .	398
9.2.1	Langmuir Model . . . . .	400
9.3	Estimating the affinity of single unknown monoclonal antibody . . . . .	406
9.4	Estimating the affinity of multiple monoclonal antibodies or polyclonal serum	409
<b>10</b>	<b>Conclusions</b>	<b>411</b>
10.1	Further Work . . . . .	414
10.1.1	Clinical Research . . . . .	414
10.1.2	SPR modelling . . . . .	415
10.1.3	Identifiability and Indistinguishability . . . . .	417
10.1.4	Therapeutic applications . . . . .	420
	<b>Bibliography</b>	<b>427</b>
	<b>Appendix A Samples of Maple and Facsimile code</b>	<b>440</b>
A.1	Maple code from Chapter 2 . . . . .	440
A.1.1	Deriving the state-space solution for the Langmuir 2 concentration model . . . . .	440
A.2	Chapter 3 . . . . .	441
A.3	Structural Identifiability . . . . .	442
A.3.1	Example 2.2 . . . . .	442
A.3.2	Measuring the time taken to find, and dimension of a solution by the TSA . . . . .	443
A.3.3	Output equation method . . . . .	447
A.3.4	Example 3.1.3 . . . . .	448
A.3.5	Output limits at discontinuity . . . . .	449
A.3.6	Measuring the time taken to find, and dimension of a solution by the BVDP and TSA at limits . . . . .	450
A.4	Derivations . . . . .	455
A.5	Identifiability analyses . . . . .	457
A.5.1	LwT Identifiability . . . . .	457
A.5.2	BERCS Identifiability . . . . .	460
A.6	Sample facsimile program . . . . .	466
A.7	Sample facsimile Logfile . . . . .	474
	<b>Appendix B Publications</b>	<b>509</b>
B.1	Paper in Human Immunology . . . . .	509
B.2	Paper at 8th IFAC Symposium on Biological and Medical Systems . . . . .	517

# List of Figures

1.1	Post transplant graft survival . . . . .	2
1.2	Diagrams of Antigen . . . . .	7
1.3	Numbers of discovered alleles by year . . . . .	9
1.4	Diagram of HLA-A2 . . . . .	10
1.5	Inteactions of phagocytes, lymphocytes, antibody, bacteria and viruses . . .	11
1.6	A general representation of the method used to produce monoclonal antibodies	12
1.7	Diagram of antibody structures . . . . .	13
1.8	Diagram of polyclonal antibody-antigen structures . . . . .	16
1.9	Distrobution of measured affinities for rabbit and mouse monoclonal anti- bodies against their target antigen . . . . .	17
1.10	Diagram of experiment design for the XPR36 biosensor . . . . .	18
1.11	Diagrams of TIR and SPR taken from (Marquart, 2013). . . . .	19
1.12	Diagram of refraction in an SPR biosensor . . . . .	21
1.13	Illustrations of two configurations of SPR experiment for investigating antibody- antigen interactions. . . . .	22
1.14	Distribution of the kinetic rates and the equilibrium binding constants, with average values marked using a dotted line and the boundaries of standard deviation highlighted in grey, adapted from Katsamba <i>et al.</i> (2006). . . . .	23
1.15	Plots of estimateds of $k_a$ and $k_d$ along with lines of isoaffinity taken from two multinational benchmark studies. . . . .	24
1.16	Plots of estimateds of $k_a$ and $k_d$ along with lines of isoaffinity for 10 inter- actions, taken from Papalia <i>et al.</i> (2006). . . . .	26
1.17	ERC model output (black lines) plotted with corresponding experimental data (grey lines) taken from Evans <i>et al.</i> (2010). . . . .	27
1.18	Illustrations of three of the common assumptions about unbound analyte concentration used in models. . . . .	30
1.19	Illustrations of light reflected in BLI experiments taken from (ForteBio, 2007).	33
1.20	Visualisation of the relationships between thesis chapters. . . . .	35
2.1	Diagrams of relationships between parameter-state, state-output and parameter- output mappings as well as their domains and images. . . . .	46
2.2	The input, output and state-space solutions of the Langmuir model for a given input $I$ and parameter vector $\mathbf{p}$ , as well as the mappings between them.	52
2.3	Compartmental diagrams of the Langmuir model . . . . .	54
2.4	The output and state-space solution curves of the Langmuir model for a given input $I$ and parameter vector $\mathbf{p}$ . . . . .	56
2.5	Phase portraits and visualisation of the two concentration Langmuir model	60

3.1	Graph showing $(C_T, k_a)$ pairs indistinguishable to those shown in Table 2.2.	77
3.2	In the L1 model indistinguishable parameters are indistinguishable because for a generic input $I$ , $X_{\Pi}(\cdot, I)$ is surjective, even though $Y_{\Sigma}(\cdot)$ is bijective.	78
3.3	Graph showing the set of parameters indistinguishable to those shown in Table 2.3 as two lines, one a reflection of the other in the plane $k_{a1} = k_{a2}, k_{d1} = k_{d2}$ , one is dashed to represent that two of the dimensions it has moved in are not shown.	80
3.4	In the L2 model indistinguishable parameters are indistinguishable because both $X_{\Pi}(\cdot)$ and $Y_{\Sigma}(\cdot)$ are surjective, for a generic input $I$ .	81
3.5	Graphs showing how the dimension of solution and the time taken to obtain the solution as numbers of derivatives of the output at the start of either phase vary.	82
3.6	Graph showing state-space trajectories of the two configuration Langmuir model with two parameter vectors with equal outputs (blue and green) with a dotted red line for $x_1 = x_2$ and a solid red line for $R = x_1 + x_2$ . The parameters used in this graph are taken from Table 2.3 and an indistinguishable parameter vector derived with Eq. 3.38.	92
3.7	The relationship between simple identifiability and identifiability for $\mathbf{m}(\mathbf{p})$ , a rational function of parameters.	102
3.8	The relationship between simple structural identifiability and structural identifiability for a model $\Sigma$ .	104
3.9	Graphs showing how the dimension of solution and the time taken to obtain the solution as numbers of derivatives of the output at the start of either phase vary.	117
3.10	Diagram of fibres of $X_{\Pi}(\mathbf{p}, \cdot)$ and $Y_{\Sigma}(\mathbf{x}(\cdot; \mathbf{p}, I))$	119
4.1	Diagram of interaction spots on the XPR 36 chips for the SN607D8 antibody vs HLA-A2 antigen experiment	136
4.2	Sensorgrams from SN607D8 antibody vs HLA-A2 antigen experiments	137
4.3	Diagram of interaction spots on the XPR 36 chips for the SN230G6 antibody vs HLA-A2 antigen experiment	140
4.4	Sensorgrams from SN230 vs HLA - A2 experiments	141
4.5	Diagram of interaction spots on the XPR 36 chips for the WK1D12 antibody vs HLA-B27 and HLA-B7 antigens experiment	143
4.6	sensorgrams from experiments reacting WK1D12 against HLA - B27 and HLA-B7 antigens	143
4.7	Diagram of interaction spots on XPR 36 chips	145
4.8	SN607 vs HLA-A2, HLA-A68 and HLA-A69	146
4.9	SN230G6 vs B57 and A2	147
4.10	Sensorgrams from purified patient protein vs A trisaccharide.	148
4.11	Diagram of interaction spots on the XPR 36 chips for the patient samples against HLA antigen experiments	150
4.12	Sensograms from processed plasma effluent reacted against HLA antigens	151

---

4.13	Sensorgrams from processed serum samples reacted against HLA antigens . . .	151
4.14	Sensorgrams of patient antibody being reacted against HLA antigen with fits	153
4.15	Mixed SN607D8 and SN230G6 vs HLA-A2 experiment design. . . . .	154
4.16	Sensorgrams from mixtures of monoclonal antibody reacted against HLA-A2 antigen . . . . .	156
4.17	Sensorgrams from mixtures of monoclonal antibody reacted against HLA-A2 antigen . . . . .	157
4.18	Maximum response for each lane in each sensorgram from the sensitised patient study. . . . .	159
4.19	Serum from LT35 reacted with the multiple antigen . . . . .	159
4.20	Serum from LT40 reacted with the multiple antigen . . . . .	160
4.21	Serum from LT65 reacted against A2 antigen . . . . .	160
4.22	Serum from LT39 reacted against B37 antigen . . . . .	160
4.23	Serum from LT43 reacted with multiple antigen . . . . .	161
4.24	Serum from LT68 reacted with multiple antigen . . . . .	161
4.25	Serum from LT42 reacted with multiple antigen . . . . .	162
4.26	Serum from LT78 reacted against A2 antigen in various densities . . . . .	162
4.27	Serum from LT2 reacted with multiple antigen . . . . .	163
4.28	Serum from LT33 reacted against B4002 antigen . . . . .	163
4.29	Serum from LT39 reacted against A2 antigen . . . . .	163
4.30	Serum from LT36 reacted against multiple antigen . . . . .	164
4.31	Serum from LT21 reacted against multiple antigen . . . . .	164
4.32	Serum from LT16 reacted against A1 antigen . . . . .	164
4.33	Serum from LT79 reacted against A2 antigen . . . . .	164
5.1	Diagrams of an IgM antibody binding to multiple antigen . . . . .	168
5.2	Phase portraits of the LwT model . . . . .	173
5.3	Compartmental diagram of the two configuration Langmuir with transport model . . . . .	173
5.4	Visualisations of a solution trajectory for the Langmuir two concentration model in the association phase . . . . .	176
5.5	Visualisations of a solution trajectory for the Langmuir two concentration model in the dissociation phase . . . . .	177
5.6	Compartmental diagram of the effective rate constant model . . . . .	178
	(c) Association phase . . . . .	180
	(d) Dissociation phase . . . . .	180
5.7	Graphs of bound antibody $x_1$ against time $t$ for the Langmuir single configuration model (green dashed line) and the ERC model for different values of $k_m$ . . . . .	180
	(a) A visualisation of the state and output variables of the two concentration ERC model. $x_1$ is shown as a dotted line, $x_2$ as a dashed line and $y/\alpha$ as a yellow line. . . . .	182

---

(b)	Phase portrait for the association phase, with nullclines for $x_1$ shown as a dotted line, for $x_2$ as a dashed line. . . . .	182
(c)	Phase portrait for the dissociation phase, with nullclines for $x_1$ shown as a dotted line, for $x_2$ as a dashed line. . . . .	182
5.8	Phase portraits and visualisation of the two concentration ERC model . . .	182
5.9	Compartmental diagram of the bivalent Langmuir model . . . . .	186
5.10	Phase portraits of the two concentration bivalent model . . . . .	187
5.11	A logarithmically scaled graph of the effective dissociation rate constant as $k_{a2}$ varies in proportion to $k_{a1}$ . . . . .	190
5.12	Compartmental diagram of the bivalent ERC model . . . . .	192
5.13	Phase portraits and visualisation of the BERC model . . . . .	193
5.14	Compartmental diagram of the BERCs model . . . . .	196
5.15	Visualisations of a solution trajectory for the Langmuir two concentration model in the association phase (black) alongside nullcline-planes $x_1$ in blue, $x_2$ in green, $x_m$ in red. . . . .	197
5.16	Visualisations of a solution trajectory for the Langmuir two concentration model in the dissociation phase (black) alongside nullcline-planes $x_1$ in blue, $x_2$ in green, $x_m$ in red. . . . .	198
5.17	A visualisation of the state and output variables of the bivalent BERCs model. $x_1$ is shown as a dash-dotted line, $x_2$ as a dotted line, $x_m$ as a dashed line and $y/\alpha$ as a yellow line. . . . .	199
5.18	Output at steady state in association phase against analyte concentration, green for Langmuir, Langmuir with transport and ERC models, and blue for the bivalent and bivalent ERC model. . . . .	204
5.19	Initial rate of change of output variable in the association phase against analyte concentration, blue for the Langmuir and bivalent Langmuir model, green for the ERC and bivalent ERC model. . . . .	207
5.20	Initial rate of change of output variable in the dissociation phase where state variables had been in a steady state against analyte concentration, red for the Langmuir, blue for the bivalent Langmuir, green for the ERC and magenta bivalent ERC. . . . .	208
5.21	Graph showing three indistinguishable state-space trajectories (red, blue and green), corresponding to parameter vectors indistinguishable to those in Table 5.3. . . . .	212
5.22	In the L2 model indistinguishable parameters are indistinguishable because both $X_{\Pi}(\cdot, I)$ and $Y_{\Sigma}(\cdot)$ are surjective, for a generic input $I$ . . . . .	212
5.23	Visualisations of a solution trajectories for the Langmuir two concentration model with identical outputs. . . . .	214
6.1	Histogram of SPR noise . . . . .	224
6.2	Two configuration bivalent fits to data with varying step functions. . . . .	227
6.3	Diagrams of data sets . . . . .	233
6.4	Scatter graph of RSS for the Langmuir and ERC model . . . . .	235

---

6.5	Comparisons between fits of the ERC model and the Langmuir model . . .	237
6.6	Estimates of association and dissociation rate constants . . . . .	241
6.7	Estimates of $k_m$ and $R$ for the L2 and ERC2 models. . . . .	243
6.8	Graphs comparing RSS. See Figure 6.4 for legend. . . . .	244
6.9	Comparisons of fit of model and residuals of model for each antibody . . . .	246
6.10	Scatter graphs of estimates of $k_{a1}$ and $k_{d1}$ , with boxplots and secondary axis representing affinity. See Figure 6.4 for legend. . . . .	249
6.11	Estimates of $k_m$ and $R$ for the BL and BERC models. . . . .	251
6.12	Scatter graph of RSS for the BERC and BERCs model. See Figure 6.4 for legend. . . . .	253
6.13	Comparisons of fit between the BERCs and BERC model . . . . .	255
6.14	Scatter graphs of estimates of association and dissociation rate constants, with box-plots and secondary axes representing affinity. See Figure 6.4 for legend. . . . .	258
6.15	Estimates of $k_m$ and $R$ for the BERCs model. . . . .	259
6.16	Fits of the Bivalent ERC model with spatial effects with systematic errors, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.3. . . . .	260
6.17	Scatter graphs of estimates of association and dissociation rate constants, with confidence intervals, boxplots and secondary axis representing affinity. . . . .	263
6.18	RSS values for the BERCs and BERCs extended models. . . . .	264
6.19	Illustrations of the fits of the extended BERCs model, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.3. . . .	266
6.20	Graphs showing the RSS and affinity estimates made by the simplified BERCs model. See Figure 6.4 for legend. . . . .	267
6.21	Diagrams of data sets . . . . .	270
6.22	Fits of the BERCs simplified model with residuals, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.21. . . . .	271
6.23	Graphs of RSS and rate constants for the BERCs simplified model. . . . .	272
6.24	Estimates of $k_m$ and $R$ for the BL and BERC models. . . . .	273
6.25	Diagram of interaction spots on XPR 36 chips. . . . .	276
6.26	Scatter Graph of RSS for both models. . . . .	277
6.27	Fits of the BERCs simplified model with residuals, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.25. . . . .	278
6.28	Estimates of primary and secondary association and dissociation rate constants made with the BERCs model. See Figure 6.26 for legend. . . . .	280
6.29	Estimates of $k_m$ and $R$ for the BERCs simplified model. . . . .	281
6.30	Graphs of parameter covariances for the parameter estimates shown in Table 6.6.1. . . . .	282
6.31	Diagram of model development. . . . .	284
6.32	Comparison of AICs. . . . .	285
6.33	Sensorgrams for interactions of 5 antibodies with antigen, taken with various SPR platforms. Adapted from Yang <i>et al.</i> (2016). . . . .	287

---

7.1	Phase portraits of the two ligand Langmuir model . . . . .	295
7.2	Compartmental diagram of the Langmuir model . . . . .	298
7.3	Diagrams of the regions in which the free paratope of a bound antigen can move. . . . .	304
7.4	Rates at which complexes grow by the addition of monomers. . . . .	307
7.5	Rates at which complexes grow by binding with other complexes. . . . .	308
7.6	Rates at which complexes shrink by the dissociation of monomers. . . . .	309
7.7	The complex represented by the paratope-epitope matrices of Eq. (7.10). . . . .	310
7.8	Compartments and flows connected to $x_M$ , where $M$ has one or more entries taking the value of two; $x_M$ represents the concentration of a specific complex containing two antigen and one or more antibody. . . . .	315
7.9	Simplified compartmental diagrams of generalised ployclonal model of antibody binding . . . . .	317
7.10	Illustrations of antibody- antigen complexes. . . . .	318
7.11	Illustrations to explain the set $T_{M,j}$ . . . . .	320
7.12	Graphs of the state and output variables for the GPE 2 antibody competitive model with parameters as shown in Table 7.2, where SN230G6 is antibody 1 and SN607D8 is antibody 2 and their epitope-paratope matrices for dimers are $J$ and $K$ respectively. . . . .	330
7.13	Graphs of the state and output variables for the GPE 2 antibody non-competitive model, where SN230G6 is antibody $j$ and SN607D8 is antibody $k$ . . . . .	333
7.14	Comparison between the tetramers with IgG and IgM bivalently bound. . . . .	335
7.15	Comparison between the IgG and IgM bound on isolated antigen. . . . .	337
7.16	Illustrations of the kinds of complex that can be made when various kinds of antigen and antibody molecules react . . . . .	342
8.1	Fits of the single concentration ERC model to separated antibodies . . . . .	347
8.2	Fits of the two concentration ERC model to separated antibodies . . . . .	349
8.3	Scatter graph of the RSS for the Fits of the the ERC2C and L2C models to the sensorgrams from Section 4.6 . . . . .	352
8.4	Alternate visualisations of the data from a single sensorgram Fig. 4.16(c). . . . .	356
8.5	Average effective $k_d$ for each mixed antibody sensorgram (Fig. 4.16 and 4.17) for $m = 200$ against time, and related graphs. . . . .	357
8.6	Average effective $k_d$ at the last 91s of the dissociation phase in the mixed monoclonal experiments . . . . .	359
8.7	Boxplot of the error of these estimates. . . . .	361
8.8	GPM2X2 fits to sensorgram data of mixed antibody from Fig. 4.16 and 4.17. . . . .	362
8.9	Total antibody concentration of 100nM . . . . .	364
8.10	Diagram of the ways the outputs from interaction spots may be grouped for modelling . . . . .	364

---

8.11	Fits of the GPM to sensorgrams with mixtures of SN607D8 and SN230G6 in the mixtures 100:0 blue, 75:25 green, 50:50 red, 25:75 cyan, and 0:100 purple . . . . .	366
8.12	Plots of the distribution of estimates of $R_b$ and $R_m$ and their logs in the mixed monoclonal fits . . . . .	367
8.13	Diagram of the antibody footprint of SN230G6 (yellow) and SN607D8 (red) and their overlap (purple) on HLA-A2 taken from Jones <i>et al.</i> (2016, personal communication) . . . . .	369
8.14	Boxplots of average effective $k_d$ at the last 400s of the dissociation phase in the polyclonal experiments . . . . .	370
8.15	AICs for the fits to the data with the four models, 2 antibody of separate size non-competitive (dark blue), 2 antibody same size non-competitive (cyan), 2 antibody competitive (yellow), BERCs model (red) . . . . .	373
8.16	Plots of fits of the GPMSS 2X2 to patient sensorgrams. . . . .	374
8.17	Plots of fits of the GPM 1X1 to patient sensorgrams. . . . .	375
8.18	Bar charts of estimates of the dissociation rate constant for each model, for each of the patient-antigen pairings . . . . .	383
8.19	Bar charts of estimates of the association rate constant concentration product for each model, for each of the patient-antigen pairings . . . . .	384
8.20	Box plots of parameter estimates for the patient data set with the polyclonal models . . . . .	385
8.21	Graphs of parameter correlations for the GPM1X1 fitted to the sensitised patient sensorgrams. . . . .	386
8.22	Graphs of parameter correlations for the GPM2X1 fitted to the sensitised patient sensorgrams. . . . .	387
8.23	Graphs of parameter correlations for the GPM2X2 fitted to the sensitised patient sensorgrams. . . . .	388
8.24	Graphs of parameter correlations for the GPMSS2X2 fitted to the sensitised patient sensorgrams. . . . .	389
9.1	Diagram of the relationship between three common techniques for affinity analysis. . . . .	394
9.2	Illustrations of ELISA methods taken from Pokhrel (2015) . . . . .	396
9.3	Diagram of BLI reaction well and derivation of equations for concentration in either phase Eq. (9.3), in terms of Avogadro's number $N_a$ . . . . .	399
9.4	Graph of bound analyte $x_1(t)$ , against time, $t$ for the L1 and L1FC models, with the parameters shown in Table 9.1 . . . . .	402
9.5	Simulated experiments with analytes that would give identical results on the SPR platform . . . . .	403
9.6	An antibody in the proximity of antigen attached by linker to the chip's surface. . . . .	406
9.7	Simulations of the BERCs model and BLI version of the BERCs model with parameters averaged from Table 6.6.1 . . . . .	408

10.1	Visualisations of IgG hexamers from Diebolder <i>et al.</i> (2014). . . . .	416
10.2	Diagrams of $X_{\text{III}}(\cdot)$ and $Y_{\Sigma}(\cdot)$ as well as their domains and images for all models as defined in Definition 2.6. . . . .	418
10.3	This image shows cancerous B cells that have been treated with rituximab. The protein CD20 (shown in green) has been drawn to the side of the cells from Kurzweil (2013). . . . .	423
10.4	Schematic of proteolytic fragmentation of immunoglobulin by brief treatment with enzymes papain and pepsin from ThermoFisher (2015). . . . .	424

# List of Tables

1.1	Table of HLA- A epitopes an antigen . . . . .	8
2.1	Comparison between units used in Marquart (2013) and this thesis. . . . .	51
2.2	Values taken by parameters and steady states in Fig. 2.2 . . . . .	53
2.3	Values taken by parameters in Fig. 2.4 . . . . .	56
2.4	Values taken by parameters in Fig. 2.5 . . . . .	60
4.1	An overview of the techniques for determining antibody-antigen affinity giving an output at a single timepoint (Adapted from Neri <i>et al.</i> (1996)) . .	132
4.2	Characteristics of IgG human monoclonal HLA-specific antibodies . . . . .	134
4.3	XPR estimates of SN607D8 kinetic constants . . . . .	147
4.4	XPR estimates of SN230G6 kinetic constants. . . . .	147
4.5	XPR estimates of kinetic constants . . . . .	154
4.6	XPR estimates of kinetic constants . . . . .	165
4.7	XPR estimates of kinetic constants . . . . .	166
5.1	Values taken by parameters in the simulation shown in Fig. 5.2. . . . .	172
5.2	Values taken by parameters in Fig. 5.4 and 5.5 . . . . .	176
5.3	Values taken by parameters in Fig. 5.7 that are not varied . . . . .	180
5.4	Values taken by parameters in Fig. 5.8. . . . .	182
5.5	Values taken by parameters in Fig. 5.10. . . . .	186
5.6	Values taken by parameters in Fig. 5.11. . . . .	190
5.7	Values taken by parameters in Fig. 5.13. . . . .	192
5.8	Values taken by parameters in Fig. 5.15-5.17. . . . .	197
5.9	Values taken by parameters in Fig. 5.10. . . . .	204
5.10	Values taken by parameters in Fig. 5.19 and 5.22 . . . . .	207
5.11	Structure of mappings for models. . . . .	219
6.1	Details of parameter space as considered by Differential Evolution. . . . .	232
6.2	Estimated parameters for each dataset for the L2 model, with parameters with no SDLN or an SDLN greater than 0.2 marked with an asterisk. . . .	238
6.3	Estimated parameters for each dataset for the ERC2 model, with parameters with no SDLN or an SDLN greater than 0.2 marked with an asterisk. .	239
6.4	Details of parameter space considered by Differential Evolution. . . . .	244
6.5	Bivalent Langmuir. . . . .	247
6.6	Bivalent ERC. . . . .	248
6.7	Bivalent ERC spatial. . . . .	257
6.8	Parameter estimates. . . . .	262
6.9	Parameter estimates made with the Bivalent ERC spatial model. . . . .	268

*LIST OF TABLES*

---

6.10	Parameter estimates made with the BERCs simplified model. . . . .	274
6.11	Bivalent ERC spatial. . . . .	279
6.12	Difference in rate constants and affinities for interaction with different alleles compared to immunising allele, with p values for significant differences between alleles. . . . .	280
7.1	Values taken by parameters in Fig. 7.1 . . . . .	295
7.2	Values taken by parameters in Fig.7.12 and Fig. 7.13 . . . . .	331
8.1	Parameter estimates from the single concentration ERC model . . . . .	348
8.2	Parameter estimates from the single concentration ERC model . . . . .	350
8.3	Parameter estimates. . . . .	353
8.4	Parameter estimates with the GPN model ran on the mixed monoclonal dataset. . . . .	365
8.5	Parameter estimates with the GPN 1X1 model ran on the sensitized patient sensorgrams. . . . .	377
8.5	Parameter estimates from the GPN 1X1 model ran on the sensitized patient sensorgrams (continued). . . . .	378
8.6	Parameter estimates from the GPN 2X1 model ran on the sensitized patient sensorgrams. . . . .	378
8.6	Parameter estimates from the GPN 2X1 model ran on the sensitized patient sensorgrams (continued). . . . .	379
8.7	Parameter estimates from the GPN 2X2 model ran on the sensitized patient sensorgrams. . . . .	380
8.7	Parameter estimates from the GPN 2X2 model ran on the sensitized patient sensorgrams (continued). . . . .	381
8.8	Parameter estimates from the GPNSS 2X2 model ran on the sensitized patient sensorgrams. . . . .	381
8.8	Parameter estimates from the GPNSS 2X2 model ran on the sensitized patient sensorgrams (continued). . . . .	382
9.1	Paramters used in Fig. 9.4 . . . . .	401
9.2	Paramters used in Fig. 9.2.1 . . . . .	404

# Acknowledgements

There is an oft repeated African proverb “It takes a Whole Village to Raise a Child”. In the process of research and writing I have discovered it takes a whole village to get a PhD.

I dedicate this to my parents; my mother who raised me to reason and debate and instilled in me my faith in God, my father who my earliest memories include reading to me from science books whilst we waited for my big sister to come home from nursery; to my big sister, Lucy, who I was always trying to catch up at school - and who’s inspired me both through her successes and kindness.

To my many aunts uncles and cousins, who I have the honour of being related to - and who have the dubious honour of being related to me; whose support guided me from infancy to PhD.

To my friends, John, Hannah, Dan, Robyn, Asa, Elsje, Vince, Jeremy, Andy and others to numerous to name. Our long walks, car trips, all night conversations, were a font of wisdom and peace on the pathway to a published thesis. I can’t thank you enough for all the times I slept in your spare rooms or on your sofas.

Special thanks goes to all those that I collaborated with at the University and the UHCW, particularly Sunil Daga who was a wonderful lab partner, my supervisors Neil evans, Daniel Zehnder and Dan Mitchel.

Finally thanks goes to God, the father and redeemer of all; who made us all from Aristotle to Zorn in the image of his love and wisdom; of whom Solomon wrote “God delights in concealing things so that kings may delight in discovering them.” (Prov 25:2); who thus

## *ABBREVIATIONS*

---

can be said to crown us all who have worked in the sciences, by sharing in the glory of discovery; whose light illuminates and shines through the many gems that I have listed here.

Thank you all.

# Abstract

Patient antibody reacting against foreign antigen is a key part of both primary and secondary immune responses. However previous methods of studying the affinity of these reactions have been semi-quantitative and do not allow for the unique identification of the affinity of interactions. The goal of this PhD project was to produce mathematical models and experimental protocols that could be used with existing biosensor technology to accurately and repeatably determine the kinetic rate constants for monoclonal antibody as well as clinical samples of patient antibody.

Surface Plasmon Resonance (SPR) experiments with human antibody were conducted and repeated. A variety of models was developed and fit alongside models from the literature, until one model was able to consistently and accurately estimate the kinetic constants of the interaction as well as minimize residuals, the Bivalent effective rate constant spacial model (BERCS). This model was applied to further biosensor experiments in which each antibody was reacted against multiple samples of antigen, each presenting the same epitope alongside different structural features. The BERCS model dramatically reduced the residual sum of squares of the fit, compared to the fits done with commercially available software and models, and produced little variation in rate constant estimates between replicate experiments. This allowed for measurements of how the rate constants varied between the different structures of antigen that present the same epitope.

However, structural identifiability analysis consistently showed that SPR technology would not be able to identify the affinity of antibody in unknown concentrations - limiting the applicability of the technology for dealing with samples taken from patients. The growing complexity of these models lead to the necessity of novel methods for approaching

structural identifiability. New algorithms for analysing structural identifiability were created alongside new concepts relating to structural identifiability allowing users to distinguish between identifiability problems arising from the underlying state-space system and problems caused by its output.

SPR experiments with antibody samples taken from sensitised kidney transplant patients were initially processed with the XPR Proteon Manager software, which measured affinities as high as  $3.7 \times 10^{-24}\text{M}$  – vastly in excess of the highest antibody affinities expected for antibodies. The sensorgrams produced by these biosensors were analysed and demonstrated to show features that had not been observed in the monoclonal experiments; most importantly very slow total rates of antibody dissociation. A model was developed to explain the emergence of these features from interactions between proteins in patient serum, the generalised polyclonal model (GPM). SPR experiments were conducted with mixtures of monoclonal antibody, to simulate possible effects of polyclonal antibody binding in the patient serum. These experiments also saw slower total rates of antibody dissociation, than would be expected given their affinities, confirming the significance of the mechanism identified as causing the slow dissociation of the polyclonal samples.

Both the mixed monoclonal and patient serum experiments were fit with versions of the GPM. For the simulated patient serum, the GPM was able to accurately calculate the concentrations of antibody in the mixture where their affinities were known from the earlier study – confirming its reliability. The analysis of the patient serum was limited by the structural identifiability of GPM and its variants so only gave minimum affinities required to produce such an output, rather than affinities for the antibody in patient samples.

New models were developed, and new experiments were designed that took advantage of the properties of other biosensors so that they could be used to identify the affinities of patient antibody- even with mixtures of antibody of unknown concentration.

The significance of the ability to study the affinity of antibody even in unknown concentrations is that high affinity may be a biomarker for problems such as autoimmune diseases and transplant rejection as well as correct immune response to pathogens and cancer. As a result it may enable better screening in a variety of medical environments. The significance of the ability to increase the apparent affinity of antibody may also be a useful tool in the developing field of antibody medicine.

# Chapter 1

## Introduction

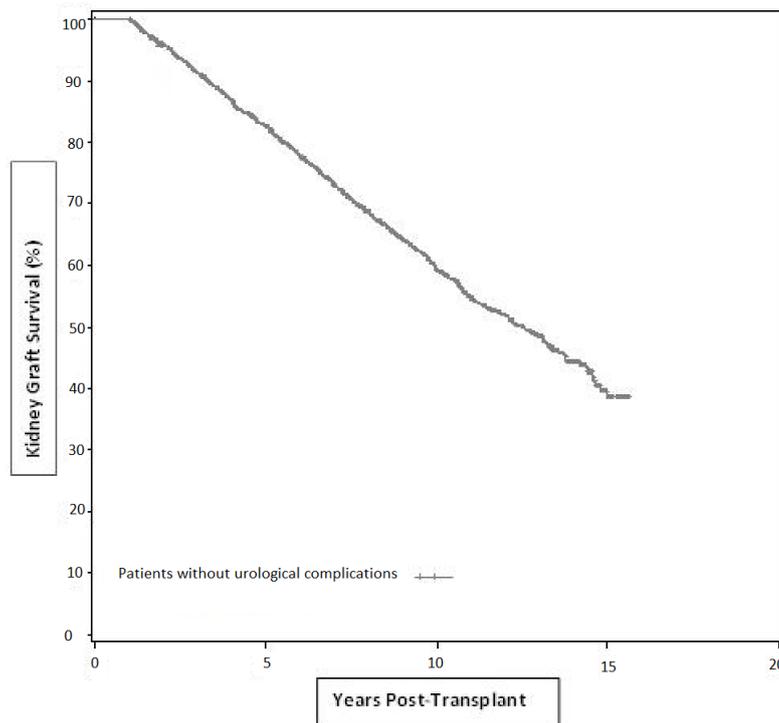
### 1.1 Transplantation

After five decades of kidney transplantation, the procedure has become the best treatment for renal failure in terms of life expectancy, quality of life and cost-effectiveness (Wolfe *et al.*, 1999). Chronic kidney disease, with its end stage being renal failure, affects 10-16% of the adult population in Asia, Europe, Australia and the United States (Nitsch *et al.*, 2013). As a result, the need for transplant organs is growing globally and doubled in the UK between 1995 and 2005 (El Nahas, 2005).

However, all transplantation carries with it the risk of rejection, so risk stratification is necessary. The complement dependent cytotoxicity (CDC) assay has remained for decades the standard test for compatibility used internationally (Patel, 1969), although it is not able to detect concentrations of antibody or their affinity for donor antigen and successful transplants have been made despite detection of donor antibody against the transplant (Higgins *et al.*, 2011).

On the basis of CDC assay, patients are divided into low-risk, non-sensitised patients and high-risk highly sensitised patients.

For low-risk patients, direct transplantation is the best treatment, for high-risk patient's removal of circulating donor-specific antibodies (DSA) and immunosuppression are asso-



**Figure 1.1:** Long term post transplant graft survival for patients without urological complications adapted from Arpali E (2016).

ciated with improved outcomes, but significantly worse outcomes than the former.

However, transplantation itself increases patient’s sensitisation. For low-risk patients, this can mean that they move into the high-risk category. For sensitised patients, repeated transplant failure can lead to a feedback loop of increased sensitisation and rejection.

This effect of increased sensitisation is compounded by the impermanence of transplant organs. The percentage of organs surviving after a given period of time is shown in Fig. 1.1. As a result, a single initially non-sensitised patient may have many organs over a lifetime, whilst increasing in sensitisation.

An additional complexity is that many patients suffer from chronic kidney disease with low or intermediate antibody levels measured by the current methods. For these patients

the best treatment is uncertain. As a result, some patients have an increased wait for transplantation or are denied it (NHSBT, 2016).

The primary goal of this thesis was to develop experimental and corresponding mathematical techniques for studying antibody-antigen kinetics for use in a clinical environment. Such investigation would allow for future medical studies into the effect of antibody-antigen kinetics on transplant outcome, and to improve organ matching.

In greater detail, the motivation for this is that the presence of antibodies directed against human leukocyte antigen (HLA) in the patient, years after the transplant has been identified repeatedly as a predictor of chronic rejection and graft loss (Mao *et al.*, 2007; Lee *et al.*, 2009) and established as a cause (Terasaki & Cai, 2008). The significance of HLA is that it is the class of antigen that includes the antigen on the tissue of the transplant that may differ from the recipient's antigen.

At the same time as HLA specific antibody has been established as a cause of rejection, there has been continued improvement in short time transplant organ survival, but long-term transplant survival has seen little change (Collins *et al.*, 2014; Meier-Kriesche *et al.*, 2004). Significantly, the rate of ten-year patient survival from deceased donor transplant in adult patients is just 75% (NHSBT, 2016).

The gap between improvements in short term and long term transplant survival has been attributed to improvements in immunosuppression, and chronic kidney disease has been highlighted as the dominant cause of long-term transplant failure (Nankivell *et al.*, 2003). Significantly Nankivell & Chapman (2006) established that late transplant failure is due to cumulative damage, and divided the mechanisms that caused damage to the organ into two groups immune and nonimmune, noting that damage was reduced by immunosuppression.

Despite the fact that post-transplant development of HLA antibodies is an indicator of

graft loss, it has been established that abundance of HLA antibody in a patient's body pretransplant is not an indicator of rejection (Higgins *et al.*, 2007). As a result, there is a need to study patient antibody not just in terms of its abundance, but its characteristics, and particularly its reactions against the transplant organ.

Intuitively we would expect that high-affinity antibody would be an indicator of a strong immune response, and it has previously been demonstrated that higher antibody affinity causes an increased pathogenic response in the autoimmune condition lupus nephritis (Williams *et al.*, 1999).

As a result, it was considered necessary to develop new techniques for studying antibody-antigen interactions. Beyond transplantation, antibody interactions are a key part of both the primary and secondary immune response. Improving technology for measuring these interactions may have many other applications, not just recognising undesirable immune responses in cases that threaten a transplant, or in autoimmune disorders, but also recognising right response to infections, pathogens and cancers.

The potential benefits of developing more accurate tools for analysing antibody-antigen affinity for clinical use are that it could allow for new studies allowing for better understanding of the risk factors of patient antibody for transplant organs; particularly by differentiating between different affinities and types of antibody against donor HLA.

By creating tools for this use, this work also paves the way for improved kidney matching, risk stratification and immunosuppression strategies.

Outside of transplantation, this research has pharmaceutical applications. Monoclonal antibody (mAb) treatments have been developed for a wide range of diseases including Alzheimer's, autoimmune disorders and cancer.

Key to the function of these treatments is the specific binding of the mAb to a target receptor, and not to other patient proteins. Highly accurate measurements of antibody-

antigen affinity measurement will help researchers develop mAb drugs that bind to the desired targets, with the desired properties.

One of the results presented in this thesis is that in some cases patient samples of antibody were measured to have much higher affinities against HLA than could be explained by a single antibody-antigen interaction. As a result, experiments were conducted with mixtures of mAb where the combination of antibodies lead to higher bound concentrations and longer lasting antigen-antibody binding.

This points to the possibility of using combinations of monoclonal antibody together to produce oligoclonal (oAb) antibody therapies, which could bind to new targets and more permanently than would be possible for mAbs.

However before these topics can be discussed some discussion of antibody, antigen and the immune system more generally is necessary.

## **1.2 Antibody, antigen and the immune system**

The immune system is the system that protects the human body from infection. It does this by recognising and destroying foreign substances. Antibody-antigen binding is one of the mechanisms necessary for this.

Any biomolecule that antibody, or the highly variable receptors of T cells or B Cells, binds to is termed an antigen. The term antigen is generally used to describe large peptides, polysaccharides or lipids - although in this thesis all the antigen considered will be peptides. The sites on the antigen that an antibody binds to are termed epitopes. Whilst each antibody will only have one epitope it has adapted to bind to, that epitope may occur on many different antigens and the antibody may bind nonspecifically to other epitopes.

Each antigen will generally have a large number of different epitopes. This means that multiple antibodies can bind to it, but a single antibody will probably not bind in multiple places.

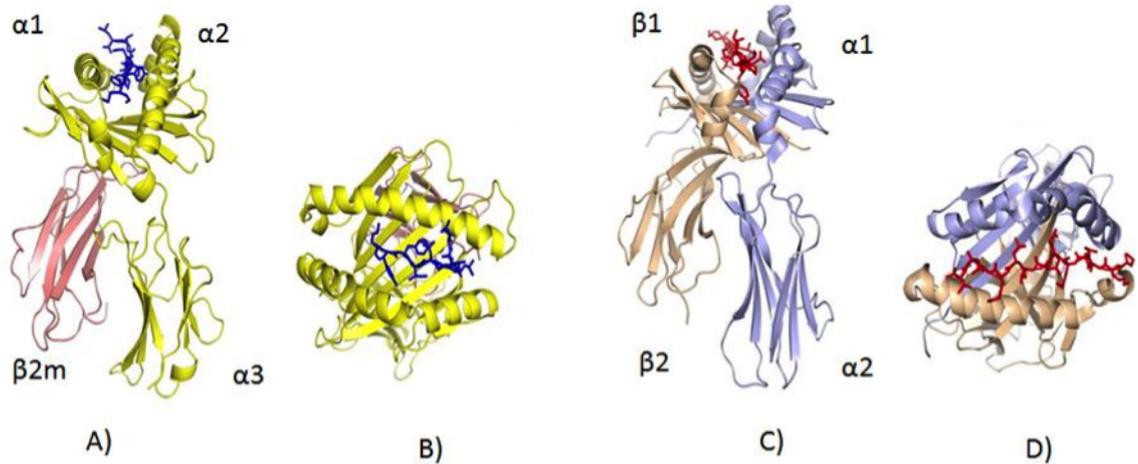
Corresponding to these epitopes the binding sites on antibody are called paratopes. They are hypervariable binding sites on the arms of the antibody molecule, that bind to the antigen-epitopes. Each antibody has multiple identical paratopes, as a result it can bind monovalently – with one of its paratopes attached to an epitope on a single antigen, or it may bind bivalently or multivalently, depending on its number of paratopes, with each of them bound to epitopes of antigen- possibly on the surface of some bacterial or cancer cell surface.

The majority of the antigens that were used in the experiments presented in this thesis were HLA. These are a set of highly polymorphic cell surface proteins that regulate the immune system, and allow the immune system to distinguish between self and non-self tissue. They are also referred to as the major histocompatibility complex (MHC).

HLA are grouped into three classes depending on their function. MHC Class 1 present peptides from inside the cell to the body's T-cells. MHC class 2 present peptides from outside of the cell. Complexes from either class are illustrated in Fig. 1.2. These functions inform the immune system of the state of the cell and surrounding tissue, for example, the presence or absence of proteins that might indicate infection or cancer.

Class 3 antigen encode the complement system - a part of the immune system that assists phagocyte and antibody in removing microbes and damaged cells from an organism.

The HLAs that will be used primarily in the experiments in this thesis are of MHC class 1. This class is divided by the genes that the proteins in the class are products of: HLA-A, HLA-B and HLA-C. Numerous alleles of the HLA genes and the resulting proteins have been classified along with their epitopes. Table 1.1 shows the result of one study which



**Figure 1.2:** Views from multiple angles of the molecular structure of Class 1 antigen in A) and B) and Class 2 antigen in C) and D) presenting a a peptide (cellular peptide in blue, extra celular peptide in red), taken from Gherardi (2008).

classified antigen and their alleles (Lowe, 2013), and Fig. 1.3 shows the number of alleles of class 1 and 2 discovered by year since they were first classified in 1989.

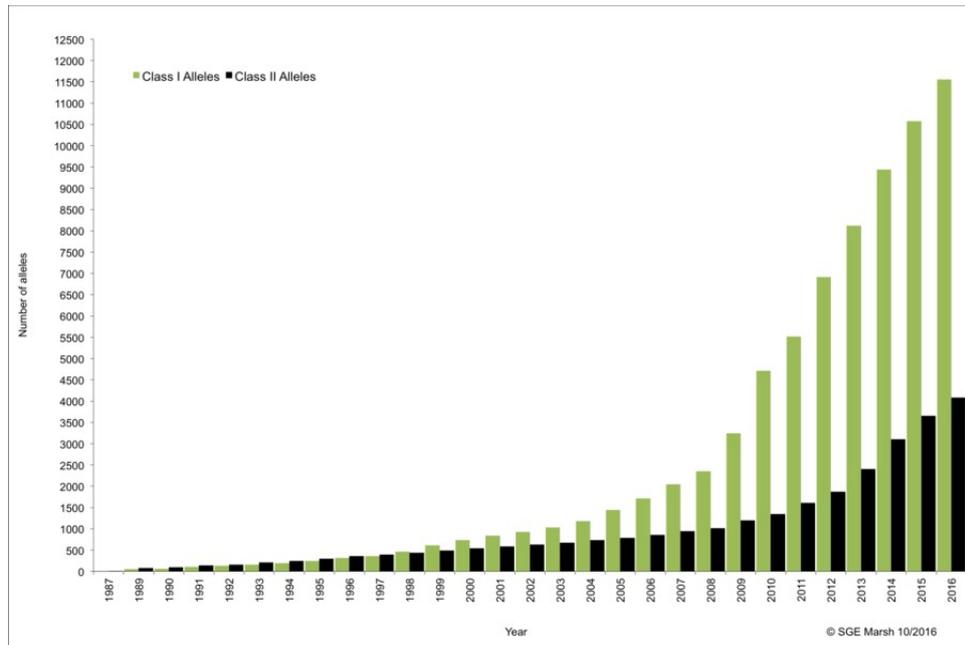
Figure 1.4 illustrates the location of the epitopes of an HLA-A2 molecule. Notably, they are not all on or near the peptide it is presenting. There are two spatial relationships that antigen binding sites can have with each other, either they can overlap or not. If they overlap then if an antibody binds to one, it may obscure the other and create steric hinderance to a second antibody binding to the second epitope. If they do not overlap, this does not happen.

However, there is a third relationship that the two epitopes may have. One epitope may be an allosteric site; that is when antibody bind to this it creates a conformational change in the antigen. This conformational change could alter the presentation of another epitope either making it more or less accessible to another antibody. (Abbas *et al.*, 2015).

When the human immune system encounters a new substance, it recognises it as foreign if molecules on the surface of the antigen differ from the body's own. This involves specialised cells such as macrophages and dendritic cells capturing the invader and breaking it down

Epitope Specificity	A1	A11	A2	A23	A24	A25	A26	A29	A3	A30	A31	A32	A33	A34	A36	A43	A66	A68	A69	A74	A80	
44K+																						
107W																						
65G																						
149T																						
62L																						
161D																						
65Q																						
163R																						
144K																						
166D+																						
76A																						
90D																						
142T+																						
127K																						
80I																						
82L+																						
149T																						
62E																						
56E+*62E+*163E+																						
19K																						
56R																						
167W																						
109L																						
56R+73T																						
253Q																						
62G+																						
127K+*151R+																						
156Q+																						
109L+*163T																						
109L+*131R																						
158V																						
144K+																						
163R+																						
76V+*144K																						
125E+*156W																						
80+																						
138M+*144Q																						
62R+																						
62R+*163T																						
90H+*171H																						
166E+																						

Table 1.1: Table of HLA- A antigen and their epitopes taken from Daga (2015).

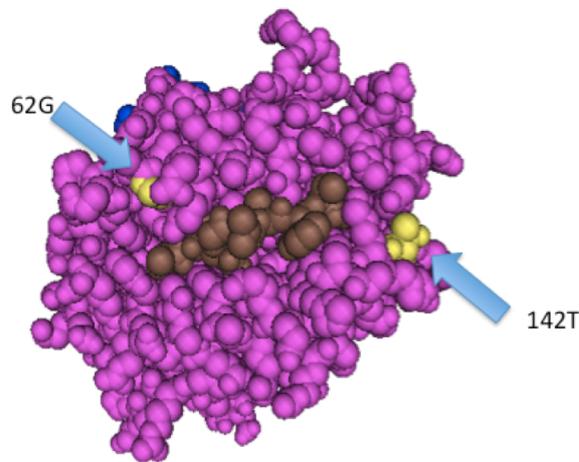


**Figure 1.3:** Numbers of discovered alleles by year, since 1989 taken from Robinson *et al.* (2015).

into parts (antigen) that can be presented to B cell lymphocytes.

When a B lymphocyte binds to an antigen, it divides and matures into a group of plasma cells and dormant memory cells. During proliferation, the B cell receptor locus undergoes somatic hypermutation. Variation in the genome of the created plasma cells is focused at 'hotspots' in the DNA, which determine the structure of the new B-cells antigen receptor, which will be identical to the paratope of the antibody it produces. As a result each B cell lymphocyte produces one unique antibody against one unique antigen epitope.

When these proliferated cells are exposed to the antigen, the variation between their antigen binding receptors will mean that the cells with receptors with the highest affinity for the antigen will spend the most time bound to the antigen, so will be selected to produce antibodies. These antibodies then bind specifically with the pathogen and allow the immune system to eliminate the pathogen from the system. At the same time the B cells that do not bind undergo apoptosis.



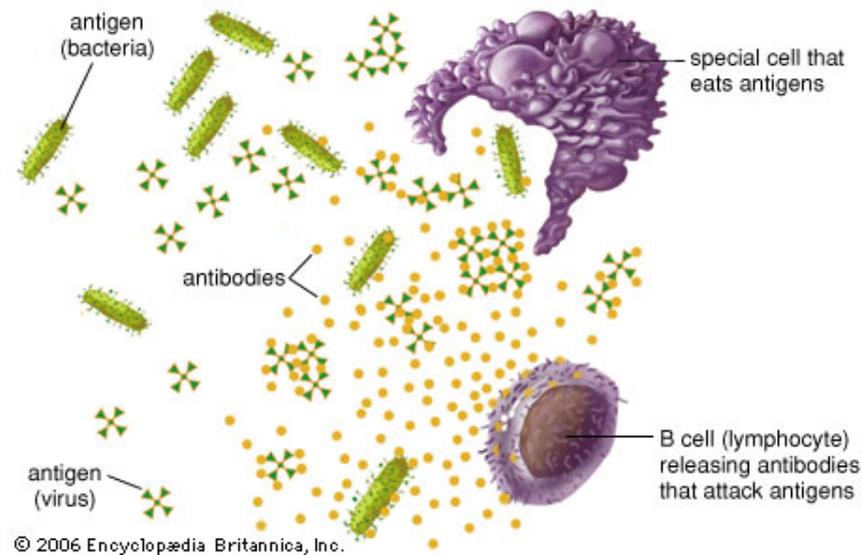
**Figure 1.4:** Diagram of HLA-A2 showing the epitopes 62G and 142T (yellow) and the protein presented (brown) from Lowe (2013).

In some cases, these antibodies can neutralise pathogens such as viruses directly due to the binding action. In other cases, such as with bacterial pathogens, these antibodies bind to surface proteins on the bacterium's surface, thereby signalling to the rest of the immune system that the pathogen should be destroyed.

After the foreign molecule has been eliminated, the immune system retains the dormant memory B cells, so that on exposure to the same antigen the immune system can reproduce the antibodies that were involved in the original response, and continue the hypermutation and selection process leading to qualitatively and quantitatively higher responses (Roitt *et al.*, 2017; Murphy, 2001; Owen *et al.*, 2013).

This process is illustrated in Fig. 1.5. Due to this process, the antibody mixtures that occur in patients are polyclonal; that is, there is a mixture of a vast number of antibody clones each with its own epitope target and affinity for that target.

Monoclonal antibodies are artificially produced. The most common method for the manufacture of monoclonal antibody is shown in Fig. 1.6. A mouse is injected with an antigen. Antibody producing B cells from the mouse's spleen are taken and fused by



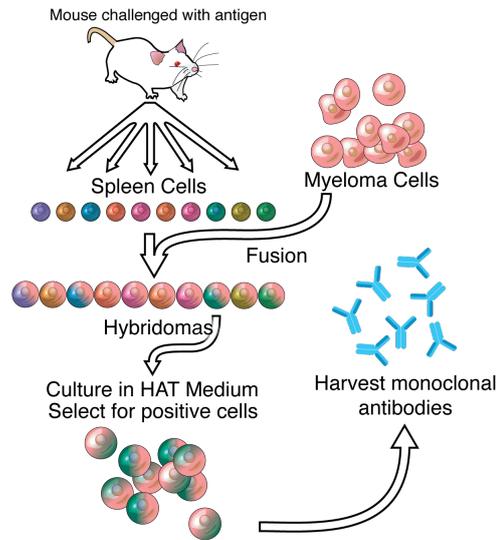
**Figure 1.5:** Phagocytic cells destroy viral and bacterial antigens by eating them, while B cells produce antibodies that bind to and inactivate antigens from Britannica (2016).

myeloma tumour cells to make hybridomas. These can then be grown in culture, each culture starting with one hybridoma which grows into a culture of genetically identical hybridomas producing identical antibodies (Liu, 2014).

Diagrams of antibody are shown in Fig. 1.7. The monoclonal antibody used in this project was in every case IgG. However, the samples taken from patients would be expected to contain mixtures of the antibody isotypes. The monoclonal samples are HLA antibody, antibody with paratopes that will bind to the epitopes of HLA.

Figure 1.7 illustrates the physical differences between antibody isotypes. IgG, IgD and IgE, as well as the IgA monomer, are all similar sizes, whereas IgM and the IgA dimer are much larger.

The larger form of IgA, the secretory form, gets its name because it is the antibody most abundant in mucous secretions (tears, saliva, milk, etc.). This happens due to a transport process, where molecules with J-chains are transported from the basal to the apical sides of epithelial cells (F.E. *et al.*, 2000). It has a molecular weight of 600 kDa. Typically they

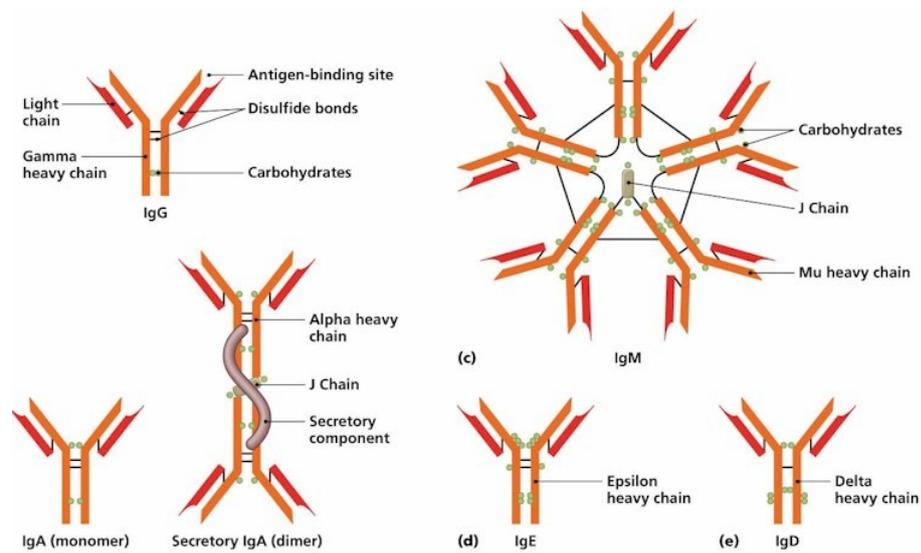


**Figure 1.6:** A general representation of the method used to produce monoclonal antibodies from Wikimedia Commons (2010).

make up 15% of total antibody in human bodies. However, in the blood, they are only found in their smaller form which weighs 160 kDa and only forms a tiny proportion of total blood antibody.

IgM is the largest form of antibody and is initially produced when the immune system is exposed to a new antigen. It is important both for protecting against infection, and for graft survival. It occurs in four distinct forms, most commonly as the pentameric unit with J-chain shown in Fig. 1.7 which weighs 970 kDa, a monomeric unit on the surface of B cells, where it functions as an antigen binding receptor and as both pentameric and hexameric units lacking J-chains.

IgE antibodies are the least common, but in people with allergies, they may be more abundant. Of the monomeric antibodies, they are the heaviest, having a molecular weight of 190 kDa. They act as a defence against parasites, particularly helminths (parasitic worms) and are found primarily in the lungs, skin and mucous membranes. In allergic individuals, they mediate the immune response to antigen (also called allergen) in the



**Figure 1.7:** Illustration of antibody structures from Cummings (2007).

substances the individual is allergic to.

IgG antibodies are the smallest, weighing from 150-170 kDa and appear in all body fluids. Through a specific pathway, they are even able to cross the placenta in a mother's womb. They make up around 75% to 80% of all of the antibodies in the body and are the predominant antibody involved in the secondary immune response, that is, the immune system's response to a foreign antigen that it has already been exposed to. As a result of this role, they are significant in dealing with infection in general but are especially relevant to the subject of long term graft survival.

IgD has a molecular weight of 180 kDa and is primarily found along with IgM in the plasma membranes of immature B cell lymphocytes, but their overall function in the immune system is unclear.

The isotypes of antibody take their names from the type of heavy chain that make up the base of the Y shaped units. The types of heavy chain are alpha, gamma, delta, epsilon, and mu, corresponding to the isotypes IgA, G, D, E, and M, respectively. The section of

the antibody made of these heavy chains is also called the Fc region. These Fc regions interact differently with Fc receptors on other immune molecules allowing each antibody isotype to have different effects, including agglutination, haemolysis, complement activation, opsonisation, mast cell degranulation, and neutralisation (Murphy, 2001).

The binding of antibody paratope to antigen epitope happens through the interaction of four (non-covalent) forces that operate over differing scales. These are: electrostatic forces and hydrogen bonds which operate over a comparatively long range, and largely determine the reaction affinity; Van der Waals forces - atoms in either structure oppositely polarising nearby atoms; and hydrophobic forces that force together hydrophobic molecules (Murphy, 2001).

Typically these forces lead to antibody antigen interactions having a  $K_D$  (a measure of affinity) between  $10E-7$  M and  $10E-9$  M in the primary immune response, and  $10E-10$  M and  $10E-11$  M in the secondary immune response (Berek, 1987). The highest antibody affinity measured are around  $5E-12$  M (Zahnd *et al.*, 2004; Abcam, 2013a).

### 1.3 Affinity and Avidity

The overarching goal of this project is to develop a way to study antibody-antigen interactions, for pharmaceutical and clinical use. However, due to the complex structures of antibody and antigen, we would expect to see a variety of effects when we react the two together. As a result, it is useful to have a language with which to discuss them.

Affinity is a measure of dynamic equilibrium concentrations of reactants and products for a reaction in which two molecules bind at a single interaction site.

For example, many antibody drugs use a piece of the antibody rather than the whole antibody, commonly the fragment antigen-binding (Fab) region - the antibody arm that ends with the paratope (Liu, 2014); if the detached Fab regions were reacted against

antigen with the corresponding epitopes they would have a simple reaction scheme:



where  $[F]$  and  $[A]$  are concentrations of the Fab regions (in M), and antigen respectively and  $k_a$  and  $k_d$  association rate constant ( $\text{M}^{-1}\text{s}^{-1}$ ) and dissociation rate constant ( $\text{s}^{-1}$ ) respectively. Notably the association rate constant is called a rate constant in the SPR literature even though it does not have a unit of ( $\text{s}^{-1}$ ) (Richard B. M. Schasfoort, 2008).

Then the affinity or dissociation constant of the reaction (measured in M) is

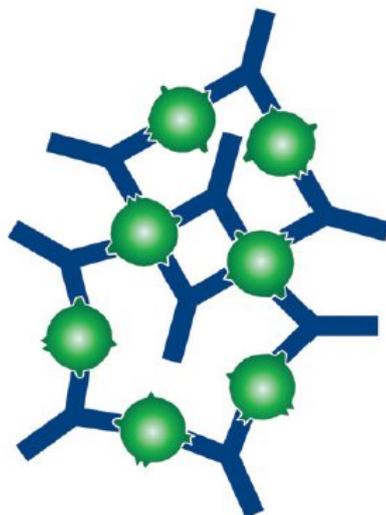
$$K_D = \frac{k_d}{k_a} = \frac{[F]^*[A]^*}{[FA]^*}, \quad (1.2)$$

where  $[F]^*$  and  $[A]^*$  and  $[FA]^*$  are the equilibrium values of  $[F]$  and  $[A]$  and  $[FA]$  respectively.

Notably whilst the dissociation constant (affinity) is defined in terms of the ratios between the equilibrium concentrations of the reactants, this is equivalent to a ratio of the rate constants. In more complex systems in this thesis, the word affinity will be used to mean the ratio of the rate constants.

Whilst affinity is a measure of the strength of interaction between two binding sites, avidity is a measure of the overall strength of the complex formed. In this thesis that complex may be a single antibody bound to multiple antigen or a group of antibody bound to multiple binding sites on multiple antigen. Avidity is the accumulated strength of multiple binding interactions acting together and is commonly referred to as a functional affinity (Rudnick & Adams, 2009; Sennhauser *et al.*, 1989).

As shown in Fig. 1.7 antibodies of all isotypes have multiple identical Fab regions. The

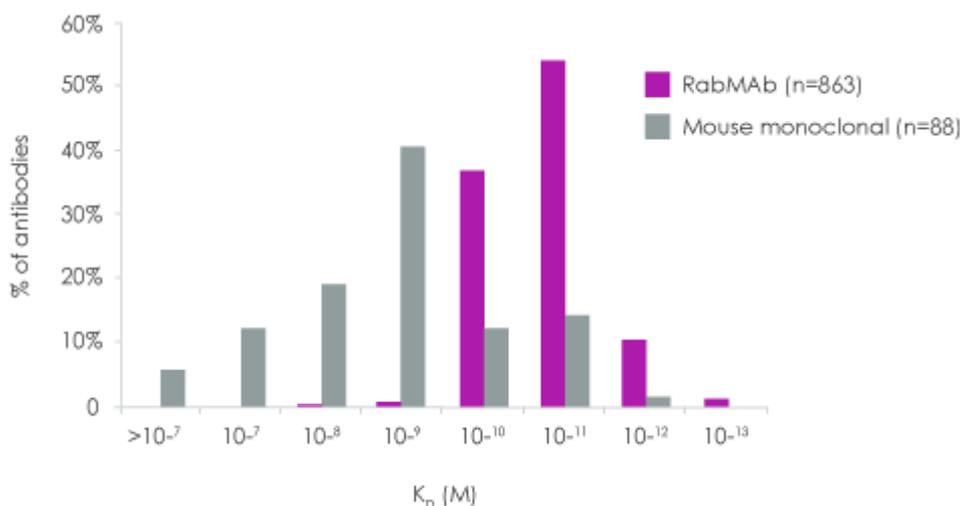


**Figure 1.8:** When an antigen is mixed with a polyclonal antibody, multivalent interactions may lead to large, stable (high avidity) structures being formed. This is because the antigen may be bound by several antibodies, each recognising a different epitope. Taken from BioRad (2010).

greater its valency (number of antigen binding sites), the greater the number of antigen an individual antibody can simultaneously bind to. Similarly, antigens with multiple epitopes are multivalent because they can bind to more than one antibody, although these antibody have to have different paratopes corresponding to the epitopes they are binding to on the antigen. As a result, large structures of antibody and antigen can develop which have much greater stability. This is illustrated in Fig.1.8.

Notably, the kinds of structure created are going to be determined not just by the properties of the reactants but their concentrations - if one antibody is in relatively low concentration as a reactant, it will not be as abundant in the structures produced (Roitt *et al.*, 2017; BioRad, 2010; Kitov & Bundle, 2003).

Figure 1.9 shows a distrobution of  $K_D$  measurements of commercially available monoclonal antibodies, taken from Abcam (2013a,b), illustrating the range of measurements we might expect from our own studies.



**Figure 1.9:** Affinity measurement distribution for abcam rabbit mAb (rabMAb) antibodies and mouse mAbs, showing 88  $K_d$  values for mouse mAbs derived from published literature, using various methods and approaches and 863  $K_d$  values for rabbit mAb antibodies taken from from abcams own measurement project. Taken from Abcam (2013a,b).

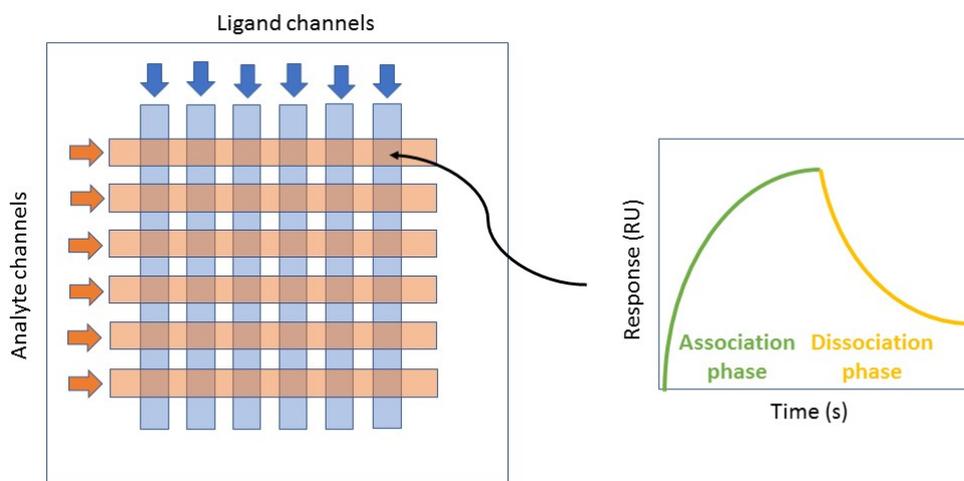
## 1.4 Surface plasmon resonance technology

Due to the complexity of the interactions being studied experimental apparatus that would provide large amounts of information was required.

In Chapter 4 the designs of the experiments will be discussed in detail, but it is useful to give some explanation here as to why surface plasmon resonance biosensors were used as this technique is focus for all of the model development and analysis.

The XPR36, a surface plasmon resonance biosensor, can measure the concentrations of molecules bound to the surface of its interaction chip at 36 spots every 0.9 s for approximately 30 minutes, allowing a single experiment to generate 72000 data points (Luo, 2013). This is diagramed in Fig. 1.10.

Figure 1.10 shows typical experiment design for the XPR36 biosensor. Six concentrations

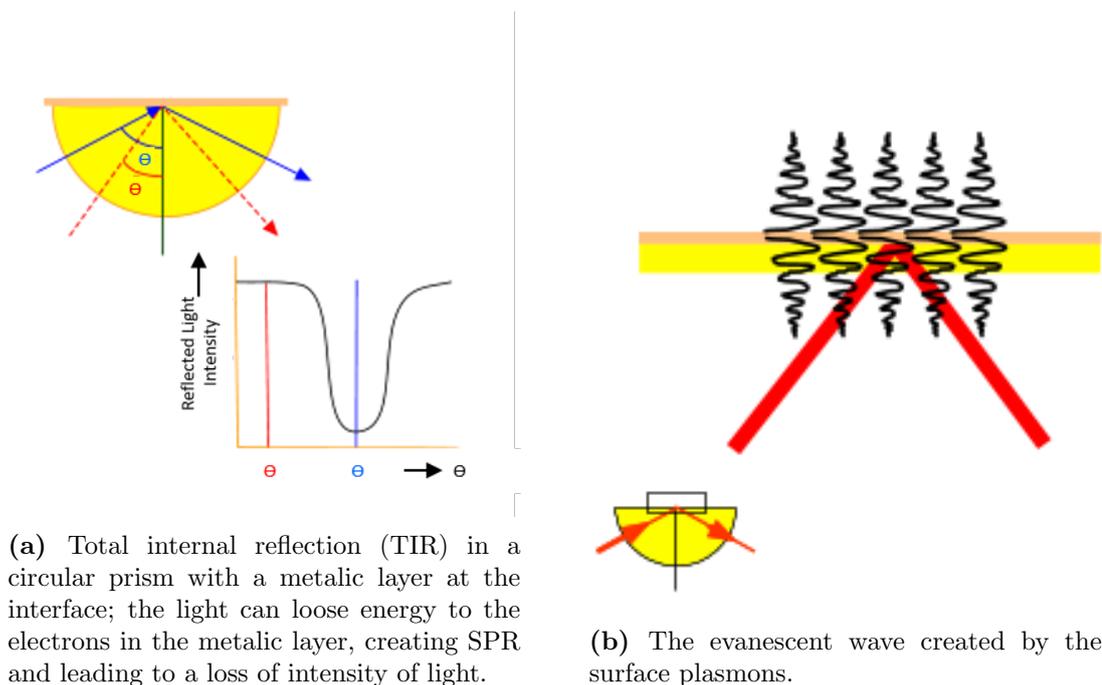


**Figure 1.10:** Diagram of typical experiment design for the XPR36 biosensor

of ligand (in this thesis this will generally be antigen) is passed over the six reaction channels of the chip and are immobilised on the dextran surface of the gold chip to create regions where the analyte can bind. The chip is then rotated, and six concentrations of analyte (in this thesis this will generally be antibody) are passed over the chip. The concentration of analyte bound to the ligand on the surface of the receptor is measured at each reaction spot, generating a response curve of the machine's measurements (in RU) against time. A group of response curves from the same experiment is referred to as a sensorgram, examples of this would include the panels of Fig. 4.2- Fig. 4.14.

Typically these experiments are split into two phases: association, where the six concentrations of analyte are passed over the chip; and dissociation where the analyte is replaced with buffer. This is so the rates at which the analyte binds to the ligand at varying concentrations can be seen as well as the rates at which the resulting complexes unbind.

The words ligand and analyte within the literature of surface plasmon resonance (SPR) experimentation refer to chemicals of interest, where ligands are always immobilised to the



**Figure 1.11:** Diagrams of TIR and SPR taken from (Marquart, 2013).

surface of the chip and analytes flow across the surface of the chip so that their reactions with the ligand can be measured. In this thesis all of the analytes used will be antibody and all of the ligand that will be used will be antigen; however many of the models and techniques discussed will not be designed for antibody-antigen interactions, so it is for the sake of accuracy necessary to use these words in some cases.

The technology and physics through which these biosensor measurements are made requires some explanation.

A plasmon is a quasi-particle resulting from the quantisation of plasma oscillations. It is to plasma as a photon is to light. Plasmons are created when electrons within a metallic surface are excited.

In SPR experiments this is done with a laser shone through a semi circular prism (see Fig. 1.12 and Fig. 10.1 a). The light is refracted by the surface of the glass producing many

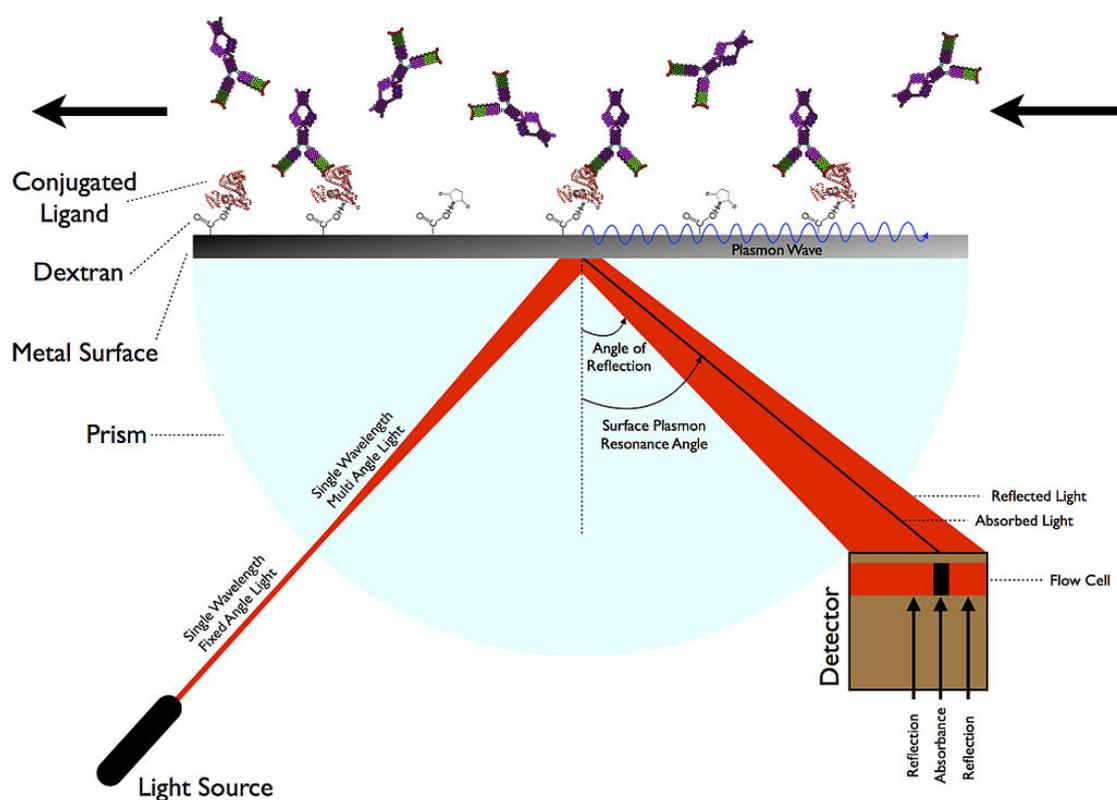
beams of light hitting the metallic layer. At the angle of surface plasmon resonance, energy from the light is absorbed rather than reflected. This energy excites the electrons in the metallic layer, creating oscillations of the electrons in the metallic layer (plasmons).

This is associated with an evanescent wave (illustrated in Fig. 10.1 b) with a wavelength equal to that of the incident light. The plasmons create a field that extends into the medium on either side of the metallic layer (including the dextran layer where the reactions take place as well as some of the analyte or buffer). This field is called the evanescent wave because the amplitude of the wave decreases exponentially away from the interface surface (without absorption), decaying over a distance of about one light wavelength (Sarid & Challener, 2010).

The way in which electromagnetic fields travel through a medium is affected by the properties of the medium. For example, the propagation of light in a medium is effected by its refractive index; light is refracted because it changes velocity as it moves from one media to another.

Similarly, the velocity (and therefore the momentum) of the plasmons is changed as the composition of the medium changes, as analyte binds to ligand. Because of the change in momentum of the plasmons, the angle of incident light at which the resonance occurs changes (Marquart, 2013). This change of angle is 122 millidegrees per  $1 \text{ ng}\cdot\text{mm}^{-2}$  of analyte bound to the sensor surface. As these units are inconvenient, RU are commonly used and will be in this thesis, where  $1000 \text{ RU} = 1 \text{ ngmm}^{-2}$  (Richard B. M. Schasfoort, 2008).

As a result biosensors like the XPR36 are able to use this physical effect to generate incredibly accurate measurements through experiments, which will be useful for studying antibody-antigen interactions. The potential of SPR technology as well as some of its current limitations are illustrated by an overview of the key antibody affinity studies in the literature.

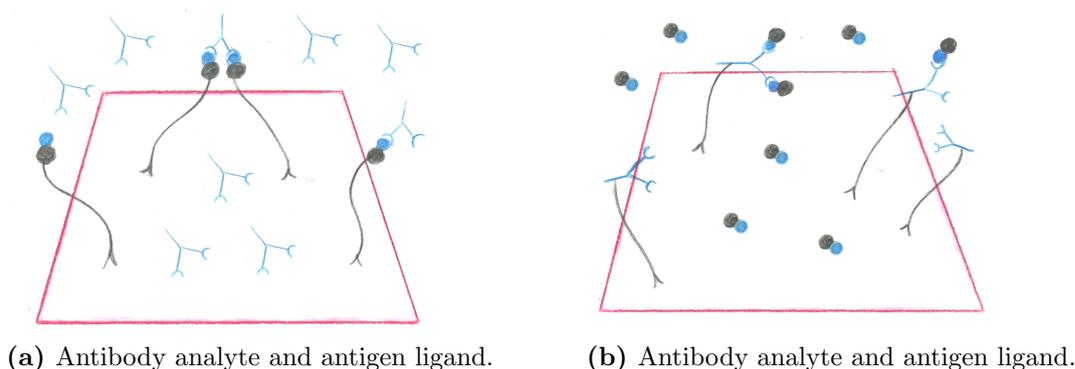


**Figure 1.12:** This figure shows an experimental configuration for an SPR chip. The metal chip (silver or gold) is prepared with a dextran surface which can bind the NH<sub>2</sub> end of protein to conjugate them to the metal surface. At the bottom a single wavelength laser beam enters a prism which results in many light angles striking the metal surface, all of them are reflected except for the angle in which the metal will absorb and turn its energy into a plasmon wave onto its outer surface, at this angle no light is reflected and thus appears with very little intensity on the detector. Since the plasmon wave propagates on the outer side of the metal, any interaction with the conjugated protein will change the resonance angle. Taken from Sabban (2011).

#### 1.4.1 SPR studies of monoclonal antibody affinity

There are two key groups of studies that are necessary to present, studies of monoclonal antibody affinities, and studies of the affinity of patient antibody - which are expected to be in unquantified concentrations of various polyclonal groups.

Monoclonal antibodies have been the subject of a number of large benchmark studies over



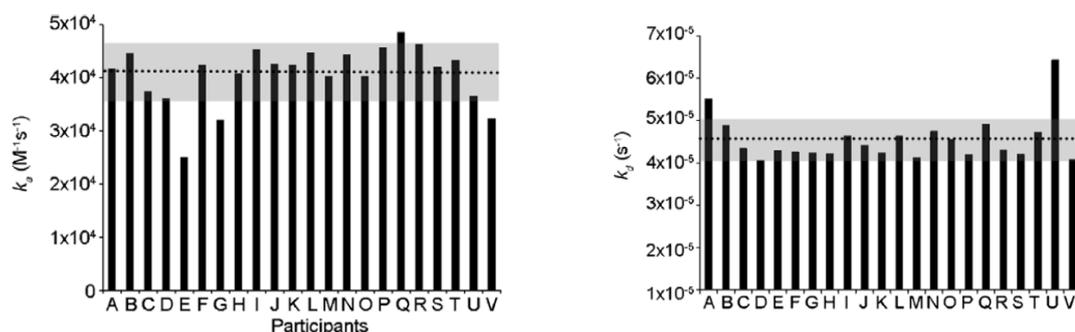
**Figure 1.13:** Illustrations of two configurations of SPR experiment for investigating antibody-antigen interactions.

the last two decades in which interactions were studied with numerous SPR platforms at different centres. In contrast, using SPR to measure patient antibody affinity is a relatively new idea and there have been a small number of published studies, each with much smaller numbers of experiments than the mAb studies.

The earliest of these studies is Katsamba *et al.* (2006) in which each participant immobilised antibody against the reaction chip surface and used antigen as the analyte. By using antigen as the analyte and antibody as the ligand, this eliminated the effect of the antibodies bivalence. That is when an antibody is the analyte it can bind to one antigen and then another, and if it unbinds from one it is still bound to the other, so remains within the evanescence field and contributes to the measurement the platform makes. In contrast, if the antigen is the analyte, although each mAb ligand may bind to multiple antigens, as each antigen is only attached to the chip via a single antibody it only needs to dissociate once to be able to leave the evanescence field, as shown in Fig. 1.13.

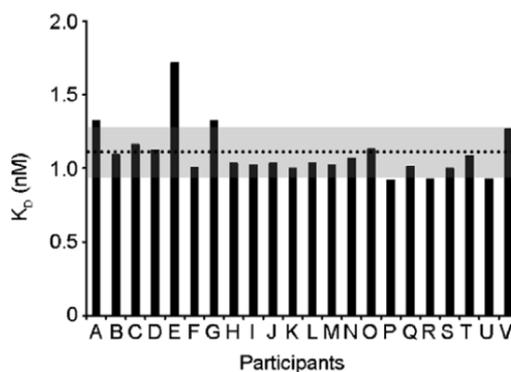
Using this experimental design Katsamba *et al.* (2006) was able to consistently and accurately estimate the affinity of the antigen-antibody interaction using the Langmuir model, as shown in Fig. 1.14.

These extremely consistent estimates have not been replicated by studies using antibody



(a) The association rate estimates. Their average is  $(4.1 \pm 0.6)E + 4M^{-1}s^{-1}$ .

(b) The dissociation rate estimates. Their average is  $(4.5 \pm 0.6)E - 5s^{-1}$ .

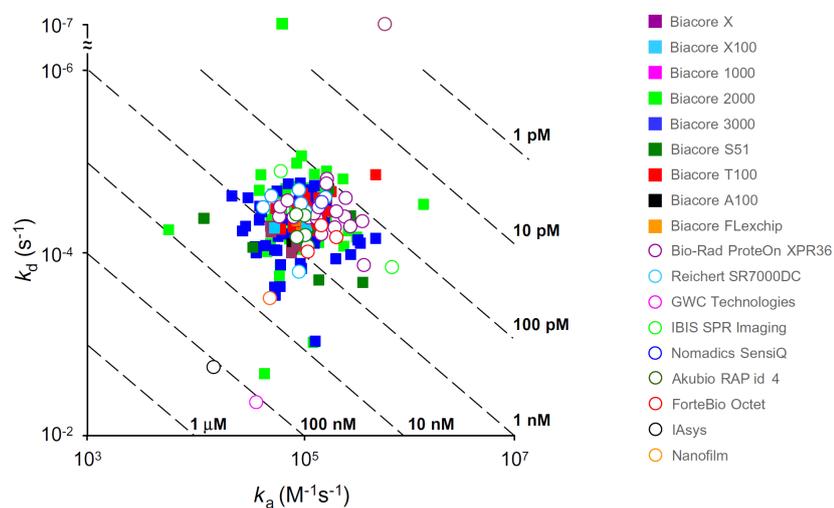


(c) The equilibrium binding constant estimates. Their average is  $(1.1 \pm 0.2)E - 9M$ .

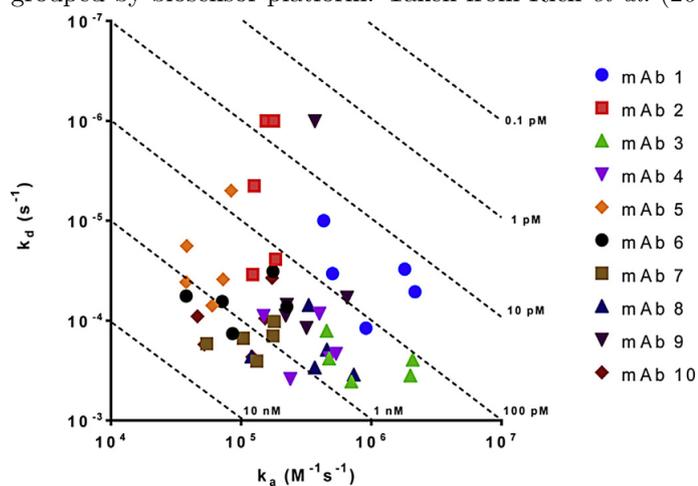
**Figure 1.14:** Distribution of the kinetic rates and the equilibrium binding constants, with average values marked using a dotted line and the boundaries of standard deviation highlighted in grey, adapted from Katsamba *et al.* (2006).

analyte and antigen ligands. Plots of affinity estimates,  $k_a$  and  $k_d$  from Yang *et al.* (2016) in which experimentalists which used antibody analyte and antigen ligand exclusively and Rich *et al.* (2009) where a combination of both methods were used but antibody analyte and antigen ligand were used primarily are shown in Fig. 1.15. For each study the estimates for a single interaction affinity and rate constant range over multiple orders of magnitude. As a result averages of estimates are not useful- as they are influenced by outliers that may be tens of thousands of times greater than the median results.

One possible reason for this is that the estimates in Yang *et al.* (2016) come mainly from the



(a) Rate constant estimates for a single monovalent interaction grouped by biosensor platform. Taken from Rich *et al.* (2009).



(b) Rate constants estimates for individual monovalent antibodies taken with multiple biosensors at multiple centers, taken from Yang *et al.* (2016).

**Figure 1.15:** Plots of estimateds of  $k_a$  and  $k_d$  along with lines of isoaffinity taken from two multinational benchmark studies.

same model as that used in Katsamba *et al.* (2006), the Langmuir model, which assumes incorrectly that a single antibody only binds to a single antigen, and that the concentration of antibody at the surface of the interaction chip is unaffected by the chemical fluid dynamics of the solution containing the antibody. Figure 1.15(a) illustrates the spread of estimates, raising the question could a model by including the more complex aspects of antibody binding, and mass transport processes make more accurate and consistent affinity estimates.

The estimates in Rich *et al.* (2009) are from models selected by the individual experimentalists in the project. The fact that there is not a single established model for antibody-antigen interactions despite the growing use of antibodies in medicine illustrates the need for the development of one.

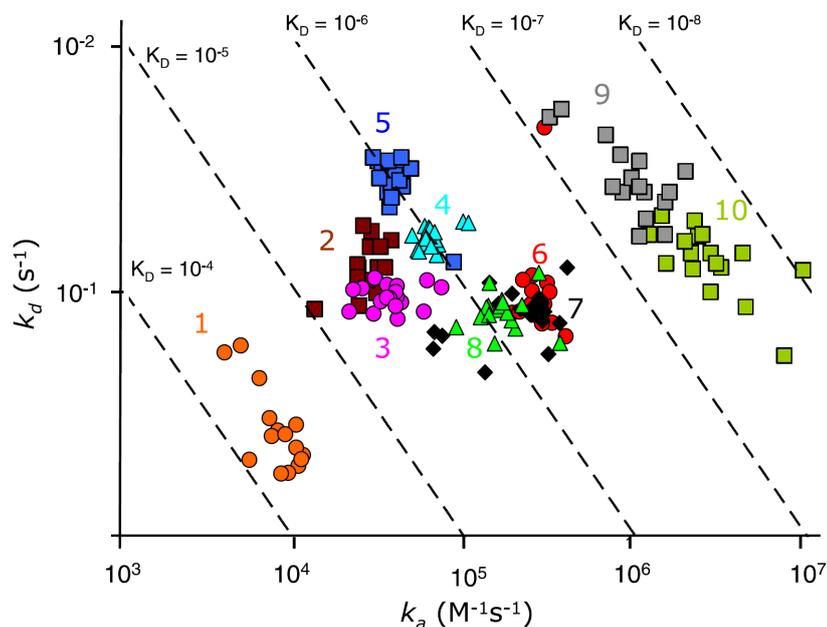
These studies contrast to non antibody studies like Papalia *et al.* (2006), which showed high levels of consistency for estimates of kinetic rate constants for their interactions, despite using the Langmuir model. the same model that was used in Yang *et al.* (2016) and predominantly in Rich *et al.* (2009) and produced such inconsistent estimates.

This is because in the case of both of these studies their experimental conditions were similar to the assumptions of the model - single analyte molecules bound to single ligand molecules, and the concentration of analyte able to bind to ligand remained close to constant.

Again, this highlights the need for a model that could be used for data from studies like Rich *et al.* (2009) and Yang *et al.* (2016) to produced consistent accurate results.

#### **1.4.2 SPR studies of patient antibody affinity**

Previously SPR experiments with polyclonal antibody were used to detect whether a patient had developed an antibody against a specific antigen (Ayela *et al.*, 2007), toxins

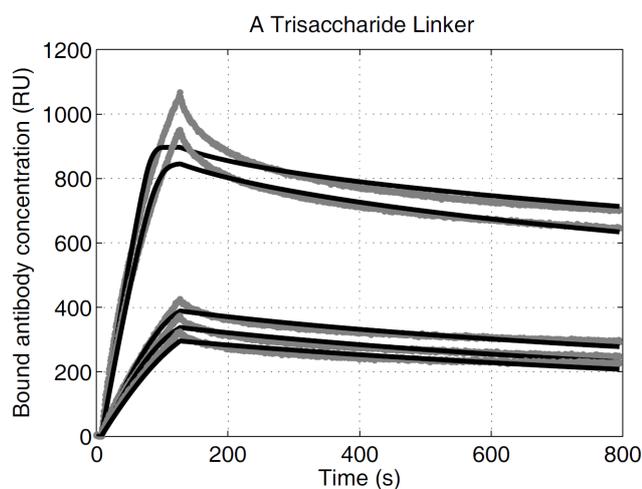


**Figure 1.16:** Plots of estimateds of  $k_a$  and  $k_d$  along with lines of isoaffinity for 10 interactions, taken from Papalia *et al.* (2006).

(Nabok *et al.*, 2005) and the presence of bacterial cells (Mazumdar *et al.*, 2007). However, only one study has so far measured the affinity of polyclonal antibody, Evans *et al.* (2010). This paper presented measurements of the affinity of patient ABO antibody with two models (the Langmuir with Transport model and effective rate constant, or ERC model).

The experimental work for this project came from a collaboration between two PhD students D.Lowe and Dr A.Bentall, who motivated by successes in the literature of studies like Katsambla (2006) wanted to develop tests for patient antibody interactions that would be more reliable and give more information than, haemagglutination, which detects antibody but does not indicate affinity, and despite being notoriously irreproducible has remained the standard method for detecting blood group incompatibility about 100 years (Aikawa *et al.*, 2003).

Figure 1.17 shows a typical fit from this paper, the model undercuts the data from every



**Figure 1.17:** ERC model output (black lines) plotted with corresponding experimental data (grey lines) taken from Evans *et al.* (2010).

interaction spot at its peak and runs close to the data from 300s onwards but in every case diverges from the data at the end of the phase. This demonstrates again the significance of developing novel models for the accurate and repeatable estimation of antibody-antigen affinity.

Figure 1.17 raises new questions. The data and model behave very differently. What features of the reactants are causing this discrepancy of behaviour? One possible answer is the same process that drives haemagglutination, individual antibodies binding bivalently to multiple antigen, but because different antibody clones may have different epitope targets each antigen may also be bound to multiple antibody, and as a result the development of more complex structures containing multiple antibody and antigen. Although this study became part of the PhD thesis Bentall (2014), there was no large scale follow up to it.

## 1.5 Model development and analysis

The complexity of the possible interactions of antibody and antigen as well as the SPR biosensor require the use of mathematical models to process the data generated (Marquart, 2013). As a result methods for fitting and comparing models were needed.

There are a variety of models that have been developed to describe the dynamics of a reaction between a flow of antigen across a ligand bound SPR sensor surface.

The most complex of these involve a system of partial differential equations (PDEs) to model the chemical fluid mechanics of the reactants (illustrated in Fig.1.18 (a)); these models have been used as stepping stones to the development of simpler ordinary differential equation (ODE) models that can be fitted to data. Examples of this include Myszka *et al.* (1998b) and Edwards *et al.* (1999) who consider the effects of a boundary layer between the well-mixed analyte and the ligand where reactions take place, and use this as a boundary condition for the transport equation (illustrated in Fig.1.18 (c)); and Yarmush *et al.* (1996) and Edwards. (2001.) who both consider the effects of three dimensional receptor layer that analyte must diffuse through before reacting.

The simplest model, the Langmuir model, uses a single ODE for analyte-ligand interaction assuming that throughout the experiment analyte is well mixed, and that reactant binding is solely dependent on the laws of mass action(Marquart, 2013) (illustrated in Fig. 1.18(b)). However, it has been demonstrated that the interactions measured are determined not only by the availability of both reactants but the effects of transport processes (Myszka *et al.*, 1998b). As a result, the use of these models in parameter fitting can introduce systematic errors in affinity estimation(Chaiken *et al.*, 1992).

Myszka *et al.* (1998b) offers a compromise between the two levels of detail, by assuming that the analyte on the boundary layer is in a quasi-steady state, a model may be

constructed consisting of only a single ODE in terms of bound analyte concentration. This model was shown to be a good approximation of the physical system under certain conditions (Edwards., 2001.). The models discussed were designed to deal with experiments where only one binding reaction is taking place, but antibodies have multiple binding domains and can bind in multiple configurations (with any specific arm or with any specific combination of arms).

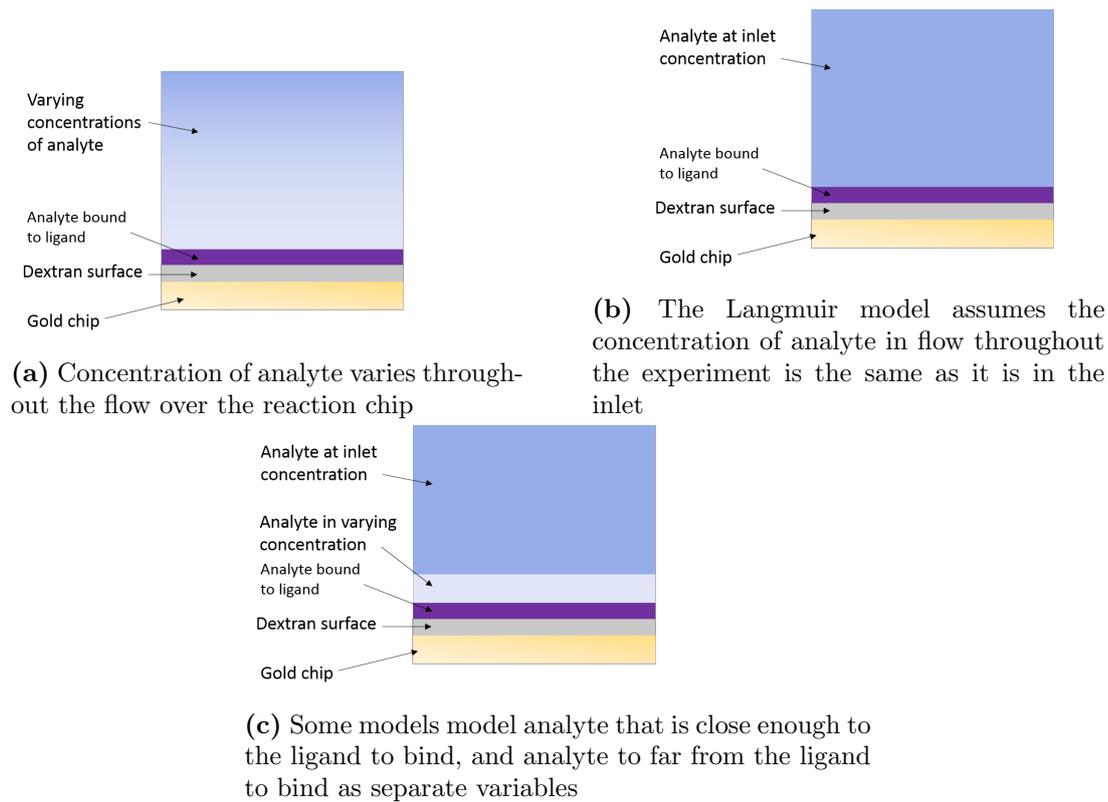
There is a model that has been developed specifically for bivalent binding analytes like antibody (BIAcore, 1998), however, it assumes that analyte throughout the experiment is well mixed, so is likely to suffer from the same systematic flaws as the Langmuir model.

Hearty *et al.* (2012) used a different approach for antibody-antigen experiments, adjusting the experimental conditions so that the Langmuir model would accurately describe interactions. They immobilised the antibody and passed the antigen in flow over the surface of the chip, meaning that the analyte could only bind monovalently.

A similar approach could not be used here, because this project is designed to have clinical application, and immobilising mixtures of patient antibody on the SPR chip would present whole new challenges. Additionally, patient antibody, as previously mentioned is expected to include clones against multiple antigen epitopes and a variety of isotypes, which will present new modelling challenges.

As a result, this thesis has a large focus on developing models that can adequately deal realistically with the inhomogenous concentrations of antibody in flow over the immobilised antigen, and the bivalent binding of the antibody to the antigen, as well as the more complex interactions of polyclonal antibody.

Whilst antibody-antigen SPR experiments are expected to include a complex interplay of bio-chemical and physical interactions as described above, the biosensors output propor-



**Figure 1.18:** Illustrations of three of the common assumptions about unbound analyte concentration used in models.

tional only to the concentration of analyte bound to the sensor. It is necessary to consider the identifiability of models as well as fitting them to data; that is, to establish whether there are multiple sets of parameters that would give an identical output.

Even in the case where each parameter of a model is physically meaningful, there may not be enough information in the output to uniquely determine the parameters for a model (Jacquez., 1996). Additionally, because the models we are developing are for clinical data where the concentration of analyte may be unknown special care has to be taken to ensure that their parameters can be meaningfully estimated from the output of an experiment where the analyte concentration would be unknown.

This problem of whether parameters of a model can be uniquely estimated regardless of

the quality of the experimental data or the way in which the model is fit is called *structural identifiability*. This is because whether the parameters of a model that can be uniquely identified is determined by the structure of the model, its system dynamics and the way in which it is observed.

For non-linear systems, such as the models discussed here, methods for the analysis include the Taylor series method (Pohjanpalo, 1978b), the generating series method (Walter, 1982), the similarity transformation approach (Vajda *et al.*, 1989), differential algebra based methods (Forsman, 1991; Ljung & Glad., 1994), the direct test (Denis-Vidal *et al.*, 2001) and a method based on the implicit function theorem (Xia X, 2003).

Due to the complexity of the models that were developed, new methods for analysing structural identifiability were developed to facilitate their analysis. The most commonly used of these methods involved adapting MAPLE code initially developed by Forsman (1991). However, this method was not able to analyse the most complex models.

This limitation coupled with the need to create experiments coupled with models that could be used to analyse unknown concentrations and mixtures of antibody motivated a series of new definitions and new ways of thinking about identifiability. In some cases, parameters could not be identified regardless of the way the system was observed, and in other cases, parameters were unidentifiable because of a combination of the system dynamics and observation.

This observation allowed for more straightforward tests for identifiability and motivated new experimental designs coupled with models in which previously unknown parameters were identifiable, even when concentrations of antibody were unknown.

This meant that in the body of the thesis there is an extensive network of definitions which parallel those in the literature, but was motivated by making a distinction between parameters that could be identified if the system were observed differently, and parameters

that could not be.

These new experimental designs involved experiments conducted with the analyte in a well rather than in flow over the chip. The proposed tool for these experiments was the Octet Biolayer interferometry (BLI) system.

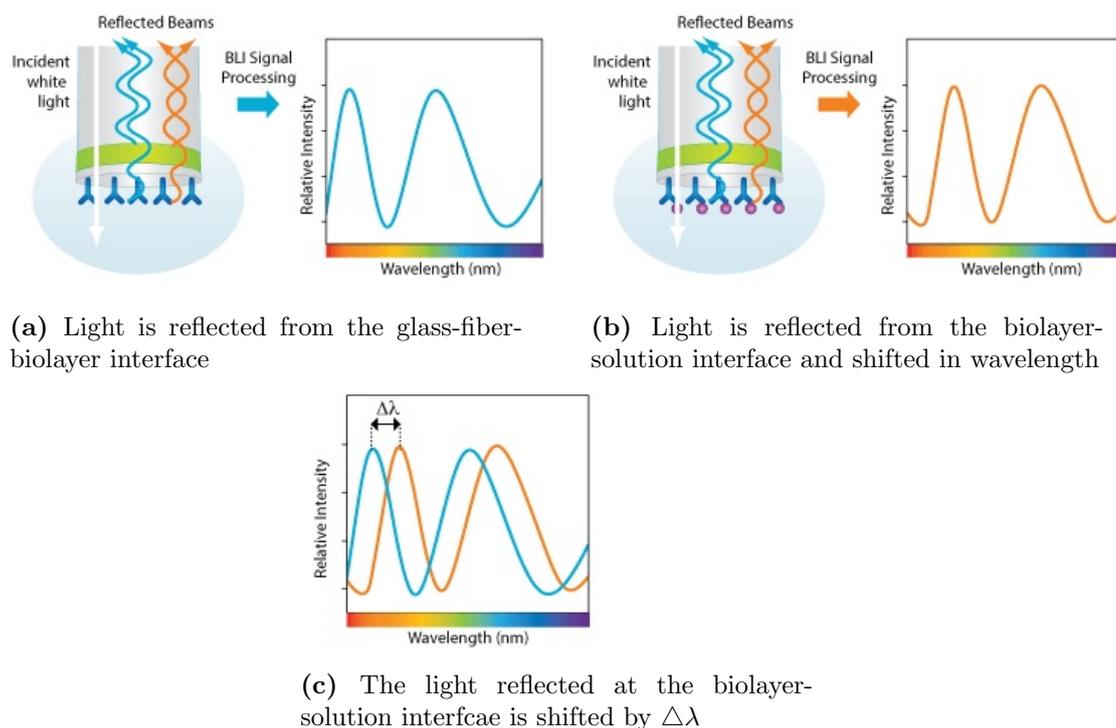
## 1.6 Bio layer interferometry technology

BLI is a technology similar to SPR that allows for label-free measurements of biochemical interactions via light and the changing optical properties of the reaction layer. The key difference in how measurement is conducted is that it uses reflection rather than refraction, although the relevance of the biosensor to this work is that the reactions take place in wells rather than in flow.

White light is sent down the glass fibre biosensor and reflected back at two interfaces, that of the glass fibre and the biolayer, and that of the surface bound reactants on the biolayer and the solution (Fig. 1.19 a). Much like SPR any change in the chemicals bound to the biolayer causes a change in the light escaping the interaction with the surface layer (Fig. 1.19 b). In the case of BLI what is measured is the interference pattern created by the reflections at the two interfaces.

The binding between a ligand immobilised on the biosensor tip surface and an analyte in solution produces an increase in optical thickness at the biosensor tip, shifting the wavelength of the light reflected at the biolayer solution interface (Fig. 1.19 c). This shift changes the relative intensity of the light reflected at each wavelength allowing for a near instantaneous measurement of the change in thickness of the biological layer to be made (in Drug Discovery, 2006; Rich, 2007; Fang, 2007).

The benchmark study Yang *et al.* (2016) used BLI and SPR technology for antibody affinity estimation. However due to the use of the Langmuir model, as discussed in Section



**Figure 1.19:** Illustrations of light reflected in BLI experiments taken from (ForteBio, 2007).

1.4, which assumes incorrectly that IgG antibody would exclusively bind monovalently. estimates of antibody affinity were also not consistent.

This highlights the need for a model that could be used for antibody experiments on BLI. As this experimental platform uses an enclosed chamber where the reaction takes place this model should be different to the equivalent SPR model, which assumes a constant flow of analyte maintaining it at constant concentration (even if not at the surface of the reaction chip). As models for BLI should be different to SPR models may allow for the identification of parameters that are not identifiable for SPR. As a result the technologies may be useful together for determining interaction properties of antibody and antigen, both in a clinical and pharmaceutical setting.

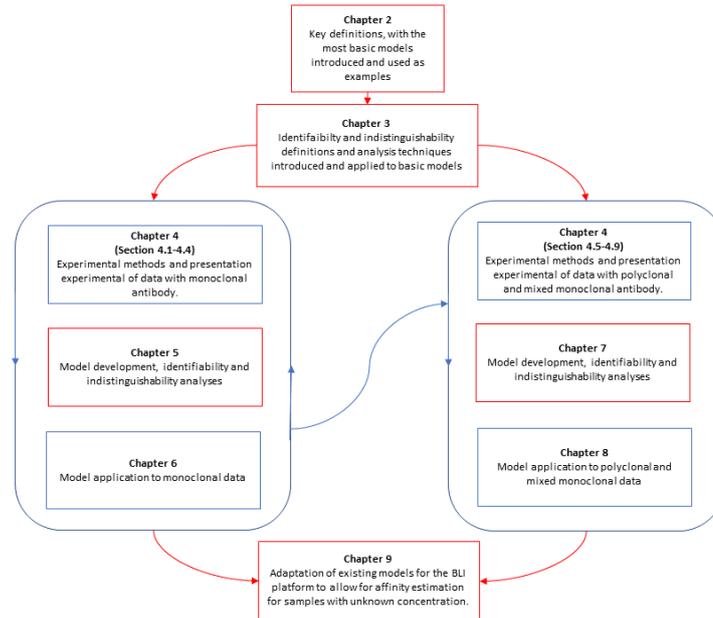
## 1.7 Aims and Objectives

The overarching goal of this thesis is to develop models and fitting procedures that could be used to accurately estimate antibody-antigen affinity. This chapter has centred on how this problem is related from material from a number of disciplines. As a result the objectives subsidiary to this aim span these disciplines.

1. To develop models based on the chemical interaction schemes of both monoclonal and polyclonal antibody against antigen, that include the effects of antibody not being well mixed.
2. To analyse the identifiability and indistinguishability of these models, and assess whether they would be useful for a pharmaceutical or clinical setting.
3. To develop methods sufficient for this analysis, if current methods do not work.
4. To participate in designing and conducting monoclonal and polyclonal antibody-antigen SPR experiments that will allow for these models to be tested against each other and those in the literature.
5. To fit the models to the data, analyse whether they improve goodness of fit, and the consistency of affinity estimates.
6. If antibody samples being in unknown concentration limits the ability of SPR to identify antibody-antigen affinity to investigate alternative experimental methods.

## 1.8 Thesis Overview

The relationships between the content of the chapters are shown in Fig. 1.20. Whilst most material is structured sequentially, there were iterative cycles of experimentation, model



**Figure 1.20:** Visualisation of the relationships between thesis chapters.

development and model fitting for both polyclonal and monoclonal antibody, that create more complex links within two groups of chapters.

Chapter 2 introduces the key mathematical definitions that will be used in the thesis; Chapter 3 introduces the concept of structural identifiability and indistinguishability, as well as algorithms that can be used to test for them, as well as the other key algorithms used in the project. A key element of these existing chapters is the division of the concept of a model into a system and an output - this allows for faster tests for structural identifiability, using the novel concept of simple structural identifiability.

Additionally, new methods and algorithms for structural identifiability analyses of systems with step inputs (such as SPR) are developed.

Chapter 4 explains the experimental techniques used and shows the data collected. As this data is involved in both iterative loops of the project it is split up into two boxes in Fig. 1.20 - one for monoclonal research and one for polyclonal and mixed monoclonal

research.

Again, because of these iterative loops, some of the decisions about experimental procedures were motivated by assumptions of models presented in chapter 5 or 7. To deal with assumptions from pre-existing models, Section 1.5 was written. This means that the key concepts could be referred back to in the introduction. Similarly, to deal with more complex chemical interactions that had not been used in SPR models before this thesis discussion of multi-antibody multi-antigen structures were included in Section 1.3. Together their inclusion in the introduction allows for the thesis to have a linear flow despite including these cycles.

A key result of Chapter 4 is that some of the patent polyclonal samples reacted in ways entirely unlike the monoclonal antibody- showing almost negligible amounts of total dissociation over the dissociation phase of the experiments. This observation motivated most of the polyclonal model development (Chapter 7) which in turn motivated experiments designed to test the new model's assumptions - i.e. that the dynamics emerged from multi-antibody multi-antigen complexes formed by simultaneous the antibody in the mixture binding to separate epitopes and then rebinding to epitopes on the neighbouring antigen to produce larger complexes which were more stable than those that could be produced with single monoclonal antibody. As a result, a third set of experiments is included, experiments made with mixtures of monoclonal antibody, which were used to see if these dynamics could be duplicated.

Chapters 5 and 6, as well as Chapters 7 and 8, are symmetrically paired; that is Chapters 5 and 7 both deal with model development and Chapters 6 and 8 deal with fitting those models to the data obtained in chapter 4. In either case, as data was collected models were developed and fit to the data. Problems with model fitting motivated the development of new models which were also fit to the data, and the cycle continued.

The cycle between experimentation with monoclonal antibody, model development and

fitting in Chapters 4-6 lead to the co-first author publication *direct quantitative measurement of the kinetics of HLA-specific antibody interactions with isolated HLA proteins* (Daga *et al.*, 2017) and its follow up Daga *et al.* (2018) in which the models developed and fit in this thesis were used with sensorgrams from Daga (2015) to determine antibody-antigen affinities were determined with a much lower degree of variance per interaction than those in the literature (see Section 1.4.1). Notably, the accuracy of these affinity estimates was so great that the difference between antibody-antigen affinity for various antigen presenting the same epitope was measured for the first time.

The cycle between experimentation with polyclonal antibody samples, model development and fitting in Chapters 4,7 and 8 lead to the publication H.A.J.Moyse *et al.* (2012) in which antibody-antigen affinities were estimated for polyclonal antibody.

Additionally, this cycle lead to the development and validation of the generalised polyclonal model (GPM) (Chapter 7). As previously mentioned this assumed the mechanism behind the extremely slow dissociation seen in some of the polyclonal samples was due to the formation of multi-antibody multi-antigen complexes. The mixed monoclonal samples were used to test this, and their analysis in Chapter 8 showed that the mixed monoclonals dissociated slower than the unmixed one. Supporting this hypothesis, and yielding one of the key results of the thesis.

Chapter 9 deals with overcoming the limitations of SPR. A key problem with all the models in the literature and developed in this thesis is that they cannot be used to identify the affinity of antibody-antigen interactions where the antibody is in an unknown concentration. BLI is selected as an alternative technology that could solve this problem. Models developed for SPR are redeveloped for BLI, and structural identifiability analyses are conducted. These show that affinity of antibody-antigen interactions where the antibody is in an unknown concentration can be identified.

# Chapter 2

## Theory - Basic Concepts

In this chapter, the more abstract concepts required for this thesis will be introduced. These will lay the foundation for the next chapter which will deal with identifiability and indistinguishability, as well as the model fitting conducted in Chapters 6 and 8.

As mentioned in the introduction one of the ideas that arose from this work is that there is a difference between parameters that are indistinguishable because of the system they are part of and those that are indistinguishable because of the relationship between the system and the way it is observed.

In some cases, definitions will be novel, or definitions are going to be presented with adaptation, so this facet of modelling can be discussed. However, these definitions will be made to reflect the common usage of terms in the literature.

### 2.1 Concepts from statistics

To be able to fit mathematical models to data, and estimate their parameters, and thus the physical constants and rates these parameters represent, some well known definitions from statistics are required.

**Definition 2.1.** *The residual of an experimentally observed value  $\mathbf{Y}_k$ , given a model predicting the value taken  $\mathbf{y}(t_k)$  is the difference between the observed value and the value*

the model predicted:  $\mathbf{r}_k = \mathbf{Y}_k - \mathbf{y}(\tau_k)$ , and  $\tau_k$  represents the time of the  $k^{\text{th}}$  observation.

In the above definition, the notation used reflects the fact that the experimental equipment only takes measurements at discrete time intervals, whereas the models used in this thesis will be continuous in time.

We measure the goodness of fit of a model with residual sum of squares (RSS) (Draper, 1998).

**Definition 2.2.** A *residual sum of squares* is the sum of the squares of the residuals of a model for an experiment

$$\text{RSS} = \sum_{i=1}^q \sum_{t \in T} \left( \frac{r_{it}}{\sigma_i} \right)^2 \quad (2.1)$$

in which  $q$  is the dimension of the  $\mathbf{y}(t)$ ,  $T$  is a set containing all time points where a measurement was taken,  $r_{it}$  is the  $i^{\text{th}}$  element of the vector of residuals at time  $t$ , and  $\sigma_i$  is the standard error of the  $i^{\text{th}}$  component of the data.

RSS is a useful tool as it provides a numerical value that quantifies the relative goodness of fit of a model to data.

It is also useful to use this concept alongside that of the distribution of residuals. When a fit is made, plots of residuals allow the user to review whether a model fits one part of the data well and another badly and if the residuals follow a normal distribution.

The distribution of residuals is significant because in the modelling process we assume the data comes from a mathematically describable system but has some measurement error. If we can find a model for the data that minimises the RSS we want that to be because it reduces the residuals to the point by point measurement error. As a result, its residuals will follow the same distribution as the errors in measurement.

We assume that measurement errors at each time point are independent samples from the

same distribution, which we expect to be normal due to the central limit theorem.

The models used in this thesis will vary greatly in their complexity. As a result, it is useful to have a tool for model selection that weighs goodness of fit against the complexity of a model. The Akaike information criterion (AIC) is such a tool (Akaike, 1974).

**Definition 2.3.** *Akaike's information criterion (AIC) is given by*

$$AIC = n \log \left( \frac{\text{RSS}}{n} \right) + 2k \quad (2.2)$$

where  $k$  is the number of model parameters,  $n$  is the number of measurements, and the residuals are independent and identically distributed random variables taken from a normal distribution.

## 2.2 State-space models

The goal of the mathematical modelling presented in this thesis is to estimate parameters representing the physical properties of an experimental system, but before this can be discussed adequately, it is necessary to develop terminology for discussing these ideas.

These definitions parallel those in Sontag (1998), although emphasis will be put on a set of mappings associated with the model, its input, states and output that will allow us to distinguish between parameters that are unidentifiable because of the system, or because of the outputs of the system. To do this, They allow us to extend the discussion of input-output relationships already existing in the literature (Evans *et al.*, 2013; Bellu *et al.*, 2007), to that of input-state-output mappings.

**Definition 2.4.** *A system or underlying system  $\Pi$  consists of:*

- A parameter vector  $p \in \Omega \subset \mathbb{R}_{>0}^n$  for some  $n \in \mathbb{N}$ , where  $\Omega$  is the set of

*meaningful parameter vectors.*

- An **input function**  $\mathbf{I} : \mathcal{T} \times \Omega \rightarrow \mathcal{U}, \mathbf{I} : (t, \mathbf{p}) \mapsto \mathbf{I}(t, \mathbf{p})$ , where  $\mathcal{T}$  is an interval on the real line representing the time period the input function is defined for,  $\mathcal{U} \subset \mathbb{R}^q$  is the **input-value space of the system**,  $q \in \mathbb{N}$ , and  $\mathbf{I} \in \mathcal{J}_{\Pi}$ , where  $\mathcal{J}_{\Pi}$  is the **set of admissible input functions**, which is a subset of the set of piecewise-smooth real vector-valued functions.
- A **state variable**  $\mathbf{x} : \mathcal{T} \times \Omega \times \mathcal{J} \rightarrow \mathcal{X} \subset \mathbb{R}_{\geq 0}^m : (t, \mathbf{p}, \mathbf{I}) \mapsto \mathbf{x}(t; \mathbf{p}, \mathbf{I})$  where  $\Omega$  is a subset of the set of real valued vectors, and  $\mathcal{X} \subset \mathbb{R}_{\geq 0}^m$  for some  $m \in \mathbb{N}$ , is the **state-space**.
- A **state-space equation**, of the form

$$\frac{d}{dt}\mathbf{x}(t, \mathbf{p}) = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I}(t, \mathbf{p})), \quad (2.3)$$

where  $\mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I}(t, \mathbf{p}))$  is a vector of rational functions of the components of both  $\mathbf{x}(t, \mathbf{p})$  and  $\mathbf{I}(t, \mathbf{p})$ , such that  $\mathbf{x}(t; \mathbf{p}, \mathbf{I})$  is always piecewise continuous and smooth in time whenever  $\mathbf{I}(t, \mathbf{p})$  is.

- An **initial condition**  $\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p})$  (Compare Sontag 1998; Def 2.1.3).

In some contexts  $\mathbf{x}(t; \mathbf{p}, \mathbf{I})$  will be abbreviated to  $\mathbf{x}(t, \mathbf{p})$  or  $\mathbf{x}(t)$ . It is often useful to distinguish between the value taken by  $\mathbf{x}(t; \mathbf{p}, \mathbf{I})$  at a time  $t$  and the function itself. As a result  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I})$  will be used to represent the function, in some cases this will be abbreviated to  $\mathbf{x}(\cdot; \mathbf{p})$ .

This definition of a system contains most of the concepts we would associate with models. In fact, we will associate each model, in this thesis with an underlying system. As a result, this definition is going to be used later to form part of the definition of a model, but splitting this part of the abstract definition of a model is also going to be useful for generating a new conceptual framework for discussing models and doing mathematics

related to modelling.

This thesis contains outputs from experimental systems, and mathematical conceptualizations of these experimental systems, that will be used to make predictions about how the internal states of the system will behave depending on the physical and chemical parameters of the system and the way the user controls it.

Notably, due to the design of the equipment, the users control of the system is very limited. For SPR at the start of the experiment, the times at which the contents of different wells are fed through the inlet tube across the reaction chip and ligand are preset. In the case of the experiments we are interested in where there is only one analyte, either a mixture of patient antibodies or a monoclonal antibody, the input function will be a vector valued step function. In the association phase, the input function will be a vector of the dilutions of the sample in each well; in the dissociation phase, it will be a vector of the corresponding length of zeros, representing the sample being replaced with buffer (see Fig. 1.10).

In the manufacturer's data fitting software, the models for BLI use the same input functions as SPR models (Tobias & Kumaraswamy, 2013) and these models are used throughout the literature (Abdiche *et al.*, 2008; Sun *et al.*, 2015). In this case, the input function represents the concentration of the sample present as unbound analyte in the well at the start of an experimental phase

As we are using these systems to make predictions, about how experiments will behave with biochemicals with different properties, it is useful to make some novel but intuitive definitions that will later be able to use to discuss the relationship these predictions have with each other.

**Definition 2.5.** *Given a system  $\Pi$  as defined above, with the rates of change:*

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{p}, \mathbf{I}), \mathbf{p}, \mathbf{I}(t))$$

and the initial conditions  $\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p})$ :

1. The solution  $\mathbf{x}(\cdot; \mathbf{p})$  to this initial value problem is called the **state-space solution**.
2. We define the **set of state-space solutions** for this system as:

$$\mathcal{F}_\Pi = \{\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}(\cdot; \mathbf{p})) : \mathbf{p} \in \Omega, \mathbf{I} \in \mathcal{J}_\Pi\}. \quad (2.4)$$

3. We define an **input-state mapping** for  $\Pi$  as  $X_\Pi : \Omega \times \mathcal{J} \rightarrow \mathcal{F}_\Pi : (\mathbf{p}, \mathbf{I}(\cdot; \mathbf{p})) \mapsto \mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}(\cdot; \mathbf{p}))$ .
4. We define an **parameter-state mapping** for  $\Pi$  given the input  $\mathbf{I}(\cdot; \mathbf{p})$ , as  $X_{\text{III}} : \Omega \rightarrow \mathcal{F}_\Pi(\mathbf{I}) \subset \mathcal{F}_\Pi : \mathbf{p} \mapsto \mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}(\cdot; \mathbf{p}))$ .

That is, each system is associated with an input-state mapping  $X_\Pi$ , which establishes the relationship between its parameters, input function and the curve through time and parameter space made by its state variables. In a later section, these simple concepts will be used to develop faster computational methods for analysing the identifiability and indistinguishability of models as well as establishing when parameters are unidentifiable because of the underlying system, rather than the relationship between it and the output. As a result, we will be able to investigate whether or not parameter identification may require different kinds of changes to the experimental system.

A useful part of this definition is  $X_{\text{III}}$ , because the experimentalist interacting will generally know the input - what samples were put in, or when certain mechanism were activated, but not the parameters that describe the system, so this will allow us to consider them separately.

We will now define a model in relation to our earlier definition of its underlying system.

**Definition 2.6.** A *model*  $\Sigma$  consists of:

- A system  $\Pi$  as defined above, on the time domain  $(0, t_1) \subset \mathbb{R}$
- An **output variable**,  $\mathbf{y}(t) \in \mathcal{Y} \subset \mathbb{R}_{\geq 0}^n$ , where  $\mathcal{Y}$  is the **output variable space**,
- An **output equation** of the form  $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p})$  where  $\mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p})$  is a smooth function, termed the **output mapping** (Compare Sontag 1998; Def 2.1.3).

We will sometimes consider an experiment on a single well of a BLI machine or on a single interaction spot of an SPR machine. As a result a special case of the above will be important. A model  $\Sigma$  is **unitary** if its output variable  $y(t) \in \mathcal{Y} \subset \mathbb{R}_{\geq 0}$ , is a scalar, and in which the underlying system, the input function  $I : \mathbb{R}^{n+1} \rightarrow \mathbb{R}, I : (t, \mathbf{p}) \mapsto \mathbf{I}(t, \mathbf{p})$  is a scalar valued function.

In this thesis, the goal will not just be to create models with similar input-output behaviour to the physical experiments, but to that combine this with state-space behaviour that matches our theoretical understanding of the physical and biochemical processes involved. As a result, a lot of this thesis will consist of the analysis of the state-space behaviour of these models.

Just as we have formalised the concept of a model and a system, it is useful to make novel definitions of sets and functions relating to the model which will be useful in theorems presented later.

**Definition 2.7.** *Given a model as defined above, with the state variables  $\mathbf{x}(t, \mathbf{p})$ , in  $n$  dimensions, with the rates of change,  $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{I}(t), \mathbf{p})$  and the output  $\mathbf{y} = \mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p})$  and the initial conditions  $\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p})$ , and state-space solution  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I})$ , where  $\mathbf{p}$  is the parameter vector:*

1. The **output solution** is defined as the mapping  $y(\cdot; \mathbf{p}, \mathbf{I})$  defined at each point in time by  $y(t, \mathbf{p}, \mathbf{I},) = g(\mathbf{x}(t, \mathbf{p}, \mathbf{I}))$ .
2. We define the **set of output solutions** for this system as  $\mathcal{G}_\Sigma = \{y(\cdot; \mathbf{p}, \mathbf{I}) : \mathbf{p} \in$

$\Omega, \mathbf{I} \in \mathcal{J}_\pi\}$ .

3. We define the **state-output mapping** for  $\Sigma$  as  $Y_\Sigma : \mathcal{F}_\Pi \rightarrow \mathcal{G}_\Sigma : \mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}) \mapsto \mathbf{y}(\cdot; \mathbf{p}, \mathbf{I})$ .
4. We may define a **input-output mapping** for  $\Sigma$ , as  $Z_\Sigma : \Omega \times \mathcal{J} \rightarrow \mathcal{G}_\Sigma : (\mathbf{p}, \mathbf{I}(\mathbf{p}, \cdot)) \mapsto \mathbf{y}(\cdot; \mathbf{p}, \mathbf{I})$ .
5. We may define a **parameter-output mapping** for  $\Sigma$ , as  $Z_{\Sigma I} : \Omega \rightarrow \mathcal{G}_{\Sigma(\mathbf{I})} \subset \mathcal{G}_\Sigma : \mathbf{p} \mapsto \mathbf{y}(\cdot; \mathbf{p}, \mathbf{I})$ .

It may be observed that the input-output mapping  $Z_\Sigma$  is the concatenation of the state-output mapping and input-state mapping,  $Y_\Sigma$  and  $X_\Pi$ . It is useful to break down the input-output mapping for a model into these two separate mappings because later we will be able to propose tests using these functions, for the identifiability of models.

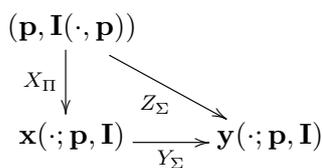
The relations between these three maps is illustrated in Fig. 2.1(a).

Just as the parameter-state mapping,  $X_{\Pi\mathbf{I}}$ , was defined because the input function is generally known, we also define  $Z_{\Pi\mathbf{I}}$ . The relationship between these functions is shown in Fig. 2.1(b).

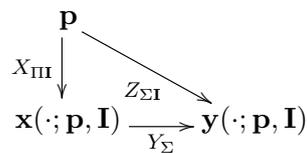
It should be noted that the mapping  $Y_\Sigma$  is quite different from the mapping in the output equation  $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p})$ ;  $Y_\Sigma$  maps from a domain containing vector valued functions, to a codomain containing vector valued functions, whereas  $\mathbf{g}(\mathbf{x}, \mathbf{p})$  is a function mapping from a domain of vectors to a codomain of vectors.

**Remark 2.1.** For a model  $\Sigma$  with an output equation  $\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t, \mathbf{p}), \mathbf{p})$  and a state-output mapping  $Y_\Sigma$ , if  $\mathbf{g}(\cdot, \mathbf{p})$  is bijective  $Y_\Sigma$  also is; but  $Y_\Sigma$  may be bijective even if  $\mathbf{g}(\cdot, \mathbf{p})$  is not.

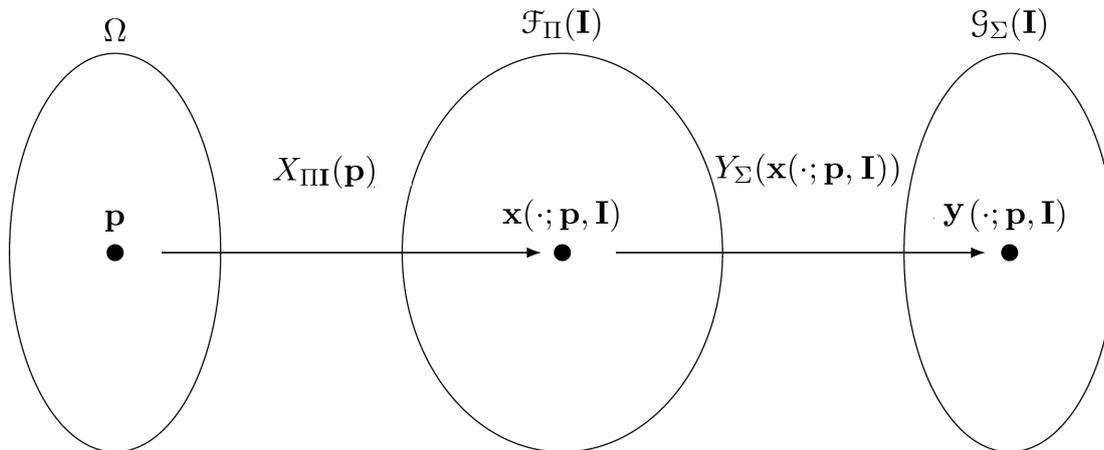
The first case of the above emerges from Def. 2.7. If  $\mathbf{g}(\cdot, \mathbf{p})$  associates each point in state-space with a unique point in output-space by  $\mathbf{g}(\cdot, \mathbf{p})$  then  $Y_\Sigma$  associates each unique



(a) An illustration of the elements being mapped by the input-state, state-output and input-output mappings.



(b) An illustration of the elements being mapped by the parameter-state, state-output and parameter-output mappings, for a specific input  $\mathbf{I}(\cdot, \mathbf{p})$ .



(c) Diagram of domains and images of parameter-state, state-output and parameter-output mappings, for a specific input  $\mathbf{I}(\cdot, \mathbf{p})$ .

**Figure 2.1:** Diagrams of relationships between parameter-state, state-output and parameter-output mappings as well as their domains and images.

state space solution to a unique output solution. A model in which  $\mathbf{g}(\cdot, \mathbf{p})$  is not bijective and  $Y_{\Sigma}$  is, is the Bivalent Langmuir model, which will be introduced in Section 5.3.4 and analysed in Section 5.1.4.

Models in this thesis have a purpose - that is to create input/output relationships similar to those observed with experiments, so chemical properties like association and dissociation rates can be estimated. To do this, computational integration methods are used. This means that the models used in this thesis, and in particular their state-space and output solutions are going to have certain well-known features. They will be both continuous and

piecewise smooth in time (as guaranteed by Definition 2.4, for step function inputs.)

It is useful to consider some brief examples to explain why the unusual choice of defining both model and its underlying system has been made, and the various mappings have been defined in such formality. In a large number of problems, to take two examples outside of this thesis, in the fields of ecology and pharmacokinetics, it is possible to obtain many different kinds of output, for example measuring the amount of a kind of drug or the various signalling molecules that may be stimulated by it, at different points in the test subject's body, or it may be possible to measure the population of a number of different species interacting in one environment. However, it may be impossible to change the underlying system of the test subject's body or the underlying ecological system.

As a result in many cases it is important to distinguish between properties of a model that arise from the underlying system, its output mapping or a combination of the two. The language of the underlying system and the various input-state-output mappings will allow for a discussion of this and make appropriate choices of experimental equipment.

Notably, in the previous two definitions, two concepts have not been explored: that of a state-space solution  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I})$  and an output solution,  $\mathbf{y}(\cdot; \mathbf{p}, \mathbf{I})$ . Although the output solution is referred to here as a solution, it is not indicated here what it is the solution to. Unfortunately the mathematics required to derive the initial value problem (IVP) to which it is a solution requires more definition and explanation, which means it will come much later in this chapter. For now, it is sufficient to think of both of these functions as the solutions to initial value problems, in the case of  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I})$ , the IVP of the state-space equations and the systems initial condition  $\mathbf{x}(0, \mathbf{p}) = \mathbf{x}_0(\mathbf{p})$ , and for the output variable, some unknown system of equations and  $\mathbf{y}(0, \mathbf{p}) = \mathbf{g}(\mathbf{x}_0(\mathbf{p}))$ .

The models we will be concerned with all fit the definition previously made. However, because of the design of the experimental equipment, it is useful to also consider a special case of this definition.

**Definition 2.8.** An *autonomous model*  $\Sigma$  is a model where the input-value space of the system,  $\mathcal{U}$ , contains only a single member. (Compare Sontag 1998; Def 2.1.8)

This is equivalent to the set of admissible input functions  $\mathcal{J}_{\Pi}$  containing only a single member, which itself is a constant function.

Sontag (1998) notes that this kind of model may be thought of as a dynamical system with outputs.

We can think of SPR or BLI experiments as being a series of coupled dynamical systems, where the input is constant in each phase and changes instantaneously between the phases. This makes sense in the context of this definition because once the experiment is begun the user has no controls, that is the user cannot affect the input function. In an SPR experiment the user cannot change the dilution of the analyte being fed into the flow from the sample wells, so it remains constant; when conducting a BLI experiment, the user cannot change the concentration of the sample in the well that the reaction takes place in, so it too remains constant. As a result in either case, in the association phase, the input-value space of the system,  $\mathcal{U}$  contains only a single vector of constants, representing the original dilutions of the sample; and in the dissociation phase it  $\mathcal{U}$  contains only a single vector of zeros.

**Remark 2.2.** Consider a generic model  $\Sigma$  a step valued input function  $\mathbf{I}(t, \mathbf{p})$  taking the values  $\mathbf{I}_a$  and  $\mathbf{I}_d$  in the time periods  $T_a = [t_0, t_1]$  and  $T_d = (t_1, t_2]$  which are termed the association and dissociation phase and with the state-space equation  $\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I}(t, \mathbf{p}))$ .

This model can effectively be split into two autonomous models the **association model**  $\Sigma_a$  and the **dissociation model**  $\Sigma_d$  - one model for either phase. These two new models inherit from  $\Sigma$  Its parameter vector, state space, state variable, output variable, output variable space and output equation. For the sake of legibility, we write the state variables

of these two models as  $\mathbf{x}_a$  and  $\mathbf{x}_d$ .

These new autonomous models have the state-space equations:  $\frac{d}{dt}\mathbf{x}_a = \mathbf{f}_a(\mathbf{x}_a(t, \mathbf{p}), \mathbf{p})$  and  $\frac{d}{dt}\mathbf{x}_d = \mathbf{f}_d(\mathbf{x}_d(t, \mathbf{p}), \mathbf{p})$  respectively; where  $\mathbf{f}_a(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I}_a)$  and  $\mathbf{f}_d(\mathbf{x}(t, \mathbf{p}), \mathbf{p}) = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I}_d)$ .

$\Sigma_a$  also inherits  $\Sigma$ 's initial condition, whereas  $\Sigma_d$  has as its initial condition  $\mathbf{x}_d(t_1, \mathbf{p}) = \mathbf{x}_a(t_1, \mathbf{p})$ .

As a number of convoluted definitions have been made, it is useful to illustrate them with an example from the literature.

**Example 2.1.** The **single configuration Langmuir model** is the classic model for SPR experiments, and is used predominantly in the literature, as discussed in Subsections 1.4.1 1.4.2. It was the starting point for the development in Section 1.5, and as a result is the simplest of the models we will review.

Unlike more complex models, it assumes the concentration of the analyte in the sample is the same as the concentration at the reaction chip's surface, where it can bind to the ligand, and because it assumes a single binding reaction takes place.

In keeping with the definitions we will first define the underlying system within the model, and then combine that with its output equation to give the model as a whole. This will then be used to illustrate the various concepts that have been defined so far, and later can be used to illustrate the concepts to be introduced later.

The underlying system is defined as:

$$\Pi = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx}{dt} = k_a I (R - x) - k_d x, \\ x(0; \mathbf{p}, I) = 0; \end{cases} \quad (2.5)$$

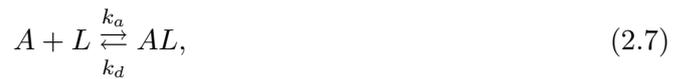
where the parameter vector is defined as  $\mathbf{p} = (k_a, k_d, C_T, R)$ ,  $k_a$  and  $k_d$  are the association rate constant and the dissociation rate constant respectively,  $C_T$  is the concentration of analyte flowing over the sensor chip and  $R$  is the maximum density of analyte that can attach to the ligand on the sensor chip; the state variable representing analyte bound to ligand on the biosensors surface is  $x(t; \mathbf{p}, I) \in [0, R]$ ;  $I(t)$  is the input function representing concentration of analyte at the machines inlet and  $T_a$  and  $T_d$  are intervals on the real line representing the association and dissociation phase, defined as  $T_a = [t_0, t_1]$ ,  $T_d = [t_1, t_2]$  respectively.

The model is defined as:

$$\Sigma = \begin{cases} \Pi \\ y(t; \mathbf{p}, I) = \alpha x(t; \mathbf{p}, I) \end{cases} \quad (2.6)$$

Where the underlying system  $\Pi$  is as previously defined and  $\alpha$  is a known constant.

The state-space equation of the model comes originally from Langmuir. (1916), and the model is named after the papers author. It is derived from the reaction equation:



where  $A$ ,  $L$  and  $AL$ , represent analyte and ligand monomers, and an analyte-ligand dimer

Parameter	Unit in Marquart (2013)	Unit used here
$x$ and $R$	RU	$\text{pg nm}^{-2}$
$I$ and $C_T$	nM	nM
$k_a$	$\text{M}^{-1}\text{s}^{-1}$	$\text{M}^{-1}\text{s}^{-1}$
$k_d$	$\text{s}^{-1}$	$\text{s}^{-1}$
$\alpha$	-	$\text{RU pg}^{-1} \text{nm}^2$

**Table 2.1:** Comparison between units used in Marquart (2013) and this thesis.

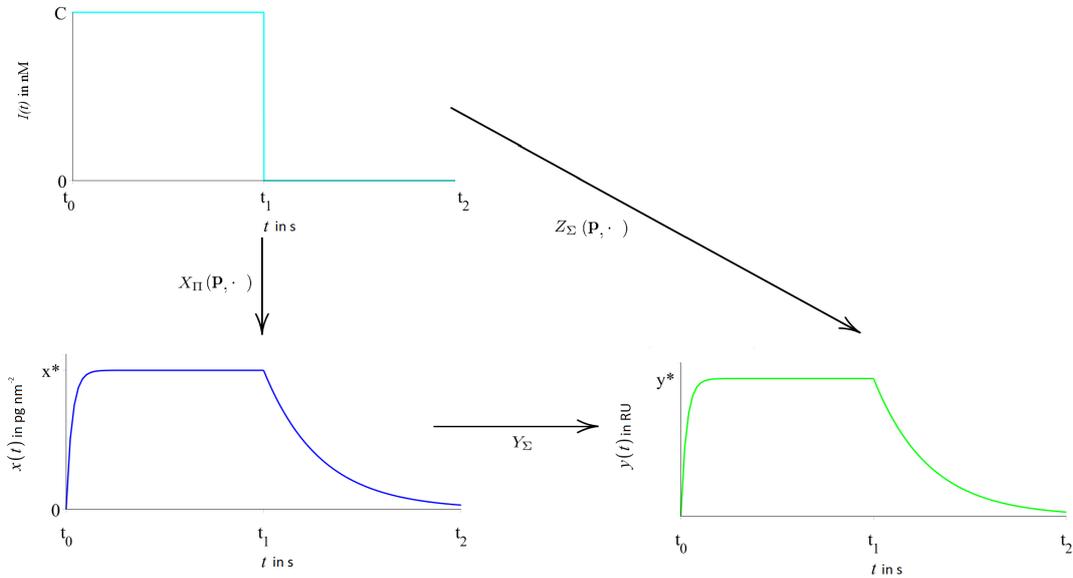
respectively.

In Eq. 2.5 the concentration of the analyte,  $A$ , is the input function  $I$ . The concentration of ligand  $L$  is total ligand concentration minus bound ligand concentration,  $(R - x)$ . Their product multiplied by the association rate constant,  $k_a$  is the rate at which new dimers are formed. The concentration of bound ligand  $AL$  is  $x$ , and the rate at which it unbinds is  $k_d x$ .

As with parameters and variables in general, the units chosen to measure them in may vary depending on their use. The units of the parameters and variables in this thesis will generally be the same as those used in Marquart (2013). They are shown in Table 2.1, where RU are a unit of refraction angle change commonly used in SPR experiments (see 1.4 ).

The only difference is that Marquart (2013) uses a simpler version of the model in which the state variable and output variable are the same. In this thesis they are related by Eq. 2.6 where  $\alpha = 1\text{RU pg}^{-1} \text{nm}^2$  (Richard B. M. Schasfoort, 2008). This minor change was made so that every model presented in this thesis will have its internal states measured in consistent SI units.

Because this underlying system and its input is relatively simple, the state-space solution



**Figure 2.2:** The input, output and state-space solutions of the Langmuir model for a given input  $I$  and parameter vector  $\mathbf{p}$ , as well as the mappings between them.

evaluated at a general time-point  $t$  found, and is given by:

$$\mathbf{x}(t, \mathbf{p}, I) = \begin{cases} \frac{k_a C_T R}{C_T k_a + k_d} (1 - e^{-(C_T k_a + k_d)t}) & : t \in T_a \\ x(t_1; \mathbf{p}) e^{-k_d(t-t_1)} & : t \in T_d. \end{cases} \quad (2.8)$$

It may be observed from the above that the state-variable converges asymptotically towards a fixed point in the association phase. We call this  $x^*$  where  $x^* = \frac{k_a C_T R}{C_T k_a + k_d}$ .

The input, internal states and the output of the model are visualised in Fig. 2.2, and the parameters used for these visualisations are shown in Table 2.2, and these graphs are used to illustrate the relationships between the functions shown in Fig. 2.1(a).

There are three objects which will be mapped from and onto, the pair of the parameter vector and input  $(\mathbf{p}, I)$ ; the state-space solution,  $x(\cdot; \mathbf{p}, I)$ ; and the output solution,  $y(\cdot; \mathbf{p}, I)$ .

Parameter	$k_a$	$k_d$	$R$	$C_T$	$t_1$	$t_2$
Value	$6\text{E-}4\text{M}^{-1}\text{s}^{-1}$	$7\text{E-}3\text{s}^{-1}$	$500\text{pg nm}^{-2}$	$100\text{nM}$	$500\text{s}$	$1000\text{s}$
	Steady state	$x^*$	$y^*$			
	Value	$448\text{ pg nm}^{-2}$	$448\text{ RU}$			

**Table 2.2:** Values taken by parameters and steady states in Fig. 2.2

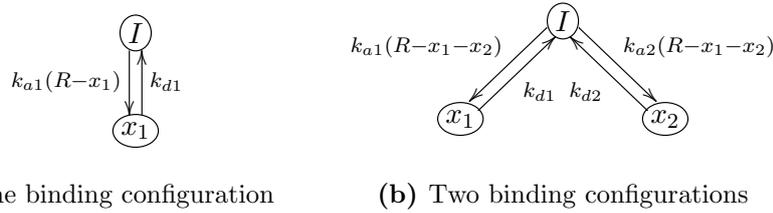
The input-state mapping  $X_{\Pi}$  takes the parameter vector and input pair,  $(\mathbf{p}, I)$  and maps it onto the state-space solution  $x(\cdot; \mathbf{p}, I)$ . The state-output mapping  $Y_{\Sigma}$  takes state-space solution  $x(\cdot; \mathbf{p}, I)$  and maps it to the output solution  $y(\cdot; \mathbf{p}, I)$ . The composition of these maps is  $Z_{\Sigma}$ .

This model would be somewhat appropriate for a solution containing a single protein as the analyte that binds at a single point to the ligand protein, for example a separated antibody paratope and an antigen that it only binds to a single epitope on. However there is no way of guaranteeing the solution has the same concentration at the reaction layer as it has in the sample so Chaiken *et al.* (1992) note that parameter estimates using it can exhibit systematic errors for the kinetic rate constants.

**Example 2.2.** It is useful to have a more complicated example so both examples can be used to illustrate the techniques that will be discussed in Chapter 3.

The **two configuration Langmuir model** is a simple variation of the previous model that is commonly available in the software used to model SPR experiments and widely used (Bio-Rad, 2010; Evans *et al.*, 2013). It is a little more complex as it has two state variables, and its output mapping is non-bijective, as a result, it will be useful as an example for illustrating techniques for model analysis that will be introduced later.

Like the single configuration Langmuir model, this model assumes that the concentration of analyte that is fed into the machine is identical to the concentration of the analyte in the vicinity of the interaction spots, and thus able to bind (see Fig. 1.18 (a)). Unlike the single configuration Langmuir model, it assumes the interaction of an analyte and ligand



**Figure 2.3:** Compartmental diagrams of the one and two configuration Langmuir model with analyte concentration  $I$  and concentrations of bound analyte  $x_1(t)$  and  $x_2(t)$ .

can form two different configurations of dimer. This is illustrated by its reaction equation:



where  $A$ ,  $L$ ,  $AL_1$ ,  $AL_2$ , represent analyte and ligand monomers, and analyte-ligand dimers in either configuration respectively.

Notably, this is not a model specifically developed for antibody-antigen interactions, so although two species of analyte-ligand products come from the reaction of analyte and ligand they are both dimers, neither represent a trimer, an antibody bivalently bound to two antigens.

A diagram for both forms of the Langmuir model are shown in Fig. 2.3. These provide a visual way to distinguish between complex models. In the case shown the input and state variables are the compartments, these are shown in circles, arrows between these circles show the flow between these compartments. In later chapters the compartments will also include more complex functions of the variables.

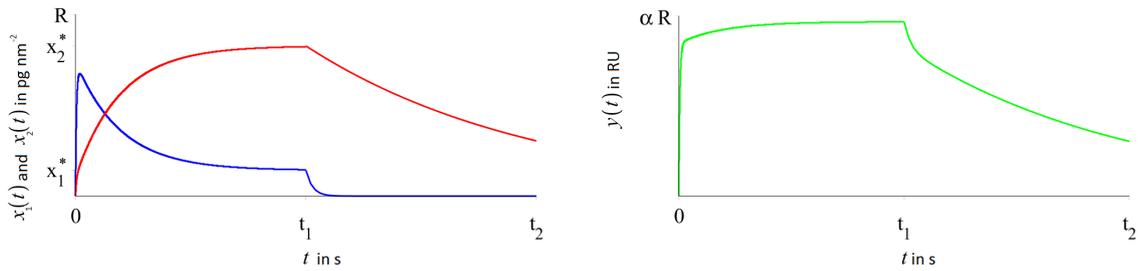
The two configuration Langmuir model is defined as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = k_{a1}I(R - x_1 - x_2) - k_{d1}x_1 \\ \frac{dx_2}{dt} = k_{a2}I(R - x_1 - x_2) - k_{d2}x_2 \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, p) = \alpha (x_1(t, \mathbf{p}) + x_2(t, \mathbf{p})), \end{cases} \quad (2.10)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R)$ . The parameters  $k_{a1}$  and  $k_{a2}$ ;  $k_{d1}$  and  $k_{d2}$ ; are the kinetic rate constants for association and dissociation of the two binding species respectively measured in  $\text{M}^{-1}\text{s}^{-1}$  and  $\text{s}^{-1}$  respectively,  $C_T$  is the concentration of analyte flowing over the sensor chip in nM and  $R$  is the maximum density of analyte that can attach to ligand on the sensor chip in  $\text{pg nm}^{-2}$ . The state variable  $(x_1, x_2)^\top = \mathbf{x}(t; \mathbf{p}, I) \in \mathcal{X}$ , represents the concentrations of two configurations of analyte-ligand dimers in  $\text{pg nm}^{-2}$  and the state-space is defined as  $\mathcal{X} = [0, R] \times [0, R]$ . The input function  $I(t)$  represents concentration of analyte at the machine's inlet and  $T_a$  and  $T_d$  are intervals on the real line representing the association and dissociation phase, defined as  $T_a = [0, t_1], T_d = [t_1, t_2]$  respectively, and  $\alpha$  is a known constant  $1 \text{ RU pg}^{-1} \text{ nm}^2$ .

Like the previous model, it may be observed from the above that the state variable converges asymptotically towards a fixed point in the association phase. Due to the complexity of the fixed point it is not written out. Maple code for it's derivation as well as the general solution of the model, is show in Appendix A.1.1.

In general, the fixed points will not be written out for this reason. Generally, they will be written as  $\mathbf{x}^* \in \mathbb{R}^n$ , where  $\mathbf{x}^* = (x_1^*, x_2^*, \dots, x_n^*)^\top$ , which represent the values the state



(a) Graph of the state-space solution,  $\mathbf{x}(t; \mathbf{p}, I)$ , against time  $t$ ; with association phase fixed point  $x^*$

(b) Graph of the state-space solution,  $y(t; \mathbf{p}, I)$ , against time  $t$ ; with association phase fixed point  $y^*$

**Figure 2.4:** The output and state-space solution curves of the Langmuir model for a given input  $I$  and parameter vector  $\mathbf{p}$

Parameter	$k_{a1}$	$k_{d1}$	$k_{a2}$	$k_{d2}$	$R$	$C_T$	$t_1$	$t_2$
Value	$3\text{E-}3\text{M}^{-1}\text{s}^{-1}$	$7\text{E-}2\text{s}^{-1}$	$5\text{E-}3\text{M}^{-4}\text{s}^{-1}$	$2\text{E-}3\text{s}^{-1}$	$500\text{pg nm}^{-2}$	100nM	500s	1000s

**Table 2.3:** Values taken by parameters in Fig. 2.4

variables take at the steady state.

The internal states and the output of the model are visualised in Fig. 2.4 for the same input function shown in Fig. 2.2.

This model would be somewhat appropriate for a solution containing a single protein as the analyte that binds at two points on the ligand protein. One example of this would be, a separated antibody paratope and an antigen that it binds to two epitopes on. Similar to the single concentration Langmuir model, it assumes the solution containing the analyte is well mixed, and there is no way of guaranteeing the solution has the same concentration at the reaction layer as it has in the sample. As a result parameter estimates made with it may also exhibit the systematic errors for the kinetic rate constants pointed out by Chaiken *et al.* (1992).

## 2.3 Dynamical systems with inputs, and concepts from calculus

The models we will build and compare are effectively dynamical systems coupled to input and output functions, so it is useful to use some of the language and basic concepts from dynamical systems theory to discuss and analyse them. As the inputs for the models in this thesis are piecewise constant, these definitions are analogous to those in the field of switched systems (Z.Sun & M.Sarabia, 2011), but so they are easily applicable to the models in this thesis, they will be given explicitly in terms of input functions.

Because the models we are concerned with represent similar physical systems, they all have similar properties. For example, all systems will be initiated at  $t = 0$  and with all state variables taking the value of 0, and the values of their state variables will increase until either the end of the phase or they reach the steady state of that phase. In the next phase, the system will evolve starting with the state variables taking the values they had at the end of the previous phase. Generally, the values taken by the state variables will decrease until they all reach 0 again.

Because of these similarities, it is only necessary to use a small vocabulary of terms when describing these systems.

**Definition 2.9.** *Given a model with the state variable  $\mathcal{X} \subset \mathbb{R}_{>0}^n$ , with the rates of change,  $\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I})$  a **nullcline** at the time  $t$  is a locus of points where  $0 = f_i(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I})$  for some  $i \in \mathbb{N}$  where  $i \leq n$  (Sternberg 2011 Subsection 11.1.1).*

We use the idea of a nullcline, because when our input changes, we expect the nullclines also to change.

In many cases, the formula for the null-cline of a variable is too long to print. In this

thesis, the value taken by a variable in steady state will be denoted by the parameter with an asterisk next to it, for example, a steady state of  $x_1$  will be denoted  $x_1^*$ .

In some cases, the nullclines of all variables will intersect. As a result, it is necessary to define a steady state.

**Definition 2.10.** *Given a model with the state variable  $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}_{\geq 0}^n$ , with the rates of change,  $\frac{dx}{dt} = \mathbf{f}(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I})$  a **steady state** is a point in state space where  $0 = f_i(\mathbf{x}(t, \mathbf{p}), \mathbf{p}, \mathbf{I})$  for  $\forall i \in \mathbb{N} : i = 1, \dots, n$  (Sternberg 2011 Subsection 11.1.1).*

A similar concept that will be used in this thesis is that of a quasi-steady-state (QSS). In some of the systems, we will be concerned with the rate of change of one variable will be orders of magnitude greater than the rate of change of another variable. As a result, the system will quickly move onto the nullcline of the faster-changing variable, where it will stay.

As a result the algebraic equation given by the nullcline can be used to remove one of the variables. This process generates a new system with fewer variables, potentially it may also make numerical integration simpler by eliminating a variable that needs to be considered on a different time scale to the others (compare Schauer & Heinrich 1983).

**Definition 2.11.** *If  $\bar{\Pi}$  is the system  $\Pi$  under the quasi-steady-state assumption for  $x_n$ , and the state variable of  $\Pi$  is in  $n$  dimensions, then they are identical except that  $\bar{\Pi}$  has the state variable  $\bar{\mathbf{x}} = (x_1, \dots, x_{n-1})$  where  $(x_1, \dots, x_n)$  are the components of the state variable of  $\Pi$ ; and  $\bar{\Pi}$  has the state equation:*

$$\frac{d}{dt} \bar{\mathbf{x}}(t; \mathbf{p}) = \bar{\mathbf{f}}(\bar{\mathbf{x}}(t; \mathbf{p}) \mathbf{p}), \quad (2.11)$$

where

$$\bar{f}_i(\bar{\mathbf{x}}, \mathbf{p}) = f_i((x_1, \dots, x_{n-1}, x_n^*), \mathbf{p}) \quad (2.12)$$

and  $x_n^*$  is the value taken by  $x_n$  such that the state variable is on the nullcline of  $x_n$  when the other parameters take the values  $x_1, \dots, x_{n-1}$  (Murray, 2002, Chapter I, Definition 11.4.).

**Definition 2.12.** The Jacobian of a smooth function  $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is the matrix:

$$\mathbf{J}(\mathbf{f}(\mathbf{x})) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \dots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \dots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} \quad (2.13)$$

(Müller & Kuttler 2015 Subsection 2.3.2).

**Example 2.3.** The **two configuration Langmuir model** may now be analysed in terms of the previous definitions.

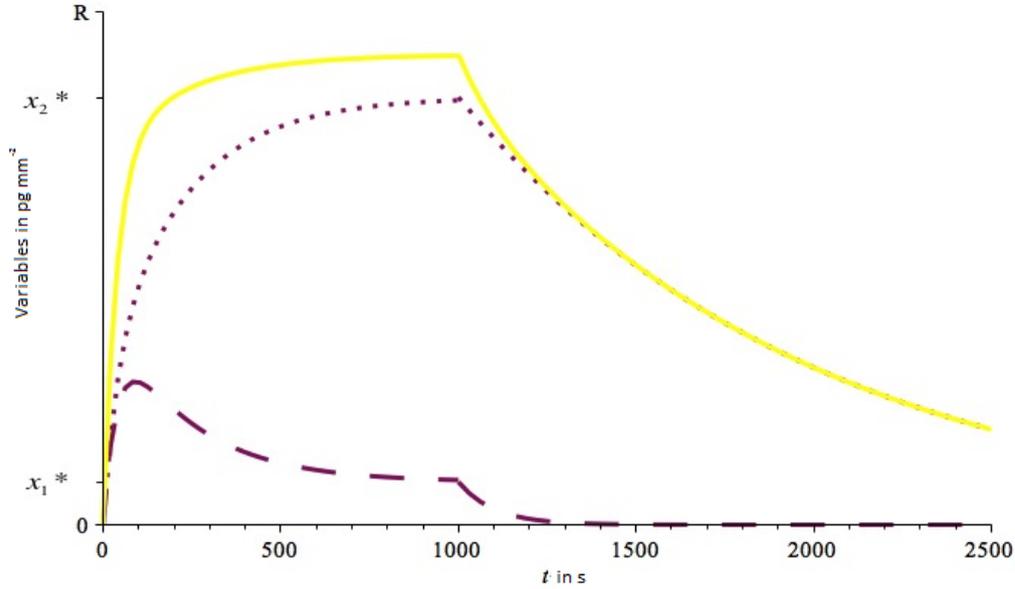
Figure 2.5 shows phase portraits of the model in either phase, with nullclines and steady states, marked as well as a trajectory of a single solution, and a graph of the state and output variables of this model against time, for the parameters shown in Table 2.4. The significance of these plots is that they can be used to develop a detailed understanding of what models predict about the inner states of the systems and as a result how models differ from each other, not just in terms of their output solutions, but their internal states.

In Fig. 2.5 (a) we see initially a rapid increase in the amount of each dimer as well as the output signal. The increase in output signal begins to slow as free ligand becomes less available.

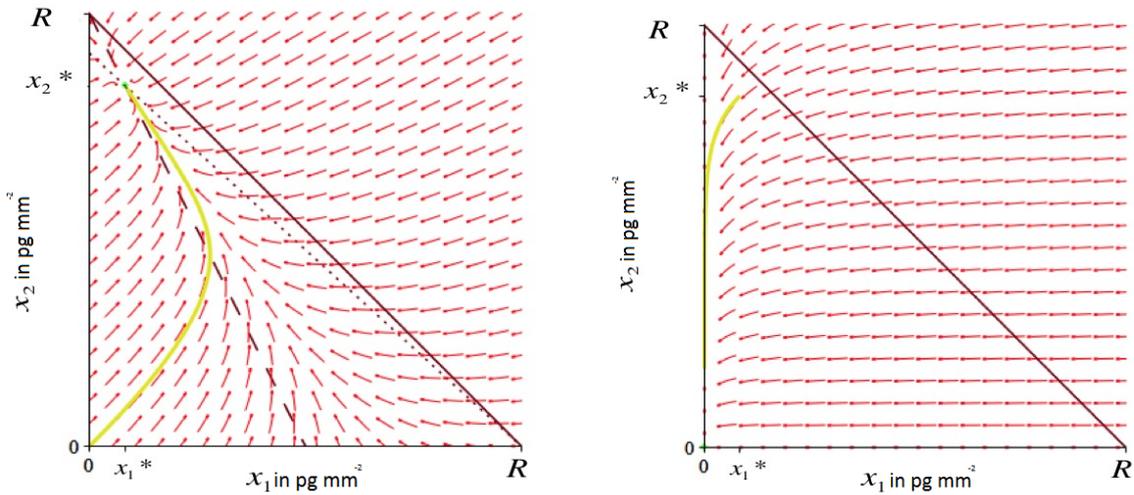
By drawing phase portraits like Fig. 2.5 (b) we can observe what is happening to the internal states of the system - generally only one will be shown for each phase, and it will have parameters chosen similar to those estimated by the model fitting conducted in later

Parameter	$k_{a1}$	$k_{d1}$	$k_{a2}$	$k_{d2}$	$R$	$C_T$
Value	$1\text{E-}3\text{M}^{-1}\text{s}^{-1}$	$7\text{E-}2\text{s}^{-1}$	$5\text{E-}4\text{M}^{-4}\text{s}^{-1}$	$2\text{E-}3\text{s}^{-1}$	$500\text{pg nm}^{-2}$	$100\text{nM}$

**Table 2.4:** Values taken by parameters in Fig. 2.5



(a) A visualisation of the state and output variables of the Langmuir model against time. The state variables  $x_1$  and  $x_2$  are shown as a dotted line and a dashed line respectively. Total concentration of bound analyte,  $y/\alpha$ , is shown as a yellow line.



(b) Phase portrait for the association phase, with nullclines for  $x_1$  shown as a dotted line, and for  $x_2$  as a dashed line.

(c) Phase portrait for the dissociation phase, with nullclines either state variable obscured by the axes

**Figure 2.5:** Phase portraits and visualisation of the two configuration Langmuir model.

chapters.

For the two configuration Langmuir model the binding configuration with the greatest association rate constant (in Fig. 2.5  $k_{a2}$ ) increases fastest in the early part of the association phase. As the amount of free antigen decreases the binding configuration that increases fastest becomes the one with the greatest ratio between its on rate constant and off rate constant ( $k_{a1}/k_{d1}$  or  $k_{a2}/k_{d2}$ ).

In Fig. 2.5 (b) this is illustrated by the two state variables, both having equal association rates increasing equally initially, but  $x_1$  slowing faster than  $x_2$ . This continues until the solution curve crosses the null-cline of  $x_1$ , and the amount of dimer bound in the second configuration becomes constant before beginning to decrease. In the latter part of the association phase,  $x_1$  decreases rapidly, but the output signal continues to increase because of the increase in the amount of antigen bound in the first configuration. As the phase continues, the solution approaches the steady state, but does not reach it.

When the input function changes at  $t_1$  from  $C$  to 0 the null-clines of the system move to being along the axis, and the steady state moves to the origin. We see this in Fig. 2.5 (c). Both state variables rapidly decrease in value, and move towards the new steady state.

The system reaches a steady state just before the end of the association phase. In the dissociation phase, both null-clines move to the axis, moving the steady state to the origin (Fig. 2.5 c). We see exponential decay in the number of dimers of either configuration. Because of the faster rate of dissociation of dimers of the first type, their population quickly becomes negligible, and the output becomes dominated by the signal from the second configuration.

## 2.4 Gröbner bases, concepts from abstract algebra and the Wronskian

Some of the identifiability analyses that will be used on the models later in this thesis will rely on concepts from abstract algebra. So they can be discussed it is necessary to make some definitions.

**Definition 2.13.** A *polynomial ring*  $\mathbb{P}(\mathbf{x})$  in  $\mathbf{x}$  over  $\mathbb{R}^n$  for some  $n \in \mathbb{N}$  is a set of polynomials in the components of  $\mathbf{x} \in \mathbb{R}$ , closed under the commutative operations of addition, multiplication and scalar multiplication; containing additive inverses for all elements; and in which multiplication is distributive over addition (Awange & B.Paláncz., 2016, Definition 3.3).

In this thesis the set of polynomials that will generally be considered is the real valued polynomials on  $x \in \mathbb{R}^n : n \in \mathbb{N}$ . This is a polynomial ring, commonly referred to as  $R[\mathbf{x}]$ .

Another kind of set we will be interested in is the set of polynomials that can be obtained by algebraic operations on a small number of polynomials. Each of these sets is an ideal.

**Definition 2.14.** A *polynomial ideal*  $I(\mathbf{x}) \subset \mathbb{P}(\mathbf{x})$  is a subset of the polynomial ring containing 0, that is closed under addition and multiplication by elements of the ring i.e. for all  $p \in I(\mathbf{x})$  and  $q \in \mathbb{P}(\mathbf{x})$  then  $p(\mathbf{x}) \times q(\mathbf{x}), p(\mathbf{x}) + q(\mathbf{x}) \in I(\mathbf{x})$  (Awange & B.Paláncz., 2016, Definition 4.1).

Some of these identifiability tests will involve generating ideals from algebraic operations on polynomials. As a result, it is useful to have a concept of independence. It is clear that there is something very different about the set generated from algebraic operations

on  $(x_1 + x_2, x_1)$  and one generated by the same operations on  $(x_1 + x_2, 2x_1 + 2x_2)$ .

**Definition 2.15.** *A set of polynomials  $p_1, \dots, p_n$  are algebraically **independent** if, when assembled into a vector, the Jacobean of that vector is non-singular (Lefschetz, 2005, Chapter I, Definition 11.4.).*

Equivalently, this holds if there is no annihilating polynomial, that is there is no non-zero polynomial on  $R[p_1, \dots, p_n]$  that is equal to zero.

The word independent will also be used to describe equations. Two equations will be said to be independent if they can be rewritten in the form  $0 = p_1(\mathbf{x})$ ,  $0 = p_2(\mathbf{x})$  where  $p_1(\mathbf{x})$  and  $p_2(\mathbf{x})$  are independent polynomials.

We are also interested in mappings between polynomial rings.

**Definition 2.16.** *Let  $A$  and  $B$  be commutative rings with identity. A ring homomorphism is a mapping  $\psi : B \rightarrow A$  such that for all  $x, y \in B$ :*

$$\psi(x + y) = \psi(x) + \psi(y), \tag{2.14}$$

$$\psi(xy) = \psi(x)\psi(y), \tag{2.15}$$

$$\psi(1) = 1. \tag{2.16}$$

(Forsman, 1991, Definition 1.10)

**Definition 2.17.** *Let  $a$  be an ideal in the ring  $A$  and  $\psi : B \rightarrow A$  a homomorphism, where  $A$  is a sub-ring of  $B$ . Then the **contraction** of  $a$  to  $B$  is the ideal  $a^c = \psi^{-1}(a)$  (Forsman, 1991, Definition 1.35) see also Lefschetz (2005, Chapter 3, Exercise 13).*

Calculus will also be used on elements from rings so it is useful to explore some concepts

from differential algebra.

**Definition 2.18.** A *derivation* of a ring  $A$  is a mapping  $\delta$  from  $A$  into itself satisfying:

$$\delta(x + y) = \delta(x) + \delta(y), \delta(xy) = \delta(x)y + x\delta(y). \quad (2.17)$$

A *differential ring* is a commutative ring with identity and a derivation. (Margaria, 2000, Definitions 2.9-2.10).

**Definition 2.19.** A *differential polynomial ring*  $\mathbb{P}\{\mathbf{x}\}$  for  $\mathbf{x} \in \mathbb{R}^n$  is the polynomial ring in the infinite indeterminates  $x_1, x_1^{(1)}, \dots, x_2, x_2^{(1)}, \dots, x_n, x_n^{(1)}, \dots$  where  $\delta(x_i^{(j)}) = x_i^{(j+1)}$  for  $i = 1, \dots, n$  and  $j \in \mathbb{N}_{\geq 0}$  (Margaria, 2000, Definition 2.12).

Margaria (2000, Example 2.8) notes that any polynomial ring may be converted into a differential ring by imposing the trivial derivation  $\delta(P) = 0P \in \forall \mathbb{P}\{\mathbf{x}\}$

**Definition 2.20.** A *differential polynomial ideal*,  $I\{\mathbf{x}\}$  is an ideal of a differential polynomial ring  $\mathbb{P}\{\mathbf{x}\}$  that is closed under  $\delta$  (Margaria, 2000, Definition 2.21).

The concept of a differential polynomial ideal is a very powerful tool. The state-space and output equations that this chapter has centred on may be rewritten as sets of differential polynomials equated to zero ( $0 = \mathbf{x} - \mathbf{f}(\mathbf{x})$  and  $0 = \mathbf{y} - g(\mathbf{x})$ ). The differential ideals generated by these sets include what we might think of as sets of descriptions of the behaviour of the models - the polynomials we can generate by doing arithmetic operations on these equations and their derivatives.

Now that we have the tools to create infinite numbers of polynomials we have to develop tools to sift them - i.e. to find the ones we want to work with. To do that we use Gröbner bases, but first we have to have a concept of what makes a polynomial desirable. Some desirable properties might be minimal degree in the polynomials intermediates, to think

about this we use monomial ordering.

**Definition 2.21.** A *total order* on the set  $S$  is a binary relation denoted  $\leq$  such that:

$$\begin{aligned} s_1 \leq s_2 \text{ or } s_2 \leq s_1 \\ s_1 \leq s_2 \text{ and } s_2 \leq s_1 &\implies s_2 = s_1 \\ s_1 \leq s_2 \text{ and } s_2 \leq s_3 &\implies s_1 \leq s_3 \end{aligned} \tag{2.18}$$

(Nederpelt, 2004, Chapter 20.2).

**Definition 2.22.** A *monomial ordering* is a total order on the set of all monomials (polynomials with a single term) in a given polynomial ring  $\mathbb{P}(\mathbf{x})$ , such that for any three monomials  $a$ ,  $b$  and  $c$ :

$$a \leq b \iff ac \leq bc \tag{2.19}$$

$$a \leq ab \tag{2.20}$$

(Cox *et al.*, 2007, Section 2.2).

**Definition 2.23.** In a *pure lexicographical ordering*, Monomials are compared first by their in their intermediate of highest order with ties broken by degree each other intermediate in order; i.e. for the order  $x > y$  then  $x^i y^j > x^k x^l$  if  $i > k$  or  $i = k$  and  $j > l$ . (Cox *et al.*, 2007, Section 2.2 Definition 3).

**Definition 2.24.** A *Gröbner basis*  $G$  for a polynomial ideal  $I(x)$  is a generating set for it, where the smallest polynomial ideal containing the leading terms of the polynomials in  $I(x)$  is the polynomial ideal  $(G)$ . Cox *et al.* (2007, Section 2.5 Definition 5)

Due to the desirable properties of Gröbner basis they also have these properties

- the leading term of each polynomial in  $I(x)$  is divisible by the leading term of some polynomial in the basis  $G$ ;
- the multivariate division of any polynomial in the polynomial ring  $\mathbb{P}(\mathbf{x})$  by  $G$  gives a unique remainder;
- the multivariate division by  $G$  of any polynomial in the ideal  $I$  has remainder 0.

**Definition 2.25.** A *reduced Gröbner basis* is a Gröbner basis where the leading coefficient of each element of the basis is 1 and no monomial in any element of the basis is in the ideal generated by the leading terms of the other elements of the basis (Cox et al., 2007, Section 2.7 Definition 4).

The value of Gröbner bases is that they are used as the basis of algorithms that they can be used to solve systems of multivariate polynomial equations. In this thesis these algorithms shall be implemented through the Maple package *Groebner*. However when Gröbner bases and related algorithms are used the emphasis will not be put on the individual algorithms chosen but on the choice of multivariate polynomial that are put into these algorithms.

The full documentation for these algorithms may be found in the Maple manual (Maplesoft, 2005-2015).

In some cases we will want to check the independence of the monomial terms of these differential polynomials, for this we use the Wronskian.

**Definition 2.26.** The *Wronskian* for a group of  $n$  real valued functions, differentiable

on the interval  $T$  is the function  $W(m_1, \dots, m_n)$

$$W(m_1, \dots, m_n)(t) = \det \left( \begin{bmatrix} m_1(t) & m_2(t) & m_3(t) & m_4(t) \\ m_1^{(1)}(t) & m_2^{(1)}(t) & m_3^{(1)}(t) & m_4^{(1)}(t) \\ \vdots & \vdots & \vdots & \vdots \\ m_1^{(n-1)}(t) & m_2^{(n-1)}(t) & m_3^{(n-1)}(t) & m_4^{(n-1)}(t) \end{bmatrix} \right), \text{ for } t \in T$$

(Bender & Orszag, 1978, Page 9).

If the functions in the Wronskian  $m_i$  are linearly dependent, then so are the columns of the matrix, so the Wronskian vanishes. Because of this, the Wronskian can be used to show that a set of differentiable functions is linearly independent on an interval by showing that it does not vanish identically, although it may, however, vanish at isolated points.

## 2.5 Lie derivatives, fibres and continuity of solutions

The models combine vector fields and differential operators, so Lie-derivatives will be useful. A Lie derivative is a coordinate-invariant evaluation of the change of a vector or scalar field, along the flow of another vector field.

**Definition 2.27.** *The **Lie-derivative** of a scalar function  $g(x)$  on a vector field  $\mathbf{f}$  is*

$$L_{\mathbf{f}}g = \nabla g \cdot \mathbf{f}$$

where  $\nabla g$  is the gradient of  $g$  and  $\cdot$  is the dot product operator (Gilkey et al., 2015, Section 7.3.1).

**Definition 2.28.** *The **repeated Lie-derivative** of a scalar function  $g(x)$  on a vector*

field  $\mathbf{f}$  is defined iteratively as

$$L_{\mathbf{f}}^i g = L_{\mathbf{f}} \left( L_{\mathbf{f}}^{i-1} g \right)$$

where  $\nabla g$  is the gradient of  $g$  and  $\cdot$  is the dot product operator (Gilkey *et al.*, 2015, Section 7.3.1).

As the models are associated with functions it is useful to develop some basic concepts related to functions to discuss them.

**Definition 2.29.** 1. A function is **injective** if no two elements of its domain are mapped to the same element of its codomain.

2. A function is **bijective** if it is injective and its codomain is its image (Halbeisen, 2012, Section 3.3, Axiom 6).

3. For a map  $f : X \mapsto Y$  the **fibre** of an element  $y \in Y$ , commonly denoted by  $f^{-1}(y)$ , is defined as the set of elements in  $X$  that are mapped by  $f$  onto the same element of  $Y$  (Weisstein, 2010).

In the next chapter all of these concepts will be brought together to make tools for identifiability analyses, which will be carried out in the rest of the thesis.

# Chapter 3

## Theory - Identifiability and Indistinguishability

### 3.1 Structural Identifiability

The models considered in this thesis are intended to be fitted to time series data so that parameters may be estimated. This is done by finding parameter values for which the residual sum of squares for the model is at its global minimum. However, there may be multiple parameter vectors for which a model provides the same minimal value because they have the same output.

In more precise language, let  $\mathbf{p} \in \Omega \subset \mathbb{R}^n$  represent a vector of all the unknown parameters of the model, and  $\Omega$  the set of physically meaningful parameters for the model. We denote the output of the model at the time  $t \in T$  as  $y(t; \mathbf{p})$  where  $T$  is the interval representing the time for which data are collected.

**Definition 3.1.** *The parameter vectors  $\mathbf{p}, \bar{\mathbf{p}} \in \Omega$  are **indistinguishable** if and only if  $y(t; \mathbf{p}, \mathbf{I}) = y(t; \bar{\mathbf{p}}, \mathbf{I}), \forall t \in T$ , for a generic input  $\mathbf{I}$ , this is written as  $\mathbf{p} \sim \bar{\mathbf{p}}$ .*

**Definition 3.2.** *The equivalence class of a parameter vector  $\mathbf{p} \in \Omega$  is  $E(\sim, \mathbf{p})$ , where  $E(\sim, \mathbf{p}) = \{\bar{\mathbf{p}} \in \Omega : \mathbf{p} \sim \bar{\mathbf{p}}\}$ .*

In terms of Definition 2.29 and Definition 2.7  $\mathbf{p}$  and  $\bar{\mathbf{p}}$  are indistinguishable if and only if, for a generic input  $\mathbf{I}$  the pairs  $(\mathbf{p}, \mathbf{I})$  and  $(\bar{\mathbf{p}}, \mathbf{I})$  are in the fibre of the input-output mapping  $\mathbf{Z}_\Sigma$ . In this thesis the usage of the word indistinguishable will be extended to the state-space solutions of a model.

**Definition 3.3.** *The state-space solutions  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}), \bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I}) \in \mathcal{F}_\Pi$  are **indistinguishable** if and only if  $Y_\Sigma(\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I})) = Y_\Sigma(\bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I}))$ , for a generic input  $\mathbf{I}$ , this is written as  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}) \sim \bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I})$ . (Sontag, 1998, Example 6.1.2, Lemma 6.1.3).*

When state-space solutions meet this criteria they will be described as indistinguishable solutions of a model. Through this concept of indistinguishable state-space solutions we can develop a concept of observability.

**Definition 3.4.** *A system is observable if and only if  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}) \sim \bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I})$  is equivalent to  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}) = \bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I})$ . Compare Sontag (1998, Definition 6.14).*

That is given an input, the state-space solution equivalence classes  $E(\sim, \mathbf{p})$  all singletons. This also means that the state of the system may be determined throughout the time it is observed.

As so many of the objects we are concerned with consider a given input its useful to create subsets of the sets from Definition 2.5 and 2.7 associated with a generic input.

**Definition 3.5.** *We define  $\mathcal{F}_\Pi(\mathbf{I}) \subset \mathcal{F}_\Pi$  by  $\mathcal{F}_\Pi(\mathbf{I}) = \{\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}(\cdot; \mathbf{p})) : \mathbf{p} \in \Omega\}$  for some specified  $\mathbf{I} \in \mathcal{I}_\Pi$ , and  $\mathcal{G}_\Sigma(\mathbf{I}) \subset \mathcal{F}_\Sigma$  by  $\mathcal{G}_\Sigma(\mathbf{I}) = \{y(\cdot; \mathbf{p}, \mathbf{I}) : \mathbf{p} \in \Omega\}$  for some specified  $\mathbf{I} \in \mathcal{I}_\Sigma$ .*

As we are concerned with input output behaviour, it's useful to consider sets of state-space solutions and output solutions that can be achieved from a given input.

**Definition 3.6.** *A system is observable if and only if  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}) \sim \bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I})$  is equivalent to*

$\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}) = \bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I})$ . Compare Sontag (1998, Definition 6.14).

From the concept of indistinguishable parameter vectors we can develop categories for the identifiability of individual parameters:

**Definition 3.7.** For a generic parameter vector,  $\mathbf{p} \in \Omega$ ,  $p_i$  is **globally identifiable** if and only if for any  $\bar{\mathbf{p}} \in \Omega$  such that  $\mathbf{p} \sim \bar{\mathbf{p}} \Rightarrow \bar{p}_i = p_i$ .

**Definition 3.8.** For a generic parameter vector,  $\mathbf{p} \in \Omega$ ,  $p_i$  is **locally identifiable** if and only if there exists a neighbourhood  $N(\mathbf{p}) \subset \Omega : \bar{\mathbf{p}} \in N(\mathbf{p}), \mathbf{p} \sim \bar{\mathbf{p}} \Rightarrow \bar{p}_i = p_i$ .

We may observe that if a parameter is globally identifiable then it is also locally identifiable.

**Definition 3.9.** For generic  $\mathbf{p} \in \Omega$ ,  $p_i$  is **unidentifiable** if and only if it is not locally identifiable.

As a result we can categorise the identifiability of models as follows:

**Definition 3.10.** A model is **structurally globally identifiable (SGI)** if all parameters are globally identifiable.

Relating this to the Definition 2.7, a model  $\Sigma$  being SGI is equivalent to its input-output mapping,  $Z_\Sigma(\mathbf{P}, \cdot)$ , being bijective, i.e., that all of its fibres are singletons.

**Definition 3.11.** A model is **structurally locally identifiable (SLI)** if all parameters are locally identifiable, but not all are globally identifiable.

We may observe that if a model is SGI then it is also SLI.

**Definition 3.12.** A model is **structurally unidentifiable (SU)** if one or more parameters are not locally identifiable.

Just as we were able to go from discussing the identifiability of the parameter vector to the identifiability of parameters we can also extend the notion of observability from the state variable to the state variable components.

**Definition 3.13.** *A state variable component  $x_i$  is observable if and only if  $\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}) \sim \bar{\mathbf{x}}(\cdot; \bar{\mathbf{p}}, \mathbf{I})$  implies  $\mathbf{x}_i(\cdot; \mathbf{p}, \mathbf{I}) = \bar{\mathbf{x}}_i(\cdot; \bar{\mathbf{p}}, \mathbf{I})$ .*

There are a number of methods for identifiability analysis in the literature that could either be used directly or adapted for use on the models presented in this thesis. However, due to the complexity of the models, computational difficulty can arise with the application of these methods. As a result only three methods are presented in detail here.

### 3.1.1 The Taylor Series Approach

Pohjanpalo (1978a) proposed the Taylor series approach (TSA), a method which uses the fact that for the outputs of two parametrisations ( $\mathbf{p}$  and  $\bar{\mathbf{p}}$ ) of the same model to be identical, the coefficients of Taylor series expansions of their outputs when the system is initialised must also be identical.

It is convenient to write this in terms of Lie-derivatives. However, to do this it is useful to introduce a more compact notation. As we will be comparing the same state-space equations with different parameter vectors, we abbreviate these by writing:  $\mathbf{f}(\mathbf{x}(t; \mathbf{p}), \mathbf{I}(t; \mathbf{p}), \mathbf{p}) = \mathbf{f}(\mathbf{x})$  and  $\mathbf{f}(\mathbf{x}(t; \bar{\mathbf{p}}), \mathbf{I}(t; \bar{\mathbf{p}}), \bar{\mathbf{p}}) = \bar{\mathbf{f}}(\mathbf{x})$  where these are defined as they were in Def. 2.4. Similarly we abbreviate the output mappings,  $\mathbf{g}(\mathbf{x}(t; \mathbf{p}), \mathbf{p}) = \mathbf{g}(\mathbf{x})$  and  $\mathbf{g}(\mathbf{x}(t; \bar{\mathbf{p}}), \bar{\mathbf{p}}) = \bar{\mathbf{g}}(\mathbf{x})$  where these are defined as they were in Def. 2.6.

Pohjanpalo (1978a) therefore requires:

$$L_{\mathbf{f}}^i \mathbf{g}(\mathbf{x}_0) = L_{\bar{\mathbf{f}}}^i \bar{\mathbf{g}}(\mathbf{x}_0) \quad \forall i \in \mathbb{N}. \quad (3.1)$$

The number of equations that need to be satisfied for two systems to have identical outputs is infinite. Sets of equations from this list can be solved to obtain relationships between the elements of  $\mathbf{p}$  and  $\bar{\mathbf{p}}$ .

Whilst the number of these equations is infinite, they may not all be linearly independent. As a result, it is important to determine if there is a dependence of higher order terms on lower order ones. If there is, then only a finite number of equations need to be considered to find the relationships required for  $\mathbf{p} \sim \bar{\mathbf{p}}$ .

In this regard, some work has been done for rational systems to give an upper bound on the number of Taylor series coefficients required. Margaria *et al.* (2001) presents a significant result - that for a model of the kind defined in Def. 2.6 with a state variable  $\mathbf{x}(t; \mathbf{p}) \in \mathbb{R}_{\geq 0}^m$ ,  $y(t) \in \mathbb{R}_{\geq 0}$ , a parameter vector  $\mathbf{p} \in \mathbb{R}_{> 0}^n$ , a state-space mapping  $\mathbf{f}(\mathbf{x}(t; \mathbf{p}))$ , an smooth input function  $\mathbf{I}(t; \mathbf{p})$  that is a quotient of polynomials in  $\mathbf{x}(t; \mathbf{p})$  and  $\mathbf{I}(t; \mathbf{p})$ , then the number of Taylor series coefficients needing to be considered is the first  $m + n$ .

A special case of this result that was earlier derived which will also be relevant (Vajda, 1984), which is that if  $\mathbf{f}(\mathbf{x}(t; \mathbf{p}))$  is linear only the first  $2m$  derivatives need to be considered.

The coefficients of the Taylor series are obtained by repeated differentiation of the output, and evaluation at the initial conditions where the state variables take known values. Because of this repeated differentiation, the coefficients of higher powers of the Taylor series are often long and contain complex non-linear relationships between parameters. As a result, there may be computational difficulty with solving these systems of equations. Limits to the number of derivatives required (Margaria *et al.*, 2001; Vajda, 1984) are significant because they limit the number of equations of the form of Eq. (3.1) that need to be considered.

It was noted in Section 2.2 that because the input function our models will take is

discontinuous, the trajectories of the state and output variables will be piecewise smooth rather than smooth with respect to time. As a result, this method would need some modification to be appropriate for the models we are considering.

*Adapting the Taylor series approach for input step functions* is relatively simple for models like the single configuration Langmuir model, where at each point in time the value of the state variable can be determined by the output variable (because the output mapping for this model,  $g(x) = \alpha x$  is bijective). In cases where the state variable can be known from the output variable, a second set of equations can be derived by considering equations of the form:

$$L_{\mathbf{f}}^i \mathbf{g}(\mathbf{x}_1) = L_{\bar{\mathbf{f}}}^i \bar{\mathbf{g}}(\mathbf{x}_1) \quad \forall i \in \mathbb{N} \quad (3.2)$$

where  $\mathbf{x}_1 = \mathbf{g}^{-1}(y(t_1))$ , and  $[0, t_1)$  and  $[t_1, t_2)$  are the intervals on which the input function takes a constant value, the association and dissociation phases,  $\mathbf{f}(\mathbf{x})$ ,  $\bar{\mathbf{f}}(\mathbf{x})$ ,  $\mathbf{g}(\mathbf{x})$  and  $\bar{\mathbf{g}}(\mathbf{x})$  continue to be state-space and output mappings.

These additional equations are helpful in two separate ways. Firstly, some parameters of the model may not be identifiable in the first interval but may be identifiable if behaviour on both intervals is considered. As a result, this adaptation of the method is necessary to ensure that the model is not wrongly categorised as SLI or SU.

Secondly for systems where  $\mathbf{f}(\mathbf{x})$  involves quotients or products of functions of  $\mathbf{x}$  we expect each successive Lie derivative to produce an equation dramatically more complicated (due to the product and quotient rules of differentiation). As a result if considering derivatives in both phases can reduce the numbers of derivatives required as determined by Margaria *et al.* (2001) and Vajda (1984), this would reduce computational time and may make otherwise intractable problems solvable.

However, the majority of the models we will consider do not have bijective output mappings. In this case Eq. (3.2) can still be derived, but in terms of a general solution to the

equation  $\mathbf{x}_1 = \mathbf{g}^{-1}(y(t_1))$ . This means that whilst these equations can still be helpful in the ways listed above, they carry with them some additional computational complexity because they also need to be solved for the new variables introduced in the general solution to  $\mathbf{x}_1 = \mathbf{g}^{-1}(y(t_1))$ .

The solution to this set of equations also gives us an additional piece of information about the model; whether there are multiple initial conditions in the state-space, which for some parameter vectors the model will give identical outputs if the dissociation phase begins on them. If it is demonstrated that there is only one point in state-space that the dissociation phase can begin with that will give an equal output solution, this is a necessary condition for the state-output map to be bijective. Conversely, if it is assumed that this map is bijective, this simplifies the solution process of these equations, because it no longer needs to be solved for the parameters introduced in  $\mathbf{x}_1 = \mathbf{g}^{-1}(y(t_1))$ .

**Example 3.1.** The **single configuration Langmuir model** (Eq. (2.5) and Eq. (2.6)) has four parameters and one variable, and was previously used as an example of a model (Example 2.1). It can now be used to illustrate the Taylor series approach. This begins with evaluating iterated Lie-derivatives of the models output ( $\mathbf{g}(\mathbf{x}) = \alpha x$ ), at the initial conditions. Conveniently its state-space equation (from Eq. (2.5)) can be rewritten in a very compact form:

$$f(x) = k_a IR - (k_a I + k_d)x. \quad (3.3)$$

Conveniently this expression is linear and when  $t = 0$ ,  $x = 0$ . This us allows us to write a compact expression for the  $i^{\text{th}}$  Lie-derivative:

$$L_{\mathbf{f}}^i \mathbf{g}(0) = \begin{cases} 0 & \text{if } i = 0 \\ \alpha k_a C_T R & \text{if } i = 1 \\ \alpha k_a C_T R (-1)^i (k_a C_T + k_d)^i & \text{if } i > 1. \end{cases} \quad (3.4)$$

This is done for two parametrisations of the model ( $\mathbf{p}$  and  $\bar{\mathbf{p}}$ ), and these Lie-derivatives of the output have to be equal for either parametrisation to give the same output (Eq. (3.1)):

$$\alpha(k_a C_T R) = \alpha(\bar{k}_a \bar{C}_T \bar{R}) \quad (3.5)$$

$$\alpha(-1)^i (k_a C_T + k_d)^i (k_a C_T R) = \alpha(-1)^i (\bar{k}_a \bar{C}_T + \bar{k}_d)^i (\bar{k}_a \bar{C}_T \bar{R}) \text{ for } i > 1 \quad (3.6)$$

These can be solved together to determine the identifiability of the system. However, as these equations form a geometric series, they can all be generated by their first two terms. This illustrates the limit to the number of derivatives from Vajda (1984), only two derivatives are required.

A simpler set of generators, equivalent to the first two terms are the initial term and the absolute value of the multiple that is applied to successive terms:

$$k_a C_T R = \bar{k}_a \bar{C}_T \bar{R}, \quad (3.7)$$

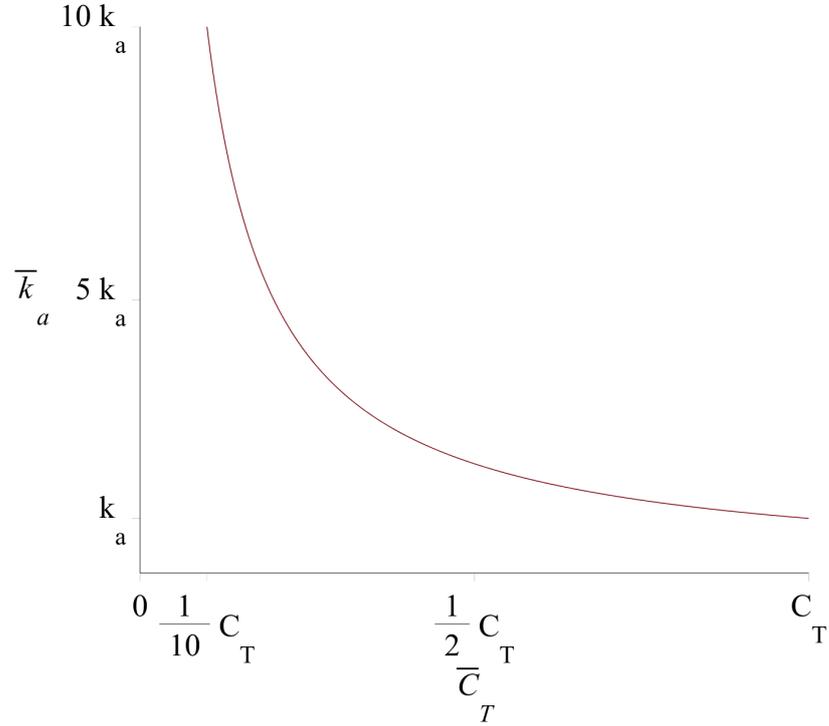
$$k_a C_T + k_d = \bar{k}_a \bar{C}_T + \bar{k}_d. \quad (3.8)$$

As a result, in the association phase for  $\mathbf{p} \sim \bar{\mathbf{p}}$ , where  $\mathbf{p} = (k_a, k_d, C_T, R)$ :

$$\bar{\mathbf{p}} = \left( \frac{k_a C_T + k_d - \bar{k}_d}{\bar{C}_T}, \bar{k}_d, \bar{C}_T, \frac{k_a C_T R}{k_a C_T + k_d - \bar{k}_d} \right), \quad (3.9)$$

where  $\bar{k}_d$  and  $\bar{C}_T$  can take any value greater than zero.

However, other parameters may become identifiable when we consider the output of the dissociation phase, so the adaptations made for systems with step function inputs are used. As there is a bijective relationship between the state and output variables for this model, the point at which it starts the dissociation phase is also known and the previous method can be repeated on the initial conditions of the dissociation phase  $x(t_1) = \frac{y(t_1)}{\alpha}$ ,



**Figure 3.1:** Graph showing  $(C_T, k_a)$  pairs indistinguishable to those shown in Table 2.2.

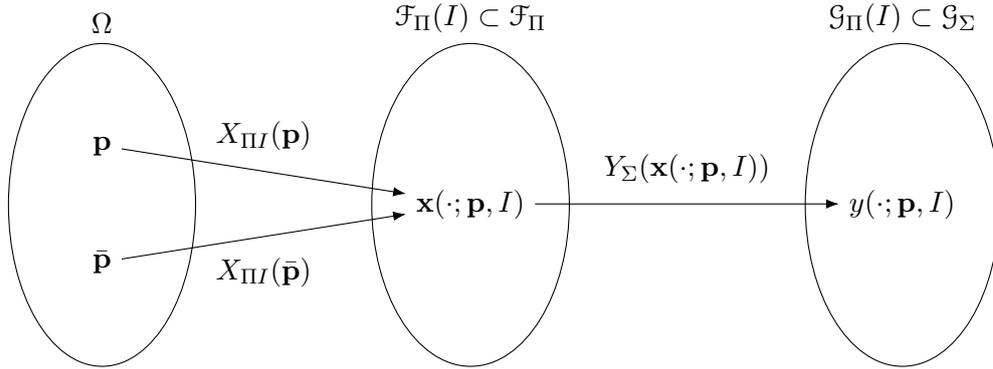
giving additional equations:

$$(-k_d)^i \frac{y(t_1)}{\alpha} = (-k_d)^i \frac{y(t_1)}{\alpha} \text{ for } i > 0. \quad (3.10)$$

This, together with the previous two equations, can be solved to determine that for  $\mathbf{p} \sim \bar{\mathbf{p}}$ :

$$\bar{\mathbf{p}} = \left( \frac{k_a C_T}{\bar{C}_T}, k_d, \bar{C}_T, R \right), \quad (3.11)$$

where  $\bar{C}_T$  can take any value greater than zero. The set of parameter vectors indistinguishable from a given parameter vector, i.e. its equivalence class can be visualised as a 1 dimensional curve in parameter space. This is shown in Fig. 3.1.



**Figure 3.2:** In the L1 model indistinguishable parameters are indistinguishable because for a generic input  $I$ ,  $X_{\Pi}(\cdot, I)$  is surjective, even though  $Y_{\Sigma}(\cdot)$  is bijective.

An interesting aside is that this is because the model's input-state mapping,  $X_{\Pi}$  is not bijective, i.e.  $X_{\Pi}(\cdot; \mathbf{p}, I) = X_{\Pi}(\cdot; \bar{\mathbf{p}}, I)$  is equivalent to Eq. (3.11) not  $\mathbf{p} = \bar{\mathbf{p}}$ ; even though its output mapping,  $\mathbf{g}(\mathbf{x})$  and its output solution  $Y_{\Sigma}$  are both bijective (see Definition 2.5 and Definition 2.7). This is illustrated in Fig. 3.2.

**Example 3.2.** The **two configuration Langmuir model** was defined in Eq. (2.10), and used as an example of a model (Example 2.2). It can now be used to illustrate some of the finer points of the way the Taylor series approach was adapted for input step functions, for models that did not have bijective output mappings.

Similar to the previous example we can generate a list of equations showing the relationships that two parameter vectors must have to have the same output in the association phase.

Unlike the previous example, if two parameterisations give the same output at the final point of the association phase their state variables cannot be deduced, because the output function for this model,  $g(\mathbf{x}) = \alpha(x_1(t_1) + x_2(t_1))$  is not bijective. However, repeated Lie-derivatives can be taken of the output along the state-space function to give a relationship

between the parameters and state variables at the start of the dissociation phase:

$$(-k_{d1})^i x_1(t_1) + (-k_{d2})^i x_2(t_1) = (-\bar{k}_{d1})^i \bar{x}_1(t_1) + (-\bar{k}_{d2})^i \bar{x}_2(t_1) \text{ for } i \geq 0. \quad (3.12)$$

Where the barred variables and parameters represent the equivalent variables and parameters of a model with an identical output.

The equations obtained for either phase can be solved together to determine the identifiability of the parameters of the system. For  $\bar{\mathbf{p}} \sim \mathbf{p}$ , where  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R)$ :

$$\begin{aligned} \bar{\mathbf{p}} &= (k_{a1}/\gamma, k_{d1}, k_{a2}/\gamma, k_{d2}, C_T\gamma, R) \\ &\text{or } (k_{a2}/\gamma, k_{d2}, k_{a1}/\gamma, k_{d1}, C_T\gamma, R) \end{aligned} \quad (3.13)$$

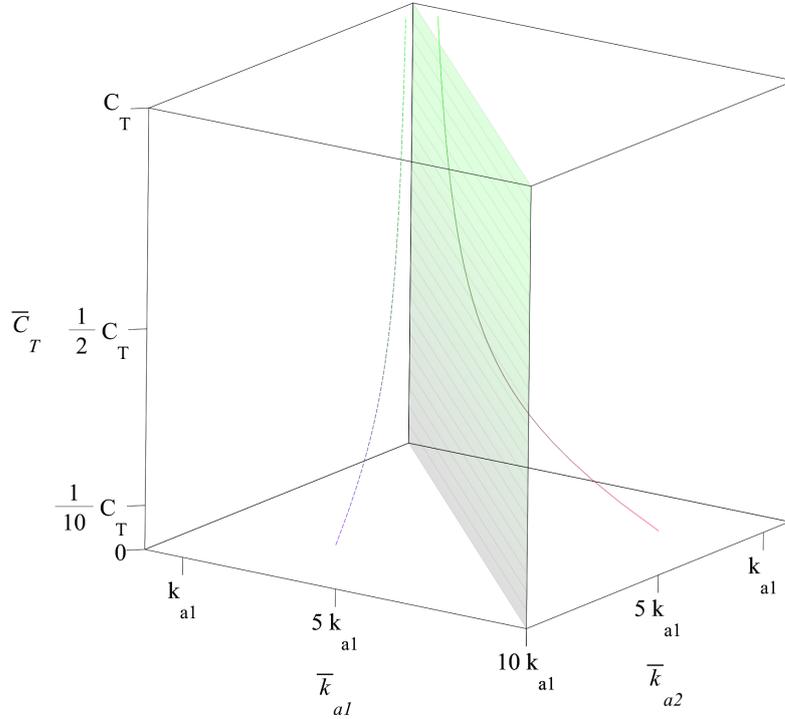
for  $\gamma = \bar{C}_T/C_T$ . Correspondingly the values taken by the state variables at the start of the dissociation phase should be related by:

$$\begin{aligned} (\bar{x}_1(t_1), \bar{x}_2(t_1)) &= (x_1(t_1), x_2(t_1)) \\ &\text{or } (x_2(t_1), x_1(t_1)). \end{aligned} \quad (3.14)$$

This means the model is SLI if and only if  $C_T$  is initially known, and SU otherwise. The set of equivalence class of a generic parameter vector can be visualised as a set of 1 dimensional curves in parameter space. These curves are shown in Fig. 3.3 for the parameters in Table 2.3.

As with the Langmuir model this arises because the input-state mapping  $X_\pi$  is not bijective, i.e.  $X_\pi(\cdot; \mathbf{p}, I) = X_\pi(\cdot; \bar{\mathbf{p}}, I)$  is equivalent to  $\bar{\mathbf{p}} = (\bar{k}_{a1}/\gamma, k_{d1}, k_{a2}/\gamma, k_{d2}, C_T\gamma, R)$  for  $\gamma = \bar{C}_T/C_T$ , not  $\bar{\mathbf{p}} = \mathbf{p}$ .

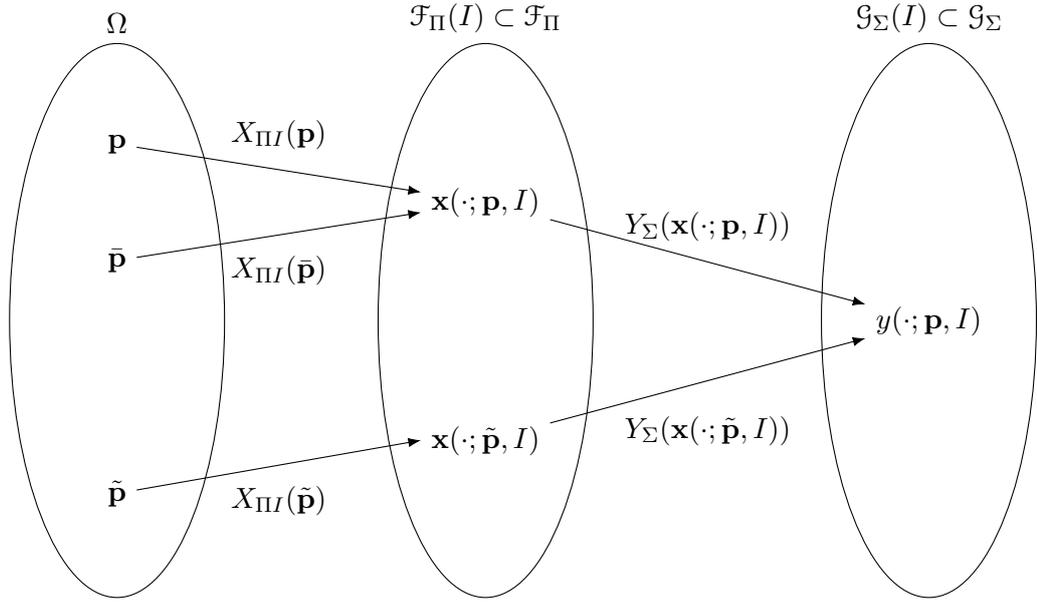
However, even if the parameters  $k_{a1}$  and  $k_{a2}$  were merged with  $C_T$  to make  $X_\pi$  bijective the model would still be only locally identifiable; this tells us that the state output mapping



**Figure 3.3:** Graph showing the set of parameters indistinguishable to those shown in Table 2.3 as two lines, one a reflection of the other in the plane  $k_{a1} = k_{a2}, k_{d1} = k_{d2}$ , one is dashed to represent that two of the dimensions it has moved in are not shown.

$Y_{\Sigma}$  is not bijective - but at this point we have not developed the tools to find fibres that are mapped by  $Y_{\Sigma}$  to the same output solution - although the form of this function is hinted at by the fact that different state-space solutions that are mapped to the same output solution start their dissociation phase with their state variables  $x_1(t_1)$  and  $x_2(t_1)$  reflected in the line  $x_1 = x_2$ .

Figure 3.4 illustrates this. Comparing Figure. 3.3, points on the same parameter-space curve are mapped onto the same state-space solution by  $X_{\Pi}$  whereas points on the reflected curve are mapped onto the same output solution by  $Y_{\Sigma}$ . This can be seen by the fact that on either curve the state-space equation ( $\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x}, \mathbf{p}, I)$ ) remains the same, but



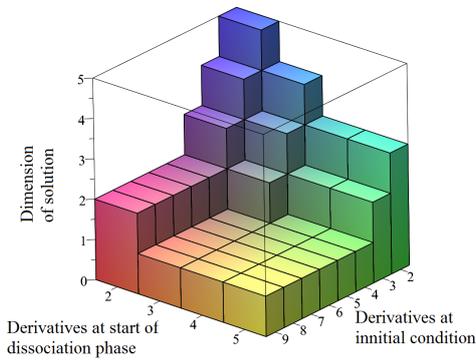
**Figure 3.4:** In the L2 model indistinguishable parameters are indistinguishable because both  $X_{\Pi}(\cdot)$  and  $Y_{\Sigma}(\cdot)$  are surjective, for a generic input  $I$ .

parameters on either curve are ultimately mapped onto the same output solution.

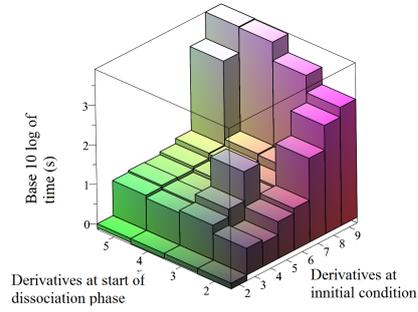
Interestingly Fig. 3.3 also illustrates how the identifiability of the model may be improved. The parameters that are only locally identifiable are only locally identifiable because the parameter vector can be on either of these curves. If we define  $k_{a1} > k_{a2}$  for the model we limit the parameter space in such a way that only one of these curves remains in it. As a result if we made this limit and  $C_T$  were known this would improve the models identifiability from SLI to SGI.

As this thesis will contain many models that have similar symmetries the fact that we can improve their identifiability by restricting parameter-space will be a common feature of their identifiability.

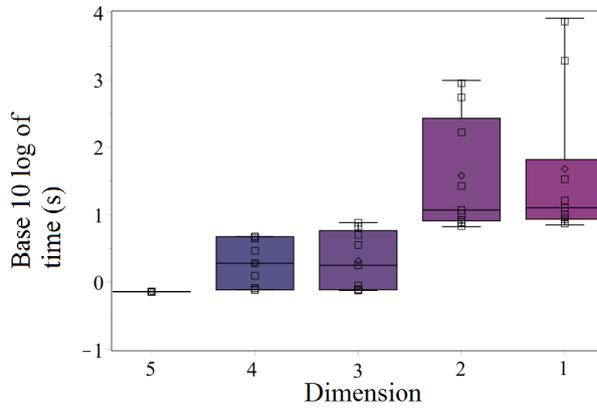
**Example 3.3.** These examples raise the question, of generally how many Lie Derivatives in either phase are required to determine the structural identifiability of the model.



(a) Dimensions of solution as numbers of derivatives at the start of either phase vary.



(b) Time taken to obtain a solution as numbers of derivatives at the start of either phase vary.



(c) Box-plots showing the distribution of time taken for solve against dimension of solution.

**Figure 3.5:** Graphs showing how the dimension of solution and the time taken to obtain the solution as numbers of derivatives of the output at the start of either phase vary.

Notably the existing upper bounds (Margaria *et al.*, 2001; Vajda, 1984) calculated are not for inputs like this, piecewise constant. As a result a series of maple programs were created to automate this, and allow for the timing of these programs (see Appendix A.3.2)

Previously the structurally indistinguishable parameters of the two example models have been visualised as single dimensional regions of parameter-space (Fig. 3.1, Fig. 3.3). The dimensionality of the region of parameter-space provides a convenient way of quantifying

its size so that the sizes of regions of parameter-space obtained by using different numbers of Lie-derivatives in either phase can be compared.

This is visualised in Fig. 3.5(a). The key element of this figure is the basin where the dimension of the solution is 1, where  $n_a > 3$ ,  $n_d > 2$ ,  $n_a + n_d \geq 8$ . In this region the solution is the one depicted in Fig 3.3, i.e. the solution is the equivalence class of  $\mathbf{p}$ .

Outside of this basin, we see the dimensions of partial solutions to the identifiability problem. Insufficient Lie-derivatives have been used, and as a result when they are solved for the parameters they give systems of equations like and except with fewer alternate parameters eliminated. For example if  $n_a = 0$ , and  $n_d > 2$  the partial solution is:

$$\begin{aligned} \bar{\mathbf{p}} &= (\bar{k}_{a1}, k_{d1}, \bar{k}_{a2}, k_{d2}, \bar{C}_T, \bar{R}) \\ &\text{or } (\bar{k}_{a2}, k_{d2}, \bar{k}_{a1}, k_{d1}, \bar{C}_T, \bar{R}) \end{aligned} \tag{3.15}$$

As the values taken by the state variables at the start of the dissociation phase are treated as parameters when in the Taylor series approach for step inputs, the four dimensional regions of parameter-space are associated with the equation:

$$\begin{aligned} (\bar{x}_1(t_1), \bar{x}_2(t_1)) &= (x_1(t_1), x_2(t_1)) \\ &\text{or } (x_2(t_1), x_1(t_1)). \end{aligned} \tag{3.16}$$

Later in this thesis when more complex models are introduced, these partial solutions will become useful, because in cases where there is not sufficient memory or time to obtain a full solution using the Taylor series approach, these partial solutions can be obtained. As no parameter vectors outside of a partial solution will be indistinguishable to a generic parameter vector, any indistinguishable parameter vector can be assumed to be from this set. This assumption therefore can speed up, and reduce the memory required for other methods, thus making intractable problems tractable.

Later in this chapter new concepts will be introduced that will allow us to determine whether sufficient numbers of derivatives have been considered, for an analysis of identifiability has been conducted for a model. However, here we may observe that the total number of derivatives required is the same number as established by Margaria *et al.* (2001).

A significant barrier to identifiability analyses is the computational time required. This is illustrated in Fig. 3.5(b) and Fig. 3.5(c). Notably even for this model with linear state-space and output equations computation times increase by orders of magnitude if sufficient derivatives are considered. This barrier will become more significant for models with products and quotients of variables in their state space equations, as the length of their derivatives will grow much faster with iterated Lie-derivatives.

These figures will also allow us to compare the time taken for an identifiability analysis with different methods.

### 3.1.2 Observability Rank Condition

Figure 3.4 raises the question of what properties must state-space solutions,  $\mathbf{x}(\cdot; \mathbf{p}, I)$ , have for them to be mapped onto the same output solution  $y(\cdot; \mathbf{p}, I)$ , and the possibility that investigating this could aid in structural identifiability analysis. One way of thinking about this problem would be to try and use the structure of the state-space and output equations to derive a point-wise mapping between the state space solution curves. Such a function has already been defined in the literature, and methods for using it to analyse the structural identifiability of models developed.

Hermann & Krener (1977) developed a concept related to identifiability, observability, and developed a related criterion, the observability rank criterion (ORC). This criterion was adapted by a series of authors for different purposes, and Evans *et al.* (2002) applied it to the identifiability of autonomous (inputless) systems. As our systems involve very simple

inputs (step functions), like the preceding method, this will require some modification to be applicable to the models we are interested in.

To do this functions have to be defined. A set of scalar valued functions, the Lie-derivatives of the output, are defined inductively with :

$$\begin{aligned}\mu_1(\mathbf{x}, \mathbf{p}) &= g(\mathbf{x}, \mathbf{p}), \\ &\vdots \\ \mu_m(\mathbf{x}, \mathbf{p}) &= L_{\mathbf{f}}\mu_{m-1}(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}\mu_{m-1}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}).\end{aligned}\tag{3.17}$$

Then they are composed into a vector field of the form:

$$\mathbf{H}(\mathbf{x}, \mathbf{p}) = (\mu_1(\mathbf{x}, \mathbf{p}), \dots, \mu_m(\mathbf{x}, \mathbf{p}))^T.\tag{3.18}$$

The ORC is satisfied if and only if, there exists an  $m \in \mathbb{N}$ , greater than or equal to the dimensionality of the state-space where the Jacobian of  $\mathbf{H}(\mathbf{x}, \mathbf{p})$  with respect to  $\mathbf{x}$ , evaluated at the initial conditions has full rank. Evans *et al.* (2005) proved that for models with indistinguishable parameter vectors ( $\bar{\mathbf{p}}$  and  $\mathbf{p}$ ), satisfying the ORC there exists a smooth map,  $\lambda : \mathcal{X} \rightarrow \mathcal{X}, \lambda : (\mathbf{x}(t; \bar{\mathbf{p}})) \mapsto \mathbf{x}(t; \mathbf{p})$ , defined on the state-space  $\mathcal{X}$ , which can be derived from the expression:

$$\mathbf{H}(\lambda(\mathbf{x}(t; \mathbf{p})), \mathbf{p}) = \mathbf{H}(\mathbf{x}(t; \mathbf{p}), \bar{\mathbf{p}}).\tag{3.19}$$

Using this function, a system of equations can be derived that can be solved to determine the relationship between two indistinguishable parameter vectors  $\bar{\mathbf{p}}$  and  $\mathbf{p}$ . We expect this function to take one of two forms:

$$\lambda(\mathbf{x}(t, \mathbf{p})) = L(\mathbf{p})\mathbf{x}(t, \mathbf{p})\tag{3.20}$$

where  $L(\mathbf{p})$  is an  $m \times m$  matrix; or

$$\lambda(\mathbf{x}(t, \mathbf{p})) = L_1(\mathbf{p})\mathbf{x}(t, \mathbf{p}), \text{ or } L_2(\mathbf{p})\mathbf{x}(t, \mathbf{p}), \text{ or } \dots, \text{ or } L_l(\mathbf{p})\mathbf{x}(t, \mathbf{p}) \quad (3.21)$$

where  $L_1(\mathbf{p}), L_2(\mathbf{p}), \dots, L_l(\mathbf{p})$  are  $m \times m$  matrices.

Evans *et al.* (2005) goes on to derive three equations in terms of this mapping  $\lambda$ , that have to be conserved for any pair of parameter vectors where  $\bar{\mathbf{p}} \sim \mathbf{p}$ . To reduce the complexity of the notation in these expressions, points on the solution curve of either model  $\mathbf{x}(t; \mathbf{p})$  and  $\mathbf{x}(t; \bar{\mathbf{p}})$  are written as  $\mathbf{x}$  and  $\bar{\mathbf{x}}$ ; and  $\mathbf{f}(\cdot, \mathbf{p})$  and  $\mathbf{f}(\cdot, \bar{\mathbf{p}})$  are written as  $\mathbf{f}(\cdot)$  and  $\bar{\mathbf{f}}(\cdot)$ . This system of equations is:

$$\begin{aligned} \lambda(\mathbf{x}_0(\bar{\mathbf{p}})) &= \mathbf{x}_0(\mathbf{p}) \\ \mathbf{f}(\lambda(\mathbf{x})) &= \frac{\partial \lambda(\mathbf{x})}{\partial \mathbf{x}} \bar{\mathbf{f}}(\mathbf{x}) \\ \mathbf{g}(\lambda(\bar{\mathbf{x}}), \mathbf{p}) &= \mathbf{g}(\bar{\mathbf{x}}, \bar{\mathbf{p}}). \end{aligned} \quad (3.22)$$

Unlike the Taylor series method, this approach relies on deriving a map,  $\lambda(\mathbf{x})$ . This can be of similar computational complexity to solving Eq. (3.1) for  $\mathbf{p}$  and  $\bar{\mathbf{p}}$ ; additionally for models containing nonlinear terms, the derivation of  $\lambda(\bar{\mathbf{x}})$  can give multiple solutions, including nested expressions in  $\bar{\mathbf{x}}$  raised to fractional powers.

Evans *et al.* (2005) propose a simplification of this criterion. If Eq. (3.22) holds, not just along the solution trajectories, but for all  $\mathbf{x} \in \mathcal{X}$  where  $\mathcal{X}$  is the state-space of the model, then the model is not structurally globally identifiable. In fact, it is generally structurally globally unidentifiable. This method is particularly interesting in the contexts of the Definitions 2.5 and 2.7. As it gives us a mapping  $\lambda$  taking points of one state-space solution and mapping them onto another. This will allow us to find and visualise state-space solutions that are mapped onto the same output solution.

*Adapting the ORC method for models with step functions as inputs* requires some modification to the notation and concepts introduced previously.

Consider a generic model  $\Sigma$ . As discussed in Remark 2.2, this model can effectively be split into two autonomous models  $\Sigma_a$  and  $\Sigma_d$ , i.e. one for either phase. with the statespace equations  $\mathbf{f}_a(\mathbf{x}(t; \mathbf{p}), \mathbf{p}) = \mathbf{f}(\mathbf{x}(t; \mathbf{p}), \mathbf{I}_a, \mathbf{p})$  and  $\mathbf{f}_d(\mathbf{x}(t; \mathbf{p}), \mathbf{p}) = \mathbf{f}(\mathbf{x}(t; \mathbf{p}), \mathbf{I}_d, \mathbf{p})$  respectively.

If the ORC is met at both initial conditions, the method can now be performed on these two models and two new output solution preserving mappings  $\lambda_a$  and  $\lambda_d$  can be obtained.

The solutions to the IVP of the state-space equations and initial condition of  $\Sigma$  are assumed to be continuous. For this to be true;

$$\lim_{t \rightarrow t_1} \lambda_a(\mathbf{x}(t; \bar{\mathbf{p}})) = \lambda_d(\mathbf{x}(t_1, \bar{\mathbf{p}})), \quad (3.23)$$

and as a result  $\lambda_a(\mathbf{x}(t; \bar{\mathbf{p}})) = \lambda_d(\mathbf{x}(t_1, \bar{\mathbf{p}}))$ . This final result may be thought of as representing the fact that the models  $\Sigma_a$  and  $\Sigma_d$  may both have many to one state-output mappings - but this only implies that  $\Sigma$  has a many to one solution mapping if at the time  $t_1$  these solutions intersect.

It is essential to find a way of *implementing the ORC method* that minimises the computational strain that comes from solving Eq. (3.19) and Eq. (3.22). A key observation that can be made here is that  $\bar{\mathbf{p}} = (\bar{p}_1, \bar{p}_2, \dots, \bar{p}_n)$  is not the only form the alternate parameter vector can take. Alternately, partial solutions from the Taylor series approach such as Eq. (3.15) could be used to reduce the number of parameters requiring consideration.

Villaverde *et al.* (2016) conceptualises parameters and variables differently. In a given model, each parameter  $p_i$  may be treated as a variable with trivial dynamics  $\frac{dp_i}{dt} = 0$ . This is done for all parameters so that identifiability is reduced to being the observability of these new constant parameters, however when we use this with the ORC we may be also

using it with partial solutions from other methods, such as Eq. (3.15) - so there is no need to do this for all parameters. In fact turning all parameters into constant variables needlessly increases the number of Lie-derivatives required for Eq. (3.19) and thus the computational time and memory required to find  $\lambda(\mathbf{x}(t; \mathbf{p}))$ . As a result where models are augmented in this way it will be done for small numbers of parameters. This means that the state variable for these augmented models is:

$$\mathbf{x}(t; \mathbf{p}, I) = (x_1, \dots, x_m, p_{i_1} \dots p_{i_q})^\top \quad (3.24)$$

where  $i_1 \dots i_q$  are the indexes of the parameters being considered as variables. This gives definitions of  $\lambda(\mathbf{x}(t; \mathbf{p}))$  of the form Eq. 3.20 and Eq. 3.21, except that these matrices are now  $(m + q) \times (m + q)$ . Notably Eq. 3.22 and Eq. (3.23) hold for this new definition of  $\lambda(\mathbf{x}(t; \mathbf{p}))$ , but for the augmented model rather than the original.

However, Rather than using them with Eq. 3.22, in this thesis the goal of their use is generally to better understand what makes output solutions indistinguishable. In particular, by obtaining the  $m \times m$  submatrix of  $L(\mathbf{p})$  that maps the original states from and to each other; or the  $m \times m$  submatrices of  $L_1(\mathbf{p}), L_2(\mathbf{p}), \dots L_l(\mathbf{p})$  that do the same. This is done by considering them as point-wise mappings between the output solutions in the fibres of the mapping  $Y_\Sigma$ .

A second way of augmenting models that is useful for the ORC method is to consider observable state variable components as additional outputs. That is if when  $\lambda(\mathbf{x}(t; \mathbf{p}))$  is determined if  $\lambda_i(\mathbf{x}(t; \mathbf{p})) = x_i$  where  $i$  is a valid index of the state variable, then the

ORC method may be repeated, this time using:

$$\begin{aligned} \nu_1(\mathbf{x}, \mathbf{p}) &= x_i, \\ &\vdots \\ \nu_l(\mathbf{x}, \mathbf{p}) &= L_{\mathbf{f}}\nu_{q-1}(\mathbf{x}) = \frac{\partial}{\partial \mathbf{x}}\nu_{q-1}(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}), \end{aligned} \quad (3.25)$$

and Eq. (3.17) to define a new version of  $\mathbf{H}(\mathbf{x}, \mathbf{p})$ :

$$\mathbf{H}(\mathbf{x}, \mathbf{p}) = (\mu_1(\mathbf{x}, \mathbf{p}), \dots, \mu_m(\mathbf{x}, \mathbf{p}), \nu_1(\mathbf{x}, \mathbf{p}), \dots, \nu_q(\mathbf{x}, \mathbf{p}))^T, \quad (3.26)$$

Which can be used to define  $\lambda(\mathbf{x}(t; \mathbf{p}))$  for a model augmented by  $q$  extra parameters being considered as constant variables and  $x_i$  as an additional output. Similarly, this gives definitions of  $\lambda(\mathbf{x}(t; \mathbf{p}))$  of the form Eq. 3.20 and Eq. 3.21, except that these matrices are now  $(m + q) \times (m + q)$ , and Eq. 3.22 and Eq. (3.23) hold for this new definition of  $\lambda(\mathbf{x}(t; \mathbf{p}))$ , but for the augmented model rather than the original.

This means that as well as using the Taylor series approach to obtain partial solutions that we can use to reduce the computational strain implementing the ORC method, we can also iterate it - by first finding observable parameters, and then using it on an augmented version of the model that includes the observable parameters as outputs.

The advantage of using augmented models is that we can use a larger number of derivatives when defining  $\lambda(\mathbf{x}(t; \mathbf{p}))$  and as a result have more simultaneous equations - thus allowing more parameters to be eliminated from the  $m \times m$  submatrices that maps the original states from and to each other. The goal of their use is to produce matrices in which all parameters that can be eliminated from the  $m \times m$  submatrices that map the original states from and to each other are eliminated.

Later on as new methods and tools are introduced it will generally be possible to implement

them together with this method, and reduce the computational bottleneck posed by solving Eq. (3.22). Although this method is hard to implement in a way that avoids computational difficulty it is also enormously valuable because it allows us to find point-wise mappings between the fibres of the state-output mappings  $Y_{\Sigma}^{-1}(y(\cdot, \mathbf{p}, \mathbf{I}))$ , and potentially identify this fibre.

**Example 3.4.** As the ORC method concerns the partial derivatives of the system the **two configuration Langmuir model** is more useful as an illustrative tool than the single configuration model which only has one state variable.

First a matrix of Lie-derivatives of the output is made:

$$\mathbf{H}(\mathbf{x}, \mathbf{p}) = \alpha \begin{pmatrix} x_1 + x_2 \\ -(C_T(k_{a1} + k_{a2}) + k_{d1})x_1 - (C_T(k_{a1} + k_{a2}) + k_{d2})x_2 + R(k_{a1} + k_{a2})) \end{pmatrix} \quad (3.27)$$

The Jacobian of  $\mathbf{H}(x, \mathbf{p})$  is:

$$\frac{d\mathbf{H}(\mathbf{x}, \mathbf{p})}{d\mathbf{x}} = \alpha \begin{pmatrix} 1 & 1 \\ -(C_T(k_{a1} + k_{a2}) + k_{d1}) & -(C_T(k_{a1} + k_{a2}) + k_{d2}) \end{pmatrix}. \quad (3.28)$$

It may be observed that this matrix has a column rank of 2, meaning that it has full rank. Notably this satisfies the ORC. This means we can now use Eq. 3.19 to find a smooth mapping,  $\lambda(\mathbf{x})$ , between points on the indistinguishable state-space solutions of the model.

There are three forms of the alternate parameter vector so far presented that could be considered for the ORC method, Eq. (3.38), Eq. (3.15) and  $\bar{\mathbf{p}} = (\bar{k}_{a1}, \bar{k}_{d1}, \bar{k}_{a2}, \bar{k}_{d2}, \bar{C}_T, \bar{R})$ . In this thesis the smallest partial solution will always be used with the ORC, in this case it is Eq. 3.38. Substituting the two cases of Eq. (3.38) into Eq. (3.19) and solving gives

two forms for  $\lambda(\mathbf{x})$ :

$$\lambda(\mathbf{x}(t, \mathbf{p})) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}(t, \mathbf{p}), \text{ or } \lambda(\mathbf{x}(t, \mathbf{p})) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t, \mathbf{p}). \quad (3.29)$$

If the other alternate parameter vectors were used the same result could be arrived at by considering Eq. 3.22 and Eq. 3.23.

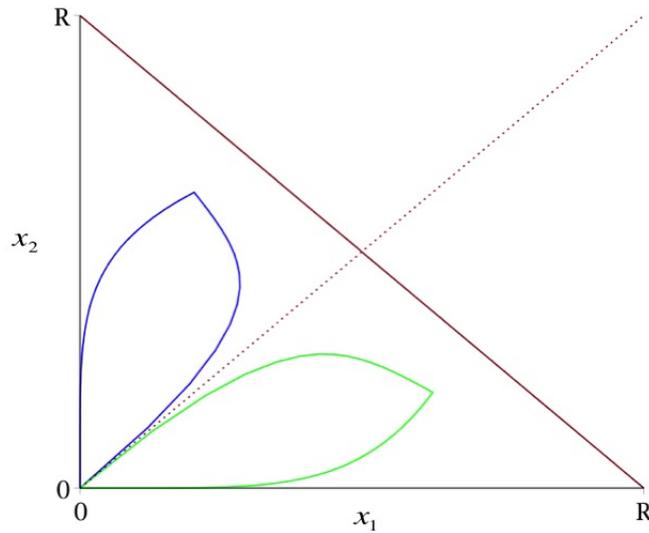
Part of the result of the ORC is a point-wise mapping between indistinguishable solutions we can now visualise the set of state-space solutions that are indistinguishable to a given state-space solution, in the same way as the equivalence class for individual parameters was previously graphed. A set of indistinguishable state-space solutions is shown in 3.6.

The advantage of trying to produce a definition of  $\lambda(\mathbf{x}(t, \mathbf{p}))$  in which all parameters that can be eliminated from the  $m \times m$  submatrices that map the original states from and to each other are eliminated is shown in Eq. 3.29 and Fig. 3.6. In this definition of  $\lambda(\mathbf{x}(t, \mathbf{p}))$ , the relationship between indistinguishable state-space solutions is very clear and as a result allows for Fig. 3.6 to be made. Notably whilst the ERC and similar methods have existed since Hermann & Krener (1977) this is the first time a graph of this kind has been published.

This may now be compared with Fig. 3.3 and Fig. 3.4. Points on the same parameter-space curve are mapped onto the same state-space solution by  $X_{\Pi}$  whereas  $X_{\Pi}$  maps points on the reflected curve to the reflected state-space solution. These two state-space solutions are mapped onto the same output solution by  $Y_{\Sigma}$ .

### 3.1.3 The Output Equation Method

The problems of implementing a general form of the ORC method that could be used on all models illustrates the need for a more robust computational method for analysing



**Figure 3.6:** Graph showing state-space trajectories of the two configuration Langmuir model with two parameter vectors with equal outputs (blue and green) with a dotted red line for  $x_1 = x_2$  and a solid red line for  $R = x_1 + x_2$ . The parameters used in this graph are taken from Table 2.3 and an indistinguishable parameter vector derived with Eq. 3.38

identifiability. A number of methods of this kind for systems with scalar outputs were developed by Forsman (1991).

He developed algorithms in Maple that could be to eliminate the state variables from the equations that define a model, as a result giving a monovariate differential polynomial (MVDP) in terms of the output variable  $y$ . So far in this thesis the phrase output solution, has been used without an explanation of what  $y(\cdot; \mathbf{p})$  is a solution to; it is the solution to this MVDP with an initial condition (for autonomous cases).

In the literature the MVDP is called the output equation, although the equation  $y(t; \mathbf{p}) = g(\mathbf{x}(t; \mathbf{p}))$  is also called the output equation. To avoid confusion the differential equation Forsman's method yields, which  $y(\cdot; \mathbf{p})$  is a solution to, will be exclusively referred to as the MVDP.

Forsman's method begins with considering equations representing the relationship between the Lie derivatives of the output and the differential of the output  $0 = \frac{d^i y}{dt^i} - L_{\mathbf{f}}^i \mathbf{g}(\mathbf{x})$ . If

these expressions have denominators, they are multiplied out, which converts them to multivariate polynomials which are labelled  $P_i(\mathbf{x}, y)$  correspondingly.

Then we are concerned with a polynomial ideal  $I(\mathbf{x}, y^1 \cdots, y^n)$  generated by  $P_i(\mathbf{x}, y^{(i)})$  where  $i \leq n$  where  $n$  is the dimension of the state variable  $\mathbf{x}$ . When a Gröbner basis for this polynomial ideal is found this allows for easy elimination of the state variables. Foresman's methods has minor variation in the way in which the basis is found and how it is used to eliminate the state variables, but they produce the same MVDP. However as their run times vary depending on the model considered, no one of the three is definitively better.

1. Find the contraction of  $I(\mathbf{x}, y, \cdots y^{(n)})$  to the polynomial ring in the intermediate  $y^{(n)}$ , with coefficients in a field extension that includes  $y, y^{(1)}, \dots, y^{(n)}$ .
2. Compute a reduced Gröbner basis for  $I(\mathbf{x}, y^{(n)})$  with a ranking in which the states,  $x_i > y^{(n)}$  for all  $i$ , and the remaining  $y^{(k)}$  are treated as parameters.
3. Compute a reduced Gröbner basis for  $I(\mathbf{x}, y, \cdots y^{(n)})$  with a ranking in which the states,  $x_i > y^{(n)}$  for all  $i$ , and  $y^{(k)}$  is ranked higher than  $y^{(k-1)}$  for  $k = 1, \dots, n$ .

These are executed with pure lexicographical ordering. That is monomials are compared first by their degree intermediate of highest order with ties being broken by the order of each intermediate in turn. That is for the pure lexicographical ordering  $x_1 > x_2$ ,  $x_1^2 x_2^2 > x_1 x_2^2$ .

Either of these methods gives a MVDP, in terms of the output variable and its derivatives, which can be written as:

$$0 = \sum_{\mathbf{v} \in \mathbb{N}^{n+1}} \left( c_{\mathbf{v}}(\mathbf{p}) \prod_{i=0}^n \left( \frac{d^{i-1}}{dt^{i-1}} y(\mathbf{p}, t) \right)^{v_i} \right), \quad (3.30)$$

where  $c_{\mathbf{v}}(\mathbf{p})$  are rational functions of the parameters, and  $v_i$  is the  $i^{th}$  component of the

vector of natural numbers  $\mathbf{v}$ , and  $c_{\mathbf{v}}(\mathbf{p})$  for the lead term is 1.

If we Let  $L_j$  represent the contraction of the ideal  $I(\mathbf{x}, y^{(1)} \cdots, y^{(j)})$  to the polynomial ring in the indeterminates  $y, y^{(1)}, \dots, y^{(j)}$ , and  $n$  is the smallest value for which  $L_j \neq 0$  then Forsman (1991, Theorem 5.3) proves that  $L_n$  is a principal ideal. As a result there are no other fundamentally different equations of minimal order describing the output relation, so the Wronskian does not need to be checked to ensure the independence of the monomial terms. This means that for two parameter vectors of the model,  $\mathbf{p}$  and  $\bar{\mathbf{p}}$ , to have identical outputs:

$$c_{\mathbf{v}}(\mathbf{p}) = c_{\mathbf{v}}(\bar{\mathbf{p}}), \quad (3.31)$$

for all  $\mathbf{v} \in \mathbb{N}^{n+1}$ .

These methods were developed applied in various ways by Meshkat *et al.* (2009) and Evans *et al.* (2013). Notably Evans *et al.* (2013) developed variations of these methods for Maple 2010, and applied them to SPR models with their simple step function inputs. The method used was to iterate Forsman's method on both phases of the experiment to give the relationships between parameters required for the outputs to be identical in either phase.

A novel variation of these methods has been developed to reduce the computational strain of this method. This will be referred to as the **iterated Gröbner method for finding the output MVDP**.

1. Begin with a set of equations including only multivariate polynomials derived as explained above from the output equation and its first derivative  $S = \{P_0(\mathbf{x}, y), P_1(\mathbf{x}, y^{(1)})\}$ . This set generates an ideal,  $I_S(\mathbf{x}, y, y^{(1)})$ .
2. Derive a Gröbner basis  $G$  for the ideal  $I_S(\mathbf{x}, y, y^{(1)})$  using a ranking in which,  $x_i > y^{(n)}$  for all  $i$ , and  $y^{(k)}$  is ranked higher than  $y^{(k-1)}$  for  $k = 1, \dots, n$
3. Whilst  $G$  does not include a non-zero MVDP in  $y$  iteratively:

- (a) take derivatives of all the elements of  $G$  and label the set of these derivatives  $D$ ,
- (b) consider a new ideal generated by the set  $S = G \cup D$ . Label this ideal  $I_S(\mathbf{x}, y, \cdot, y^{(j)})$ ,
- (c) and derive a Gröbner basis  $G$  for the ideal  $I_S(\mathbf{x}, y, \cdot, y^{(j)})$  using the same ranking.

This reduces the computational strain, because after each differentiation the length of the polynomials will be minimised before it is differentiated again, but it derives the same MVDP. This can be seen by a comparison of the methods. Foresmans method constructs an ideal,  $I(\mathbf{x}, y^1 \cdots, y^n)$  from the set  $\{P_0(\mathbf{x}, y^{(0)}) \cdots P_n(\mathbf{x}, y^{(n)})\}$  derived from the Lie-derivatives of the models output. It then derives a Gröbner basis for the ideal. The MVDP is an element of this ideal, it is the element of the Gröbner basis with the lowest monomial order.

Notably the ideal is closed under the derivation  $\delta$ , where:

$$\begin{aligned} \delta(x_i) &= f_i(\mathbf{x}, \mathbf{p}, \mathbf{I}) \\ \delta(I) &= 0 \\ \delta(y^{(i)}) &= \begin{cases} y^{(i+1)} : i < n \\ 0 : i = n, \end{cases} \end{aligned} \tag{3.32}$$

where  $n$  is the dimension of the state-variable,  $i = 1 \cdot n$  and  $f_i(\mathbf{x})$  is the  $i^{th}$  component of the right hand side of the state space-space equation with parameters  $p$  and input  $I$  as defined in Definition 2.4. When we add the derivation, the ideal becomes a differential ideal, which we write as  $I\{\mathbf{x}, y^{(1)} \cdots, y^{(n)}\}$ . Foresmans method can be thought of as a way of navigating through this differential ideal to find a specific element.

The iterated Gröbner method changes the order in which differential and the binary operations of the ideal are used to do this navigation, but searches for the same element, the element of the Gröbner basis with the lowest monomial order, the MVDP.

**Example 3.5.** The **single configuration Langmuir model**, Eq. (2.6), again is not complicated enough to properly illustrate this method. As its output variable is a constant multiple of its state variable, its state variable may be easily replaced with its output variable without need for Gröbner basis. Equation (2.6) is substituted into Eq. (2.5) to eliminate the state variable giving:

$$\frac{dy}{dt} = \alpha k_a IR - (k_a I + k_d)y, \quad (3.33)$$

and by considering this in either phase ( $I = C_T$  and  $I = 0$ ) a full analysis may be done.

**Example 3.6.** The **two configuration Langmuir model** is analysed. As  $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}_{>0}^2$ , the output equation and its first two Lie-derivatives will be needed to obtain the output MVDP for a general point in the association phase. These are:

$$\begin{aligned} 0 &= y_0 - \alpha(x_1 + x_2), \\ 0 &= y_1 - L_{\mathbf{f}}^1 \mathbf{g}(\mathbf{x}) = y_1 - \mu_{11}(\mathbf{p})x_1 - \mu_{12}(\mathbf{p})x_2 - \mu_{10}(\mathbf{p}), \\ 0 &= y_2 - L_{\mathbf{f}}^2 \mathbf{g}(\mathbf{x}) = y_2 - \mu_{21}(\mathbf{p})x_1 - \mu_{22}(\mathbf{p})x_2 - \mu_{20}(\mathbf{p}), \end{aligned} \quad (3.34)$$

where  $\mu_{ij}(\mathbf{p})$  is either the coefficient of  $x_j$  or if  $j = 0$  the constant term from the multivariate polynomial  $L_{\mathbf{f}}^j \mathbf{g}(\mathbf{x})$  the  $j^{th}$  derivative of the output. Notably as observed earlier, for this model,  $L_{\mathbf{f}}^i \mathbf{g}(\mathbf{x})$  is always linear in  $x_1$  and  $x_2$ . This notation is used here because of the length of these expressions - the first Lie-derivative earlier appeared as the second element of the vector  $\mathbf{H}$  in Eq. (3.27).

As these expressions, Eq. (3.34), are polynomial rather than rational functions there are no denominators that need to be multiplied out. So they can be labeled  $P_0$ ,  $P_1$ , and  $P_2$  respectively. We are now concerned with a polynomial ideal  $I(\mathbf{x}, y)$  containing the elements  $P_0$ ,  $P_1$ , and  $P_2$ .

For illustrative purposes the more complex of Forsman's methods will be used. The monomial ordering  $x_2 > x_1 > y_2 > y_1 > y_0$  is used. Maple's internal implementation eliminates each of  $x_2$  and  $x_1$  until what remains is the MVDP in terms of  $y_2$ ,  $y_1$ ,  $y_0$ . This MVDP is short enough to be printed in full:

$$\begin{aligned} 0 = & y_2 + (C_T k_{a1} + C_T k_{a2} + k_{d1} + k_{d2})y_1 + (C_T k_{a1} k_{d2} + C_T k_{a2} k_{d1} + k_{d1} k_{d2})y_0 \\ & - (R\alpha C_T k_{a1} k_{d2} + R\alpha C_T k_{a2} k_{d1}). \end{aligned} \quad (3.35)$$

Similarly a second MVDP can be generated for a general point in the dissociation phase:

$$0 = k_{d1} k_{d2} y_0 + (k_{d1} + k_{d2})y_1 + y_2. \quad (3.36)$$

The monomial coefficients can be used to generate a system of equations that need to be conserved for two parameter vectors to have the same output MVDP. For the parameter vectors  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R)$ ,  $\bar{\mathbf{p}} = (\bar{k}_{a1}, \bar{k}_{d1}, \bar{k}_{a2}, \bar{k}_{d2}, \bar{C}_T, \bar{R})$  these equations are:

$$\begin{aligned} C_T k_{a1} k_{d2} + C_T k_{a2} k_{d1} + k_{d1} k_{d2} &= \bar{C}_T \bar{k}_{a1} \bar{k}_{d2} + \bar{C}_T \bar{k}_{a2} \bar{k}_{d1} + \bar{k}_{d1} \bar{k}_{d2} \\ C_T k_{a1} + C_T k_{a2} + k_{d1} + k_{d2} &= \bar{C}_T \bar{k}_{a1} + \bar{C}_T \bar{k}_{a2} + \bar{k}_{d1} + \bar{k}_{d2} \\ R\alpha C_T k_{a1} k_{d2} + R\alpha C_T k_{a2} k_{d1} &= \bar{R}\alpha \bar{C}_T \bar{k}_{a1} \bar{k}_{d2} + \bar{R}\alpha \bar{C}_T \bar{k}_{a2} \bar{k}_{d1} \\ k_{d1} k_{d2} &= \bar{k}_{d1} \bar{k}_{d2} \\ k_{d1} + k_{d2} &= \bar{k}_{d1} + \bar{k}_{d2}. \end{aligned} \quad (3.37)$$

These equations have the familiar solution. For  $\bar{\mathbf{p}} \sim \mathbf{p}$ , where  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R)$ :

$$\begin{aligned} \bar{\mathbf{p}} &= (k_{a1}/\gamma, k_{d1}, k_{a2}/\gamma, k_{d2}, C_T\gamma, R) \\ &\text{or } (k_{a2}/\gamma, k_{d2}, k_{a1}/\gamma, k_{d1}, C_T\gamma, R) \end{aligned} \quad (3.38)$$

for  $\gamma = \bar{C}_T/C_T$ . Code for this example is shown in the appendix, in Subsection A.3.4

**Example 3.7.** Neither of the Langmuir models were complicated enough to make interesting examples for the Langmuir model. So it is useful to create a toy (autonomous) model specifically to demonstrate the iterated Gröbner method:

$$y = \sum_{i=2}^n x_i \quad (3.39)$$

$$\frac{dx_1}{dt} = p_1 \prod_{i=1}^n x_i \quad (3.40)$$

$$\frac{dx_i}{dt} = p_i x_{i-1} \text{ for } i > 1 \quad (3.41)$$

for some  $n > 2$ .

It is clear that taking repeated derivatives of any expression containing  $x_i$  will give expressions of rapidly increasing order and length.

We only take derivatives to obtain independent equations, that we can use to derive a MVDP in  $y$ . We obtain these by taking derivatives of any polynomial function of  $y$  and  $\frac{dy}{dt}$  that is independent of  $y$ , so we can choose one that eliminates  $x_1$  and its derivatives.

If we take repeated derivatives it is more convenient not just to eliminate  $x_1$  from any equation we are working with but the other state variables in the order  $2, 3, 4, \dots, n$ , before we take derivatives, to minimise the degree of the polynomials we are considering.

When a sufficient number of independent equations have been derived, ensuring that they

are of low order reduces the complexity of solving them.

Its easy to see that taking these simple steps to ensure that the problem doesn't become unworkable is equivalent to iteratively replacing the polynomials available with a Gröbner basis for the ideal they would generate, where the Gröbner basis orders the terms  $x_1 > x_2 > \dots > x_n$ . As high order polynomials produce the greatest problem we would also use total degree ordering.

The three key methods from the literature have now been introduced. The two methods that had not previously been adapted for step functions as inputs have been adapted. These three methods and any steps needed to deal with the step functions as inputs have been illustrated with examples.

Writing about these methods together illustrates their similarities. They all involve taking Lie-derivatives of the outputs, which are evaluated at either a general point, or the initial conditions of the phases; and algebraic operations are performed on these Lie-derivatives. In some cases there are choices about which order these algebraic operations need to be performed in. But generally these methods have computational bottlenecks that prevent the analysis of more complex models with statespace equations including rational functions and monomial terms that are nonlinear in the state variables.

## 3.2 Novel tools for analysing Identifiability

The methods previously discussed, with their adaptations are theoretically sufficient to analyse any model presented in this thesis, however in many cases this requires computational resources beyond those currently available at Warwick University. It is therefore necessary to create new tools to reduce the computational strain and allow for these analyses.

There will be four such tools introduced in this section. The first two create models

that have similar input-output relationships but are easier to analyse, the third exploits properties of the models that emerge from their having step functions as inputs to create objects similar to the Lie-derivatives, but that have algebraic properties that make the computations done with them easier to process.

However before these tools can be introduced some new definitions need to be made.

**Definition 3.14.** *An **output parameter** of a model is a parameter occurring exclusively in the output function of the model.*

**Definition 3.15.** *An output function for a model is **simple** if it is a non zero, linear, vector valued function exclusively of the output parameters and state variables.*

This definition would include all the systems considered in this thesis as well as most systems outside of it. This definition is useful outside of this thesis, firstly because for many biological or ecological systems that we might wish to study we may have some choice about the structure of the output equation but not the underlying system. Generally the output functions we can choose from will be vector valued or scalar valued linear function of the state variables.

This definition allows for making new definitions related to the identifiability of parameters.

**Definition 3.16.** *A non-output parameter or rational function of parameters is **simply unidentifiable** if there is no simple output, given the state space equations for which the parameter is locally identifiable.*

This is equivalent to the non-output parameter being unidentifiable when its state-output map  $Y_{\Sigma}(\cdot, \cdot)$  is bijective. If a model contains a parameter that is simply unidentifiable its

parameter-state map  $X_{\text{III}}(\cdot)$ , is not bijective. This is illustrated by Fig. 3.4, if the L2 model is broken down to a system and output, there is no simple output that can uniquely determined the parameter vector, because for a generic input multiple parameter vectors give the same state-space solution.

**Definition 3.17.** *The parameter vectors  $\mathbf{p}, \bar{\mathbf{p}} \in \Omega$  are **simply indistinguishable** if and only if  $x(t; \mathbf{p}, \mathbf{I}) = x(t; \bar{\mathbf{p}}, \mathbf{I}), \forall t \in T$ , this is written as  $\mathbf{p} \underset{s}{\sim} \bar{\mathbf{p}}$ , and associated with the equivalence class  $E(\underset{s}{\sim}, \mathbf{p}) = \{\bar{\mathbf{p}} \in \Omega : \mathbf{p} \underset{s}{\sim} \bar{\mathbf{p}}\}$ .*

This definition is significant because, in some cases, a property that we may wish to measure may not be measurable even if we have complete knowledge of the experimental system. As a result, we may need to put together results from multiple experimental systems in which the same property arises.

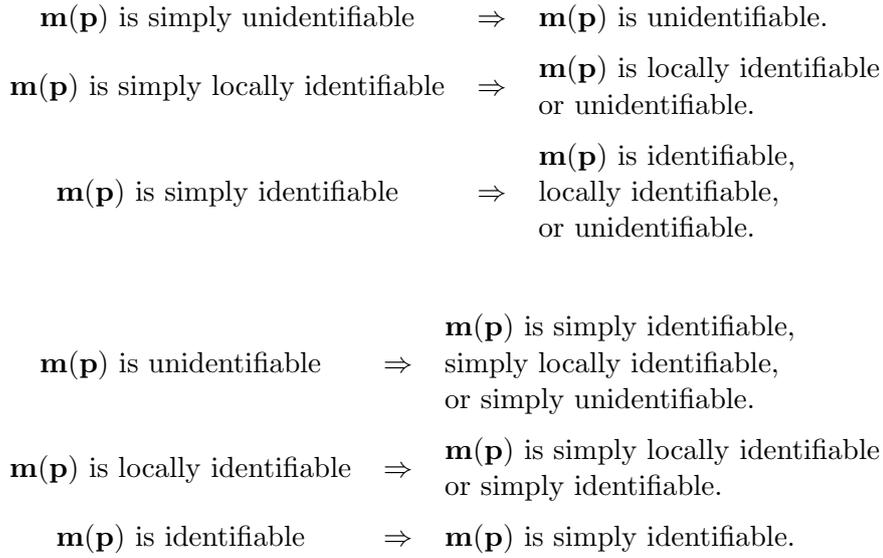
We may note because simple indistinguishability implies indistinguishability that:

$$E(\underset{s}{\sim}, \mathbf{p}) \subseteq E(\sim, \mathbf{p}). \quad (3.42)$$

Whilst a simple indistinguishability analysis hasn't yet been presented for the L2 model this makes sense of Fig. 3.3 - a subset of the parameters that are mapped onto the same output are mapped onto the same state-space solution.

**Definition 3.18.** *A parameter or rational function of parameters is **simply locally identifiable** if there is a simple output which, given the state space equations, makes the parameter locally identifiable*

**Definition 3.19.** *A parameter or rational function of parameters is **simply globally identifiable** if there is a simple output which, given the state space equations, makes the*



**Figure 3.7:** The relationship between simple identifiability and identifiability for  $\mathbf{m}(\mathbf{p})$ , a rational function of parameters.

*parameter globally identifiable.*

The relationship between simple identifiability and identifiability for a rational function of parameters is shown in Fig. 3.7. If identifiability is defined as better than local identifiability, and local identifiability defined as better than unmentionability, then these relationships may be summarised as the simple identifiability of a parameter or rational function of parameters is at least as good as its identifiability.

The statement that a parameter is simply unidentifiable implies that is also unidentifiable, so the statement is stronger. However the statement that a parameter is simply locally or simply globally identifiable is weaker than the statement that it is locally or globally identifiable. As a result, in this thesis when a parameter is simply unidentifiable it will be specified, but when it is simply locally or simply globally identifiable it generally will not.

Each of these new concepts may be thought of as a generalisation of a concept relating to a parameter of a *model* so that they can be used to discuss a parameter of the underlying *system*. Continuing this theme, we make definitions that extend the concepts of structural identifiability for models.

**Definition 3.20.** *A system is **simply structurally unidentifiable** if it has a parameter that is not locally identifiable, for any simple output. A model is simply structurally unidentifiable if it contains a simply structurally unidentifiable system.*

**Definition 3.21.** *A system is **simply structurally locally identifiable** if there is a simple output that would make all of its parameters locally identifiable. A model is simply structurally locally identifiable if it contains a simply structurally locally identifiable system.*

**Definition 3.22.** *A system is **simply structurally globally identifiable** if there is a simple output that would make all of its parameters globally identifiable. A model is simply structurally globally identifiable if it contains a simply structurally globally identifiable system.*

The relationship between simple structural identifiability and structural identifiability for a model is shown in Fig. 3.8. If structural identifiability is defined as better than structural local identifiability, and structural local identifiability defined as better than structural unmentionability, then these relationships may be summarised as the simple structural identifiability of a model is at least as good as its structural identifiability.

Of these three definitions, the first is the most useful - because if a model is simply structurally unidentifiable it is also structurally unidentifiable. Whereas a model being simply structurally locally or globally identifiable does not imply it is structurally locally

$\Sigma$ is simply structurally unidentifiable	$\Rightarrow$	$\Sigma$ is structurally unidentifiable.
$\Sigma$ is simply structurally locally identifiable	$\Rightarrow$	$\Sigma$ is locally structurally identifiable or structurally unidentifiable.
$\Sigma$ is simply structurally identifiable	$\Rightarrow$	$\Sigma$ is structurally identifiable, structurally locally identifiable, or structurally unidentifiable.
$\Sigma$ is structurally unidentifiable	$\Rightarrow$	$\Sigma$ is simply structurally identifiable, simply structurally locally identifiable, or simply structurally unidentifiable.
$\Sigma$ is structurally locally identifiable	$\Rightarrow$	$\Sigma$ is simply structurally locally identifiable or simply structurally identifiable.
$\Sigma$ is structurally identifiable	$\Rightarrow$	$\Sigma$ is simply structurally identifiable.

**Figure 3.8:** The relationship between simple structural identifiability and structural identifiability for a model  $\Sigma$ .

or globally identifiable.

These definitions have an interesting parallel with the concept of minimal output sets in Anguelova *et al.* (2012). In this paper the authors consider the question of what outputs a model needs to be identifiable given a system, or whether all such models are unidentifiable for the feasible measurement set.

The distinction this work has with that is that, whilst we consider a similar problem - that of making decisions about an output, in this thesis, these concepts will have an application to identifiability, particularly results like Eq. 3.42 will help us determine when an analysis is complete.

**Definition 3.23.** A model is *agreeable* if every parameter that is *SU*, *SLI* or *SIGI* remains *SU*, *SLI* or *SIGI* if the output is replaced with the state vector  $\mathbf{x}$ .

This, a model being agreeable, is equivalent to its solution to state-output map  $Y_{\Sigma}$  being bijective.

### 3.2.1 Simple identifiability analysis

If a model is simply structurally unidentifiable, it is also structurally unidentifiable. This is significant, because when the preceding methods are used to determine the structural identifiability of a model much of the computational strain comes from the fact that the state at a general time point may or may not necessarily be known given the input and output function of the model.

When we consider simple identifiability we eliminate this strain, because we are not asking whether a parameter can be identified given a specific output, but any simple (linear) output. As a result we can always consider a given system, not with the output equation it has in a model, but with the output equation  $\mathbf{y}(t) = \mathbf{x}(t, \mathbf{p})$  (See Definition 2.6).

When Fig. 3.3 was presented it was noted that points on individual parameter-space curves it shows are mapped onto the same state-space solution by  $X_{\Pi}$  whereas points on the reflected curve are mapped onto the same output solution by  $Y_{\Sigma}$ . When we analyse simple identifiability we are looking for the regions of parameter-space that  $X_{\Pi}$  maps onto the same solution, the equivalence class  $E(\sim_s, \mathbf{p}) = \{\bar{\mathbf{p}} \in \Omega : \mathbf{p} \sim_s \bar{\mathbf{p}}\}$ .

The state-space functions in the models used in this thesis are rational functions, as a result the state-space equations may be rewritten as polynomials in terms of the state variables, their derivatives and their parameters; i.e:

$$0 = f_{ui}(\mathbf{p}, \mathbf{x}) - f_{vi}(\mathbf{p}, \mathbf{x}) \frac{dx_i}{dt}, \quad (3.43)$$

where  $i$  is an index of a state variable and  $f_{ui}(\mathbf{p}, \mathbf{x})$  and  $f_{vi}(\mathbf{p}, \mathbf{x})$  are the numerator and denominator of the  $i^{th}$  state space function. Using the output equation  $\mathbf{y}(t) = \mathbf{x}(t, \mathbf{p}, \mathbf{I})$

(See Definition 2.6), we may now derive polynomials in the derivatives of the output, of the form:

$$0 = f_{ui}(\mathbf{y}, \mathbf{p}, \mathbf{I}) - f_{vi}(\mathbf{y}, \mathbf{p}, \mathbf{I}) \frac{dx_i}{dt}. \quad (3.44)$$

These differential polynomials can be rapidly analysed. The coefficients of their lead terms can be divided through and their Wronskian evaluated to test for the linear independence of their monomials. If their monomials are linearly independent than for two parameter vectors to have the same output, their linear terms need to have equal coefficients.

As simple identifiability is different to identifiability, the applications of doing this analysis are different.

Firstly, because the analysis is so rapid, and not dependant on computational resources, it can be done as a preliminary to an identifiability analysis- allowing for the recognition of parameters for which the underlying system is a barrier to the identification of. In particular, because of its rapidness it can be applied to numerous models and numerous experimental systems so that ones in which there is no simple identifiability barrier for parameter identification can have their identifiability analysed.

Secondly, it allows for the analysis of systems without outputs. As a result, systems where the experimenter can decide the output can be analysed - and it can be determined if given a simple output the parameters the experimenter are interested in can be identified. The experimenter can now design outputs, and conduct structural identifiability analyses of the models aware of whether parameter unidentifiability comes from the system or the output. This benefit is significantly greater for designers and manufacturers of experimental equipment - who need it to suit the needs of many experimentalists.

Thirdly for some models it allows us to decrease the number of parameters required by the Taylor series approach. Margaria *et al.* (2001) derived that for a model of the kind defined

in Def. 2.6 with a state variable  $\mathbf{x}(t; \mathbf{p}) \in \mathbb{R}_{\geq 0}^m$ ,  $y(t) \in \mathbb{R}_{\geq 0}$ , a parameter vector  $\mathbf{p} \in \mathbb{R}_{> 0}^n$ , then the number of Taylor series coefficients needing to be considered is the first  $m + n$ . If using a simple identifiability analysis we may discover that some parameters may be merged - and still allow for the same set of state-space solutions, we can correspondingly use this new value of  $n$ .

**Example 3.8.** This method can be rapidly implemented on the two configuration Langmuir model. The statespace equation is in the association phase

$$\frac{dx_1}{dt} = k_{a1}C_T(R - x_1 - x_2) - k_{d1}x_1 \quad (3.45)$$

$$\frac{dx_2}{dt} = k_{a2}C_T(R - x_1 - x_2) - k_{d2}x_2. \quad (3.46)$$

The coefficients of the monomial terms are  $k_{a1}C_TR$ ,  $-k_{a1}C_T - k_{d1}$ ,  $-k_{a1}C_T$ ; and  $k_{a2}C_TR$ ,  $-k_{a2}C_T - k_{d2}$ ,  $-k_{a2}C_T$ . The terms for the dissociation phase are  $k_{d1}$  and  $k_{d2}$ . Thus for the parameter vectors  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R)$ ,  $\bar{\mathbf{p}} = (\bar{k}_{a1}, \bar{k}_{d1}, \bar{k}_{a2}, \bar{k}_{d2}, \bar{C}_T, \bar{R})$ ,  $\mathbf{p} \sim_s \bar{\mathbf{p}}$  necessitates:

$$\bar{\mathbf{p}} = (k_{a1}/\gamma, k_{d1}, k_{a2}/\gamma, k_{d2}, C_T\gamma, R) \quad (3.47)$$

for  $\gamma = \bar{C}_T/C_T$ . This now allows us to analyse Fig. 3.3. The indistinguishability equivalence class can be neatly divided into two simple indistinguishability equivalence classes - the two curves in parameter space.

### 3.2.2 Identifiability for functions of parameters and parallel experiments

The above discussion of the identifiability of the monomial terms of the system equations motivates the idea that that it might be useful to discuss the identifiability not of parameters, but functions of parameters.

Whilst this is presented as a novel idea it is essentially a formalisation of an idea that had previously existed implicitly in the literature.

**Definition 3.24.** For  $\mathbf{p} \in \Omega$ , a rational function of parameters  $Q(\mathbf{p})$  is **globally identifiable** if and only if  $\mathbf{p} \sim \bar{\mathbf{p}} \Rightarrow Q(\bar{\mathbf{p}}) = Q(\mathbf{p})$ ; **locally identifiable** if and only if there exists a neighbourhood  $N(\mathbf{p}) \subset \Omega : \bar{\mathbf{p}} \in N(\mathbf{p}), \mathbf{p} \sim \bar{\mathbf{p}} \Rightarrow Q(\bar{\mathbf{p}}) = Q(\mathbf{p})$ ; **unidentifiable** if and only if  $\nexists N(\mathbf{p}) \subset \Omega : \bar{\mathbf{p}} \in N(\mathbf{p}), \mathbf{p} \sim \bar{\mathbf{p}} \Rightarrow Q(\bar{\mathbf{p}}) = Q(\mathbf{p})$ .

An example of these rational functions is  $k_{a1}C_T$  and  $k_{a2}C_T$  for the L2 model 3.38.

This idea is useful because it allows us to rethink what we do when we consider equations like Eq. (3.37). The Taylor series approach and the output equation method both end with the derivation of a set of rational functions of the parameters  $\{Q(\mathbf{p})\}$  that need to have the same value for two parameter vectors to give the same output for a given input, then assuming  $\bar{\mathbf{p}} \sim \mathbf{p}$  we turn these rational functions into symmetrical equations in the components of  $\bar{\mathbf{p}}$  and  $\mathbf{p}$ . Arithmetic operations are then used to solve these for each of the parameters - to determine its identifiability. Implicit in this is the concept of an identifiability ideal; an ideal within the ring of polynomial functions of parameters and their inverses, containing these sets  $\{Q(\mathbf{p})\}$  on which these arithmetic operations are done.

**Definition 3.25.** The globally identifiable ideal of a model  $I_\sigma(\mathbf{p})$  is the smallest ideal containing all globally identifiable functions of parameters, and is a subring of the ring of polynomial functions of parameters and their inverses.

This idea comes from Margaria *et al.* (2001) who considers Gröbner bases of these functions of parameters - implicitly treating them as an ideal. It is useful, because it gives us a tool for considering the identifiability of SPR experiments.

SPR experiments are conducted at many independent interaction spots simultaneously (see Section 1.4), each with its own output. This raises the question of whether this can improve the identifiability of the parameters. In the case of these parallel experiments, the state variables at each spot are unaffected by the state variables at other reaction spots, therefore each spot may be treated as an individual experiment, and may be analysed separately.

Notably SPR experiments featuring the same interaction being conducted at multiple interaction spots, with different dilutions of the sample analyte and different ligand densities, vary from the parallel experiments conducted in pharmacokinetics for which there is existing literature on parallel experiments (Nelson E, 1959; Moghadamnia *et al.*, 2003; Cheung *et al.*, 2013). This variance is due to the fact that in these kinds of experiments more complicated inputs can be used. In pharmacokinetic experiments, it is possible for the experimentalist to vary drug dosage, and timing, and these may be thought of as controls, whereas in our experiments the concentrations of the chemicals and the start and end times of the phases are fixed from the start of the experiment.

As a result whilst the concepts discussed here have some similarity- although they are much simpler.

There are two common ways in which SPR experimental set up varies between interaction spots in parallel. Either concentration of analyte is changed (by dilution) or the amount of antigen is changed, or both.

The effect of this may be considered by considering a set of simple rational functions of the parameters  $m_1(\mathbf{p}), m_2(\mathbf{p}), \dots, m_n(\mathbf{p})$  where both the numerator and denominator have only monomial terms. This set is what we might generate by applying one of the methods for analysing structural identifiability or simple structural identifiability to any of our models. These are the generators of  $I_\sigma(\mathbf{p})$ .

If we alter a parameter so it differs between interaction spots (e.g. by diluting the concentration or using less antigen) then consider the identifiability of an experiment only including output from reaction spot with the altered parameter,  $m_1(\mathbf{p}'), m_2(\mathbf{p}'), \dots, m_n(\mathbf{p}')$  where  $\mathbf{p}'$  is  $\mathbf{p}$  but with  $p_i$  replaced by  $d_i p_i$ . These are the generators of  $I_\sigma(\mathbf{p}')$ .

We can now see that the identifiability ideal for the experiment with outputs from both interaction spots,  $I_2(\mathbf{p}, d_i)$ , will have as it's generators:

$$m_1(\mathbf{p}), m_2(\mathbf{p}), \dots, m_n(\mathbf{p}), m_1(\mathbf{p}'), m_2(\mathbf{p}'), \dots, m_n(\mathbf{p}'),$$

and we can ask under what circumstances this new ideal contains the parameter that was altered at the second interaction spot,  $p_i$ .

To do this we consider Gröbner bases of  $I_\sigma(\mathbf{p})$  and  $I_\sigma(\mathbf{p}')$ , under lexicographic ordering in the parameters and  $d_i$ . As each ideal is identical- except that the variables are labelled differently, these give near identical Gröbner bases  $b_1(\mathbf{p}), b_2(\mathbf{p}), \dots, b_q(\mathbf{p})$  for  $I_\sigma(\mathbf{p})$  and  $b_1(\mathbf{p}'), b_2(\mathbf{p}'), \dots, b_q(\mathbf{p}')$  for  $I_\sigma(\mathbf{p}')$ . As many of these basis elements will be the same for both ideals, we can think of the new ideal as being generated by the Gröbner bases of  $I_\sigma(\mathbf{p})$  and whichever subset of the Gröbner bases of  $I_\sigma(\mathbf{p}')$  which were effected by replacing  $p_i$  with  $d_i p_i$ .

In the cases previously considered, the L1 and L2 models these minimal order generators containing the parameter were interested in varying were

$$k_a C_T; \tag{3.48}$$

and

$$\begin{aligned} k_{a1}C_T, \\ k_{a2}C_T. \end{aligned} \tag{3.49}$$

For a second interaction spot with a the concentration diluted by a known  $d_i$  these would become

$$d_i k_a C_T \tag{3.50}$$

for the L1 model; and

$$\begin{aligned} d_i k_{a1} C_T, \\ d_i k_{a2} C_T. \end{aligned} \tag{3.51}$$

for the L2 model, and all other elements of the bases would remain the same. As a result the ideal  $I_2(\mathbf{p}, d_i)$ , the identifiability ideal for either experiment across two interaction spots is then has the bases  $b_1(\mathbf{p}'), b_2(\mathbf{p}'), \dots, b_q(\mathbf{p}'), d_i$ , i.e. adding Eq. (3.50) or Eq. (3.51) to the identifiability ideal of either model just allows these elements to be divided by or to give  $d_i$ , and in no other way change the Gröbner basis of the ideal.

This can be generalised. If  $p_i$  only appears in terms like of the form

$$b_j = p_i^r p_k^s \tag{3.52}$$

for  $i, j \in \mathbb{Z}$ , then the ideal  $I_2(\mathbf{p}, d_i)$ , the identifiability ideal for either experiment across two interaction spots is then has the bases  $b_1(\mathbf{p}), b_2(\mathbf{p}), \dots, b_q(\mathbf{p}), d_i^t$  for some  $t \in \mathbb{Z}$  where  $b_1(\mathbf{p}), b_2(\mathbf{p}), \dots, b_q(\mathbf{p})$  are the Gröbner bases of the ideal for the experiment at one spot.

This result confirms the results of the analysis in Evans *et al.* (2013), for the models

analysed in that paper. It will be useful, because in general Eq. (3.52) will be the form that relationships between unidentifiable parameters will take in this thesis.

### 3.2.3 Output limits at the discontinuity

Remark 2.2 introduced the idea of splitting a generic model,  $\Sigma$ , into a separate model for each phase  $\Sigma_a$  and  $\Sigma_d$ . This method begins by considering the relationship between the derivatives of the outputs in either phase, and then uses these relationships to find those that are necessary for two parameter vectors to be indistinguishable.

The observation that motivates this method is that for a model with a step function input, with an output that is smooth at all points except the discontinuities of the input, the output solution can be written as a Taylor series in either phase with its coefficients in terms of the derivative of the output as the discontinuity is approached from either direction i.e.

$$y(t; \mathbf{p}) = \begin{cases} \sum_{n=0}^{\infty} \frac{y_a^{(n)}(t_1)}{n!} (t - t_1)^n & \text{if } t \in T_a \\ \sum_{n=0}^{\infty} \frac{y_d^{(n)}(t_1)}{n!} (t - t_1)^n & \text{if } t \in T_d \end{cases} \quad (3.53)$$

where  $y_a^{(n)} = L_{\mathbf{f}_a}^n g$  and  $y_d^{(n)} = L_{\mathbf{f}_d}^n g$ ; where  $t_1$  is the time point when the step happens,  $\frac{d}{dt} \mathbf{x} = \mathbf{f}_a(\mathbf{x}, \mathbf{p})$  and  $\frac{d}{dt} \mathbf{x} = \mathbf{f}_d(\mathbf{x}, \mathbf{p})$  are the state-space equations of the association and dissociation models.

The output equation method obtains a MVDP for each model in terms of its output variable  $y(t)$ . This method obtains a bivariate differential polynomials (BVDPs) in terms of the derivatives of  $y_a(t_1)$  and  $y_d(t_1)$ . This also reduces the computational strain of this task; because Lie-derivatives are being taken of two different functions,  $y_a$  and  $y_d$ , eliminating the need for higher order derivatives of each one and thus reducing the complexity of the equations being solved.

For readers unsure as to whether the global identifiability of a model can be understood only by considering its outputs rates of change at a single point - this is what the Taylor series approach did before it was adapted for models with step functions as inputs. The difference between this method and the Taylor series approach is that the single point being considered is different, and the operations being done on these rates of change are quite different.

This method is implemented in three steps:

1. Lie-derivatives of the output as  $t \rightarrow t_1$  from either side are generated.
2. A reduced Gröbner basis is found using an ordering that eliminates state -variables before derivatives of the output limits, giving a BVDP.
3. The coefficients of the monomial terms of BVDP are compared to find parameter relationships that need to be conserved for multiple parameter vectors to give the same output.

Each of these steps will require either some justification or explanation or both.

The first of these steps is, although quite convoluted, analogous to the output equation method; Lie-derivatives of  $g(\mathbf{x})$  as  $t \rightarrow t_1$  from either side are generated:

$$\begin{aligned}
 0 &= -y + g(\mathbf{x}) \\
 0 &= -y_{a1} + L_{\mathbf{f}_a} g(\mathbf{x}), \\
 0 &= -y_{d1} + L_{\mathbf{f}_d} g(\mathbf{x}), \\
 0 &= -y_{a2} + L_{\mathbf{f}_a}^2 g(\mathbf{x}), \\
 0 &= -y_{d2} + L_{\mathbf{f}_d}^2 g(\mathbf{x}), \\
 &\vdots
 \end{aligned} \tag{3.54}$$

these are rational. The denominators of each can be multiplied out to make multivariate polynomials  $P_0, P_{a1}, P_{d1}, P_a, P_{d2}, \dots$ . These polynomials form the generating set of a polynomial ideal. A monomial ordering is chosen such that state variables are ranked above the output and its derivatives  $y, y_{a1}, y_{d1}, \dots$  - this means that when a Gröebner basis is found it will include a polynomial excluding the state variables - the BVDP.

In the section 3.1.3 the iterated Gröbner method for finding the output MVDP was suggested. A similar method may be implemented here to save computational power.

Once the BVDP is derived, it can be written in a similar form to the MVDP in Eq. (3.30):

$$0 = \sum_{\mathbf{v}, \mathbf{w} \in \mathbb{N}^{n+1}} \left( c_{\mathbf{v}, \mathbf{w}}(\mathbf{p}) \prod_{i=1}^n \left( \frac{d^{(i-1)}}{dt^{(i-1)}} y_a(\mathbf{p}, t) \right)^{v_i} \left( \frac{d^{(i-1)}}{dt^{(i-1)}} y_d(\mathbf{p}, t) \right)^{w_i} \right), \quad (3.55)$$

where  $c_{\mathbf{v}, \mathbf{w}}(\mathbf{p})$  are real valued constants with values determined by the parameter vector  $\mathbf{p}$ ; and  $v_i$  and  $w_i$  are the  $i^{th}$  component of the vector of natural numbers  $\mathbf{v}$  and  $\mathbf{w}$ ;  $n$  is the dimension of the state variable  $\mathbf{x}$ , and the monomial coefficient for the lead term is 1.

The third step is to take the coefficients of these monomials and to equate them. Similar to the output equation method, this gives a set of equations that can be solved to give the identifiability of each parameter and therefore the structural identifiability of the model. For two parameterisations of the model,  $\mathbf{p}$  and  $\bar{\mathbf{p}}$  these equations are  $c_{\mathbf{v}, \mathbf{w}}(\mathbf{p}) = c_{\mathbf{v}, \mathbf{w}}(\bar{\mathbf{p}})$ , for all pairs  $\mathbf{v}, \mathbf{w} \in \mathbb{N}^{n+1}$ .

Code to implement this is presented in the appendix, in Subsection A.3.5.

In some cases the set of equations are not sufficient to derive all parameter relationships from. As a result this method may have to be implemented in conjunction with one of the other methods.

A second application of taking the Lie-derivatives of  $g(\mathbf{x})$  as  $t \rightarrow t_1$  from either side, Eq. 3.54, is that these new Lie-derivatives can be used in the Taylor series approach. This is convenient because when we equate these Lie-derivatives we get expressions of the form  $m(\mathbf{x}(t_1, \mathbf{p}), \mathbf{p}) = m(\mathbf{x}(t_1, \bar{\mathbf{p}}), \bar{\mathbf{p}})$ . When we solve enough of these expressions we will get relationships between  $\mathbf{p}$  and  $\bar{\mathbf{p}}$ , as well as  $\mathbf{x}(t_1, \mathbf{p})$  and  $\mathbf{x}(t_1, \bar{\mathbf{p}})$ .

This is especially useful because the user of an SPR or BLI machine determines when  $t_1$  is. So any relationships between the variables that are held at that point must be held throughout the association phase. Additionally because of Eq. 3.23 these relationships will generally hold for the dissociation phase too. As a result we get a mapping  $\lambda(\mathbf{x}(t, \mathbf{p}))$  between indistinguishable state-space solutions. Just like the mapping obtained in the ORC, this will show the relationships between solutions that  $Y_\Sigma$  will map to the same output solution.

Methodologically this is not dissimilar to Villaverde *et al.* (2016). The difference, other than the use of limits is that this mapping  $\lambda(\mathbf{x}(t, \mathbf{p}))$  is being used to find the fibres of  $Y_\Sigma$  - like those visualised in Fig. 3.6.

**Example 3.9.** The output limit BVDP for the **two configuration Langmuir Model** can be obtained from the equations:

$$\begin{aligned}
 0 &= y_0 - \alpha(x_1 + x_2), \\
 0 &= y_{a1} - k_{a1}I(R - x_1 - x_2) + k_{d1}x_1 - k_{a2}I(R - x_1 - x_2) + k_{d2}x_2 \\
 0 &= y_{d1} - k_{d1}x_1 - k_{d2}x_2
 \end{aligned} \tag{3.56}$$

at  $t = t_1$ .

These equations are a contrast to Eq. (3.34); whilst the first two of the three equations are the same, the third is dramatically shorter, thus allowing the system to be written in

the standard notation rather than the dramatically abbreviated one previously used.

These three equations are sufficient to eliminate the state variables, giving the output BVDP:

$$0 = C_T(k_{a1} + k_{a2})y_0 + y_{a1} - y_{d1} - C_T R\alpha(k_{a1} + k_{a2}). \quad (3.57)$$

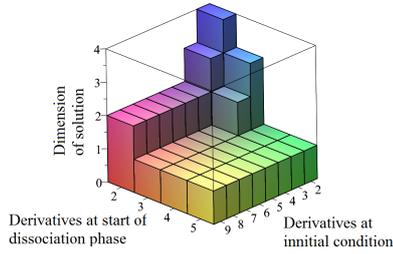
For two parameterisations to have equal outputs, their BVDPs would also have to share their coefficients. This can be seen by making  $y_{a1}$  the subject of the above. This gives us a system of equations, which have the solution

$$\begin{aligned} \bar{C}_T &= C_T \frac{(k_{a1} + k_{a2})}{(\bar{k}_{a1} + \bar{k}_{a2})} \\ \bar{R} &= R. \end{aligned} \quad (3.58)$$

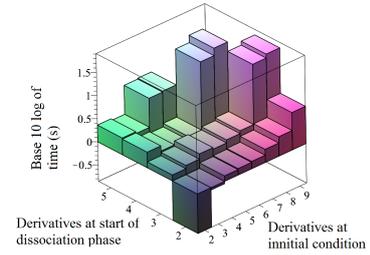
These two equations could either be used to reduce the number of parameters to solve for in either the Taylor series approach, or the ORC method.

**Example 3.10.** In Example 3.3, the length of time required to use the Taylor series to obtain a partial analysis of the L2 model was considered, as well as the dimension of the region of parameter space produced by that partial analysis.

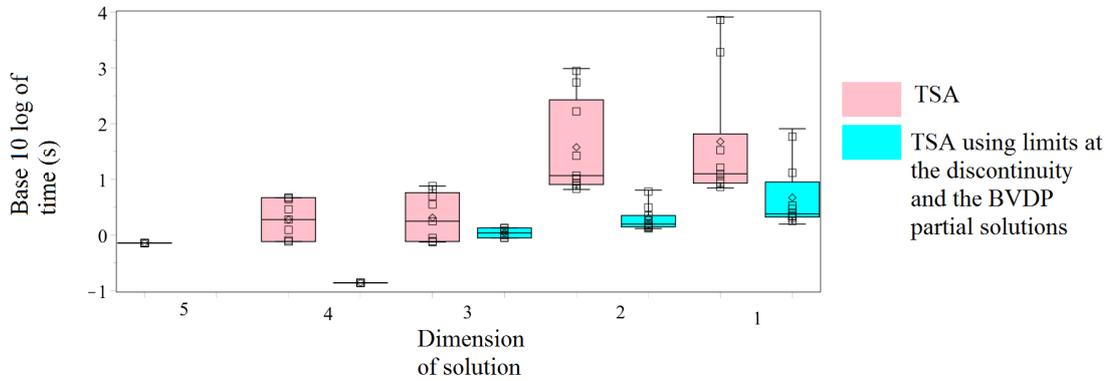
Repeating this analysis, using the partial solutions from the BVDP method, Example 3.9, and the output Lie-derivatives as  $t \rightarrow t_1$  from either side allows us to illustrate the improvements these innovations have made. To avoid extra complexity in diagrams, graphs of the partial identifiability solutions, Fig. 3.9(a), and the timings for their calculation, Fig. 3.9(b), have their number of derivatives at  $t_0$  shown on one axis and the number of derivatives at  $t_1$  approached from either side on another. This means that on this second axis each increase of 1 derivative represents two new equations being added to the system. Fig. 3.9(a) shows how the dimension of the partial solution varies as the number



(a) Dimensions of solution as numbers of derivatives at the start of either phase vary.



(b) Time taken to obtain a solution as numbers of derivatives at the start of either phase vary.



(c) Box-plots showing the distribution of time taken for solve against dimension of solution.

**Figure 3.9:** Graphs showing how the dimension of solution and the time taken to obtain the solution as numbers of derivatives of the output at the start of either phase vary.

of derivatives varies. The key element of this figure is the basin where the dimension of the solution is 1, in the corresponding figure for the Taylor Series approach used alone, this region was  $n_a > 3, n_d > 2, n_a + n_d \geq 8$ ; in this figure it has expanded to  $n_a > 1, n_d > 3$  and  $n_a > 3, n_d > 2$ . In this region the solution is the one depicted in Fig 3.3, i.e. the solution is the equivalence class  $E\{\sim, \mathbf{p}\}$ .

This figure illustrates that when methods are combined in this way, it may be possible to do a complete analysis using fewer derivatives, and derivatives of lower order.

Computational time used is shown in Fig. 3.9(b) and Fig. 3.9(c). The inclusion of

the partial solution and Lie-derivatives of limits reduces computational time by orders of magnitude for the same result. This is a strong recommendation for this methodology for the analysis of similar systems for which computational power is a limiting factor.

### 3.2.4 Combining methods

Example 3.10 illustrates that it is unclear how many derivatives are required to do an analysis when multiple analysis methods are combined. The solution to this problem is illustrated by Fig. 3.4; once simply indistinguishable parameter vector equivalence classes  $E(\sim, \mathbf{p})$ , and indistinguishable state-space solution equivalence classes  $E(\sim, \mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}))$ , then the structure of the mappings  $X_{\Pi \mathbf{I}}$  and  $Y_{\Sigma}$  are understood. This is because these equivalence classes are also the fibres of  $X_{\text{III}}$  and  $Y_{\Sigma}$ , i.e.:

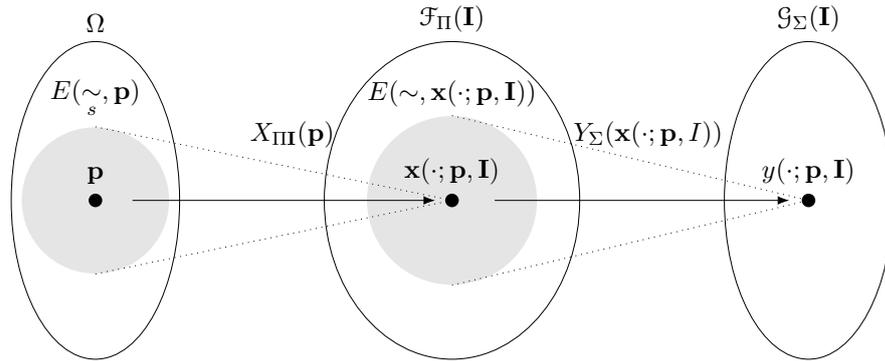
$$\begin{aligned} X_{\text{III}}^{-1}(\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I})) &= E(\sim, \mathbf{p}) \\ Y_{\Sigma}^{-1}(\mathbf{y}(\cdot; \mathbf{p}, \mathbf{I})) &= E(\sim, \mathbf{x}(\cdot; \mathbf{p}, \mathbf{I})) \end{aligned} \quad (3.59)$$

The goal of an identifiability analysis is to understanding the structure of the mapping  $Z_{\Sigma}(\mathbf{p}, \mathbf{I})$ , specifically the values of  $\mathbf{p}$  for which  $Z_{\Sigma}(\mathbf{p}, \mathbf{I})$  is invariant for a generic  $I$ . This is equivalent to finding the fibre of  $Z_{\Sigma \mathbf{I}}(\mathbf{p})$  for  $\mathbf{y}(\cdot; \mathbf{p}, \mathbf{I})$  and this fibre is by definition:

$$\{\bar{\mathbf{p}} : \bar{\mathbf{p}} \in X_{\text{III}}^{-1}(\mathbf{x}), \forall \mathbf{x} \in Y_{\Sigma}^{-1}(\mathbf{y}(\cdot; \mathbf{p}, \mathbf{I}))\}. \quad (3.60)$$

This goal could be achieved in a number of ways. Typically in this thesis the BVDP will be used to obtain some parameter relationships, then the TSA with limits around the discontinuity will be used to obtain further relationships. In most cases these two methods together will give a full result - the fibres of both mappings.

In some cases where the TSA takes to long or causes crashes, some form of the ORC



**Figure 3.10:** Diagram of fibres of  $X_{\text{III}}(\mathbf{p},)$  and  $Y_{\Sigma}(\mathbf{x}(\cdot; \mathbf{p}, \mathbf{I}))$

will be used, often with some kind of augmented system. Analysis of the system will be stopped either when all of the models simple identifiability relationships are shown to be the same as its identifiability relationships, i.e.  $Y_{\Sigma}$  for that model is bijective, equivalently  $\lambda(\mathbf{x}(t, \mathbf{p}))$ , is multiplication by the appropriate identity matrix, or when  $\lambda(\mathbf{x}(t, \mathbf{p}))$  has no dependence on parameters that are simply unidentifiable.

Unlike the numerical results in Margaria *et al.* (2001) and Vajda (1984) this allows for multiple methods to be used together, rather than one.

### 3.3 Structural Indistinguishability

Whilst structural identifiability is concerned with whether *two parameterisations* of a single model with the same general input will give the same output, structural indistinguishability is concerned with whether parameterisations of *two models* exist that, with the same general input, will give the same output. As a result structural identifiability is a special case of structural indistinguishability.

We now state definitions describing the relationship between two models.

**Definition 3.26.** *The models  $\Sigma(\mathbf{p})$  and  $\bar{\Sigma}(\bar{\mathbf{p}})$  are **output indistinguishable**, written*

$\Sigma(\mathbf{p}) \sim \bar{\Sigma}(\bar{\mathbf{p}})$ , if and only if there exists  $\tau > 0$  such that  $y(t; \mathbf{p}, \mathbf{I}(t)) = \bar{y}(t, \bar{\mathbf{p}}, \bar{\mathbf{I}}(t))$  for all  $t \in [0; \tau)$  and every admissible input function  $\mathbf{I}(t)$  (Evans et al., 2004).

If the outputs of both functions are analytic in time for all  $\mathbf{p} \in \Omega$  and  $\bar{\mathbf{p}} \in \bar{\Omega}$ , if there exists a non empty interval in time where  $y(t; \mathbf{p}, \mathbf{I}(t)) = \bar{y}(t, \bar{\mathbf{p}}, \bar{\mathbf{I}}(t))$ , then this relation holds for all time (Evans et al., 2004).

**Definition 3.27.** *The models  $\Sigma(\mathbf{p}, \mathbf{I}(t))$  and  $\bar{\Sigma}(\bar{\mathbf{p}}, \mathbf{I}(t))$  are **structurally indistinguishable**, written  $\Sigma \sim \bar{\Sigma}$ , if and only if for a generic  $\mathbf{p}$  i.e. for all  $\mathbf{p} \in \Omega$  except possibly for some vectors lying in a subset of a closed set of (Lebesgue) measure zero, there exists a  $\bar{\mathbf{p}} \in \bar{\Omega}$  such that  $\Sigma(\mathbf{p}, \mathbf{I}(t)) \sim \bar{\Sigma}(\bar{\mathbf{p}}, \mathbf{I}(t))$ ; and for generic  $\bar{\mathbf{p}} \in \bar{\Omega}$  there exists  $\mathbf{p} \in \Omega$  such that  $\Sigma(\mathbf{p}, \mathbf{I}(t)) \sim \bar{\Sigma}(\bar{\mathbf{p}}, \mathbf{I}(t))$  (Evans et al., 2004).*

In Section 3.2 concepts of identifiability were extended to systems, using the concepts of simple outputs. It is possible to make a similar extension of the concept of structural indistinguishability.

**Definition 3.28.** *The systems  $\Pi$  and  $\bar{\Pi}$  are **simply structurally indistinguishable**, written  $\Pi \underset{s}{\sim} \bar{\Pi}$ , if for any model  $\Sigma(\mathbf{p}, \mathbf{I}(t))$ , where  $\Sigma(\mathbf{p}, \mathbf{I}(t))$  is composed of  $\Pi(\mathbf{p}, \mathbf{I}(t))$  and the simple output  $\mathbf{y} = \mathbf{g}(\mathbf{x})$ ,  $\Sigma \sim \bar{\Sigma}$ , where  $\bar{\Sigma}(\bar{\mathbf{p}}, \mathbf{I}(t))$  is composed of  $\bar{\Pi}(\bar{\mathbf{p}}, \mathbf{I}(t))$  and the same simple output  $\mathbf{y} = \mathbf{g}(\mathbf{x})$ .*

*Two models are simply structurally indistinguishable if they contain simply structurally indistinguishable systems and the same output mapping.*

This concept is going to be very useful, because models containing simply structurally indistinguishable systems and models are much more closely related than structurally indistinguishable systems.

Because of the relationship between structural identifiability and indistinguishability the methods described in the previous section can be adapted to analyse the indistinguishability of two models. However, as the models used in this thesis have inputs that make their outputs continuous and only piecewise smooth, some intuitive additional work has to be done to adapt these methods.

In this thesis, the Taylor series approach will be used predominantly as a test for indistinguishability.

**Example 3.11.** So far two models have been introduced, the single configuration Langmuir model, which will be notated with barred parameters, functions and variables, and the two configuration Langmuir model, which will be notated as before. These two models will now be tested for indistinguishability. For these two models it is easiest to start by considering the Lie derivatives of the output at  $t = t_1$ , the start of the dissociation phase,

$$\begin{aligned} L_{\mathbf{f}}^i g(\mathbf{x}) &= L_{\bar{\mathbf{f}}}^i \bar{g}(\bar{\mathbf{x}}) \\ (-k_{d1})^i x_1 + (-k_{d2})^i x_2 &= (-\bar{k}_d)^i \bar{x} \end{aligned} \tag{3.61}$$

for all  $i \in \mathbb{N}$ . For these equations to hold, one of  $(\bar{k}_d = k_{d1}, \bar{x} = x_1, 0 = k_{d2})$ ,  $(\bar{k}_d = k_{d2}, \bar{x} = x_2, 0 = k_{d1})$ ,  $(\bar{k}_d = k_{d1}, \bar{x} = x_1, 0 = x_2)$  or  $(\bar{k}_d = k_{d2}, \bar{x} = x_2, 0 = x_1)$ , must be true. The first of these cases can be rejected, because they require parameters to take the value of 0. For the second two cases,  $x_1$  or  $x_2$  is required to take a zero value at the start of the dissociation phase. For this, either  $C_T k_{a1} = 0$  or  $C_T k_{a2} = 0$  respectively. As a result these cases can be rejected for the same reason. As a result none of these four cases is allowed by the definition of the models.

This is enough to demonstrate that the models are not structurally indistinguishable. That is for the two configuration model with a generic parameter vector and an input function, there is not a parameter vector for the single concentration model that when combined

with the same input function will give the same output.

This illustrates why in Definition 2.4 it was specified that parameters take non-zero values. In many cases setting a certain parameter to zero will mean that the behaviour of a more complex model becomes identical to that of a simpler model.

### 3.3.1 Limits of models

It was previously noted that as in example 3.11 two models may give the same output for the same input if a certain parameter was set to zero. In some cases the output of two models will converge as the value of a parameter grows unboundedly.

In this thesis the novel concept of one model being an infinite limit of another is introduced.

**Definition 3.29.** For two generic models  $\Sigma, \bar{\Sigma}$ , we call  $\bar{\Sigma}$  the *limit* of  $\Sigma$  as  $p_i \rightarrow q$  if for a generic parameter vector  $\mathbf{p}$ , there exists a parameter vector  $\bar{\mathbf{p}}$  such that:

$$\lim_{p_i \rightarrow q} \mathbf{f}(\mathbf{x}, \mathbf{p}) = \bar{\mathbf{f}}(\mathbf{x}, \bar{\mathbf{p}}),$$

$$\lim_{p_i \rightarrow q} \mathbf{g}(\mathbf{x}, \mathbf{p}) = \bar{\mathbf{g}}(\mathbf{x}, \bar{\mathbf{p}}),$$

$$\lim_{p_i \rightarrow q} \mathbf{x}_0(\mathbf{p}) = \bar{\mathbf{x}}_0(\bar{\mathbf{p}}).$$

where. Extending the familiar notation we write this as:

$$\lim_{p_i \rightarrow q} \Sigma = \bar{\Sigma}$$

This is useful because a parameter vector is defined as  $\mathbf{p} \in \Omega \subset \mathbb{R}_{>0}^n$  for some  $n \in \mathbb{N}$ , where  $\Omega$  is the set of meaningful parameter vectors, and this allows for discussion of cases where  $q$  is just outside of the values  $p_i$  can take. In Example 3.11 we saw that the L1 and

L2 models were distinguishable in the dissociation phase unless one of  $k_{a1}$ ,  $k_{a2}$ ,  $k_{d1}$ ,  $k_{d2}$  or  $C_T$  is 0; notably if  $k_{a1} = 0$  or  $k_{a2} = 0$  is substituted into Eq. 2.10, then it becomes simply a relabeled form of the L1 model, Eq. 2.6. This has an obvious physical interpretation - if an analyte that used to behave with L2 kinetics now binds in one configuration at a rate of  $0M^{-1}s^{-1}$  it now has L1 kinetics.

This is a trivial case, but as the state-space functions we are considering are rational there are also cases where infinite limits of the parameters are well defined. Again, this is useful because it allows us to make physical meaning ful statements about the parameters and models: for example if this rate was so fast as to be instantaneous in comparison to the other rates of the experiment, or if this volume was so large as for the other volumes we are considering to be negligible in comparison to it then this model would behave as that model.

### 3.4 Algorithms and software for model fitting

In addition to determining whether model parameters can be identified, this thesis presents data, models, and fits of models to data, resulting in parameter estimation.

Models are fitted to time series data. This is done by the coupling of two existing methods, to allow for a thorough search of a large parameter space. Differential Evolution is used to search parameter space as a whole for initial values that are passed to FACSIMILE (MCPA Software, UK). FACSIMILE then varies these parameters to find local minima. When such a minimum is found the RSS at this minimum is fed into differential evolution to be recorded alongside the parameters passed to FACSIMILE. Differential evolution then continues its search of parameter space, by passing more initial parameter values to FACSIMILE.

The goal of this is to couple differential evolution as a global search of parameter space,

with FACSIMILE as a local search method. This makes sense because although both methods are effective on their own, FACSIMILE does not search parameter space globally, and differential evolution granualises parameter space so does not work as a local search algorithm.

### 3.4.1 Differential Evolution

Differential evolution is a relatively new algorithm for solving optimisation problems. It considers a population of parameter vectors, iteratively generates new parameter vectors, whilst eliminating the parameters associated with poorer estimates.

In more detail the algorithm is:

1. An initial population of agents  $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$  with random positions in the search-space  $\Omega$  is generated. In this thesis these will represent parameter vectors, and typically,  $n = 150$ .
2. A loop updating this population is iterated until a stopping criterion is reached. In this thesis, the algorithm was generally stopped when it had been run for a certain period of time.
  - (a) In the loop, each time the population is updated a procedure is applied to each member of the population  $\{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_n\}$ .
    - i. The member of the population to be updated is named the target vector  $\mathbf{t}$ .
    - ii. Three other members of the population are picked at random without replacement  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$ .
    - iii. A random element of the target vector is selected,  $t_i$ .
    - iv. A procedure is applied to this element of the target vector, possibly updating it.

- With probability  $d$ , a user set variable called the crossover probability,  $t_i$  is updated to  $a_i + F(b_i - c_i)$  otherwise its value is unchanged.
- (b) if  $Z(\mathbf{p}) > Z(\mathbf{t})$  then the updated target vector  $\mathbf{t}$  replaces the vector  $\mathbf{p}$  in the population. In this thesis the function  $Z(\mathbf{p})$  represents the RSS of the fitting of a specified model with a parameter vector  $\mathbf{p}$  to experimental data.

This algorithm requires one additional thing to be applied to parameter fitting problems. That is, software that is capable of numerical integration that can be used to assess the RSS of a model with a given set of parameters.

There are a number of software suites capable of doing this, but FACSIMILE was chosen.

### 3.4.2 FACSIMILE

Because the models were generally non-linear, with large numbers of state variables, and large numbers of parameters which vary independently across orders of magnitude, fitting them is expected to be a stiff problem. As a result, model fitting and parameter estimation require a software package that is robust, able to deal with stiff problems and, because of the large parameter space, fast.

The commercial software package FACSIMILE (MCPA Software, UK) was chosen because it satisfies these criteria. In numerical integration it combines a backward-difference predictor-corrector method over time steps with varying sizes. As a result it is able to integrate stiff differential equations relatively quickly. To fit the model to data provided it also implements the VA05 routine from the Harwell Subroutine Library (a fortran library for numerical methods). This combines three least squares optimisation methods: the Newton-Raphson, steepest descent and Levenberg-Marquardt (Science & Council, 2015).

The package also contains parameter-fitting options, which allows the user to select parameters to be varied. These are logarithmically varied using the aforementioned VA05 routine to best fit user supplied data.

The main limitation of the FACSIMILE software is that the methods it uses are local optimisation methods. That is, methods that only consider a small region of parameter space around an initial estimate. As a result FACSIMILE is sensitive to the initial guesses and its estimates can be attracted to local minima. It therefore needs numerous wide ranging different combinations of initial guesses.

As a result FACSIMILE and differential evolution can work synergistically. Differential evolution supplies a wide range of initial estimates of parameter vectors, FACSIMILE uses numerical integration to evaluate the model at these estimates, and explores local parameter space, finally determining the RSS for the parameter estimate close to the initial estimate that best fits the data. Differential evolution stores these initial estimates and RSSs, and varies the initial estimates; allowing for a thorough search of parameter space.

### 3.5 Conclusion

In the past two chapters concepts have been introduced, some have been taken from the literature directly, many have been adapted for the specific kinds of problems that will be considered in this thesis, some have been formalisation of ideas already existing in the literature and some have been novel.

These can largely be sorted into three groups, the rigorous concept of a model and related mappings have been presented; novel tools for structural identifiability have been developed; and adaptations of these tools and existing methods for models taking step functions as their inputs.

This rigorous concept of a model and related mappings allows for many of the novel tools of structural identifiability, but also a different kind of discussion of outputs. In many cases a variety of outputs may be chosen for the same underlying system. However, with some underlying systems there are no outputs where a key parameter the investigator is interested in is identifiable. This is illustrated in Example 3.8 where  $k_{a1}$ ,  $k_{a2}$  and  $C_T$  were found to be simply unidentifiable.

Recognising this may be key to a research project. A good example of this may be an ecological research project, where a researcher is interested in the population dynamics of a number of species in the same environment. It may be that no matter how the species are observed some group of parameters like birth or death rates cannot be identified, that is that a parameter is simply unidentifiable or simply locally unidentifiable.

In some cases there may be other environments where these species can be observed separately, or in other groups, as a result model fitting could allow for the estimation of the previously unidentifiable parameters. However, for the researchers to pick which group and which method of observation to use identifiability analysis will have to be performed. The state-space tool will be useful for picking which environment research should be done in, as in some environments the desired parameter may still be simply unidentifiable or simply locally unidentifiable.

Once an environment where there is a simple output function that would make the desired parameter identifiable is found the output function has to be chosen, and as a result more generic identifiability tools can be used.

By analogy there will be similar decisions that need to be made in any other research project that involves investigating parameters of complex systems, such as those found in all kinds of mathematical biology, from pharmo-kinetics to systems biology.

Perhaps the most significant result of this chapter is the demonstration that by considering

the Lie-derivatives of the limits of a models output on either side of the discontinuity can dramatically shorten computational time for the analysis for this model, as shown in Fig. 3.9(c). The time taken for Maple 2015 to determine a solution via either method is summarised in Fig. ???. This illustrates that when the BVDP method is used to achieve a partial solution first, and the generic Taylor series (GTS) is used with limits approaching the initial condition of the dissociation phase from wither side, the time required to find a solution is generally reduced by more than a factor of 10. This is significant because it may allow for the analysis of models that could be analysed otherwise.

Additionally the work on step functions presented here will be useful for any work on autonomous systems that undergo dramatic shifts in parameter values. This could represent anything from a drought reducing the growth rate of herbivores on the African planes to changes to plankton growth rates in response to iron fertilisation, or the removal of a drug fed into a subject via an intravenous drip.

Accurate modelling of all these problems will require identifiability analysis, and could take advantage of the tools and methods presented in this chapter.

# Chapter 4

## Experimental Methods

### 4.1 Goals

The goal of this work is twofold. Firstly to produce methods to reliably estimate kinetic constants of a single species of monoclonal antibody binding against a single species of antigen, for application in the pharmaceutical sciences. Secondly to produce methods that reliably estimate kinetic constants of antibody in clinically derived samples against single species of antigen, for application in transplant medicine.

These two goals are linked in that it is assumed that creating these methods for work with monoclonal antibody will be a stepping stone towards creating equivalent methods for the much more complicated interactions we would expect to see in patient antibody samples.

In Chapter 2 and 3, the concepts of models and systems were developed, and methods of distinguishing between them were presented. Due to the complexity of antibody-antigen interactions mathematical modelling is necessary for any method for estimating these rate constants of these reactions. For this reason, a significant proportion of this thesis will be focused on models and the development of models for SPR experiments (Sections 5.1, 7.1). These models are mathematical formulations of the law of mass action and a variety of assumptions with physical meanings, which are used to predict the possible input-output

behaviour of these experiments.

However to have data to model experiments must be conducted, and experimental procedures developed. The Experiments were conducted with the goal of producing sensorgrams that could support or undermine the assumptions being considered in the modelling process. A model was seen as supported by the data if it produced good fits and consistent estimates of the kinetic parameters.

In some cases, a large number of models had to be developed before one that reliably produced good fits, as well as consistent parameter estimates, was found.

In some cases, further experiments were conducted to test the model and ensure that it would be generally applicable, and not just to the data that it was informed by.

This process of experiments being conducted to allow for model formulation, model formulation, followed by experiments conducted to test the resulting models was enacted with some variation for both reactions of single species of monoclonal antibody against single species of antigen and for clinically derived mixtures of unknown species of polyclonal antibody against single species of antigen.

Experimental data came from a number of sources. Some experiments were designed and conducted specifically for this PhD project (the mixed monoclonal experiments of 4.7), others were designed and conducted collaboratively with Dr Sunil Daga (Sections 4.3-4.4, 4.6 and the serum sample experiments of 4.9) and one set of three sensorgrams was taken from the work of Dr Andrew Bentall (Sections 4.5).

## 4.2 Selecting experimental methods

Other than SPR, there are a number of techniques that can be used to estimate the affinity of a protein- protein interaction. Neri *et al.* (1996) summarised these, and their table is

reproduced in Table 4.1.

The key way that they differ from SPR is that, whereas SPR gives thousands of measurements of the surface concentration of the analyte bound to the reaction chip these make single measurements (for example the amount of analyte that crosses a membrane). As a result, they give a much smaller amount of data for the amount of time and resources spent in conducting them.

A second hurdle is that samples of antibody from patients will be polyclonal. At minimum, antibodies produced by different B Cell lineages will be present in each sample, antibodies from separate lineages will have differences in their hypervariable regions and as a result slightly different paratopes, with different kinetic constants for their interaction with the same antigen. Possibly samples will include antibody with paratopes interacting with other antigen epitopes and multiple antibody isotypes.

These samples may also be unrefined, that is they may also include non-antibody molecules.

In contrast, the techniques of Table 4.1 were designed for determining the affinity of a monoclonal antibody for an antigen. It is unclear how they could be adapted for use with samples of patient antibody.

Two potential adaptations would be to allow them to take material that was not purified or to label individual patient antibody from the sample. In either case, the variety of molecules within each sample would present one set of difficulties, and the variability between patients would present a second set of difficulties.

The two techniques that initially seemed adaptable were competition ELISA (enzyme-linked immunosorbent assay) and SPR equilibrium analysis.

The mathematics of SPR equilibrium analysis are discussed in more detail in Subsection 5.2, but the technique is shown not to improve the identifiability of models. ELISA was judged not to be helpful due to complexity of the labeling process and the potential

Method	Measurement	Sample requirement	Sensitivity
Equilibrium dialysis	Equilibrium across a dialysis membrane, free and bound analyte measured	Small antigen that can cross the dialysis membrane	1 mM to 1 pM
Fluorescence quench	Fractional fluorescence quench upon binding	Pure material needed	1 $\mu$ M to 1nM
Isothermal titration calorimetry	Heat produced on binding of increasing titration of antibody to concentrated antigen solution	Pure material needed in large quantities	1 mM to 100nM Higher affinity can be measured by increasing the temperature
Band-shift on Gel	Detection on free and bound antibody on gel	Radioactive or fluorescent labelling of either antibody or antigen	10 nM to 10pM
Physical separation	Antibody in the complex is separated by chromatography or magnetic beads and detected by fluorescence, or electrochemiluminescence	Labelling of antibody is required	10 nM to 10 pM
Competition ELISA	After incubation of antibody with antigen, the unbound antibody is detected by ELISA	Purified antigen or antibody to immobilise on solid surface	50 nM to 1 nM
SPR Equilibrium analysis	Change in refractive index upon binding is measured	Purified antigen or antibody to immobilise on sensor chip	100mM to 1 nM (kinetic analysis can measure higher affinity)

**Table 4.1:** An overview of the techniques for determining antibody-antigen affinity giving an output at a single timepoint (Adapted from Neiri *et al.* (1996))

interactions of polyclonal antibody and antigen, but became a key part of the motivation for Chapter 9.

### 4.3 Original Monoclonal Experiments

Before the PhD project had started the two configuration versions of the Langmuir and ERC model (Subsections 5.1.1 and 5.1.3) had been developed, and presented in Evans *et al.* (2013). These differed as the former assumed that the concentration of antibody at the surface of the chip is the same as its concentration in the inlet being fed into the machine (as shown in Fig. 1.18 (b)) whereas the later assumed that the concentration of the antibody at the chip was determined by the rate at which antibody moved into the volume of liquid in contact with the surface of the chip (as shown in 1.18 (c)).

Experiments were designed with a wide range of antibody concentrations and antigen densities. The goal of this was to influence the rate at which antibody close to the surface of the chip would bind to the antigen. If this rate was slow enough the difference between the state-space solutions and output solutions predicted by either model would be small.

In the literature, it is noted that the Langmuir model, wrongly identifies the affinity of the analyte by assuming the concentration of the antibody at the surface of the chip is identical to that in the sample, and that this is exacerbated by high sample concentrations. However the rate at which analyte binds is also proportional to the density of available ligand, so conducting the reactions with high ligand density to should have similar exacerbating effects, and low ligand density should reduce these effects. If the novel models presented in Chapter 5 can produce consistent results across a wide range of analyte concentrations and ligand densities it will help illustrate their value for estimating affinity.

The antibodies used were, three human monoclonal HLA-specific antibodies SN607D8, SN230G6 and WK1D12. These were kindly provided by the A Mulder/ F Claas research

Clone (ID)	Immunising Ag	HLA-specificity defined by CDC	Epitope	Isotype	Lot No
SN203G6 (IHB-Hu-033)	A2/B57	A2/B17 (B57,58)	62G	IgG1, L	7403.1
SN607D8 (IHB-Hu-081)	A2/B57	A2/A28 (A68,69)	142T -H	IgG1, K	26789.1
WK1D12 (IHB-Hu-059)	B27	B27,7,60	163EW+self73TE	IgG1, K	7688

**Table 4.2:** Characteristics of IgG human monoclonal HLA-specific antibodies

group at the University of Leiden, cultured as described in Mulder *et al.* (2003);). Although these three experiments were analysed together, there were some differences in their design, as some points of the experimental protocol were sbeing refined.

These three human monoclonal HLA-specific antibodies were secreted by human hybridoma produced by EBV transformation of B-lymphocytes from HLA antibody seropositive multi-parous women and sub-cloning of antibody producing EBV lines. Only human monoclonal antibodies of the IgG class were used.

These monoclonal IgG antibodies were provided as culture supernatants in RPMI / IMDM with 10% FCS, 50 uM 2-ME; 0.02% sodium azide and stored at 4°C. Samples were decanted into dialysis tubing with 8,000 MWCO (Spectra/Por® Dialysis membrane) and incubated with stirring in a cylinder of the desired buffer (e.g. PBS) overnight at 4°C. The antibody were isolated and purified from the supernatants by HLA protein affinity chromatography (McMurtrey *et al.*, 2014;) and Ion exchange chromatography. The purification was confirmed using SDS-PAGE (sodium dodecyl sulphate - polyacrylamide gel) 4 -12% gel electrophoresis and reactivity tested using a Luminex Single antigen bead assay (One Lambda) as per manufacturer’s instruction. The purified monoclonal HLA specific antibodies were quantified using Lowry’s assay.

The details of these antibodies are summarised in Table 4.2.

ProteOn XPR was used to study simultaneous interactions between HLA-protein and

HLA-specific antibodies. Its design allows for six analytes to be each passed over six ligands. In this case, the analytes used were different concentrations of the three human monoclonal antibodies, and the ligands were varying densities of immobilised biotinylated HLA-proteins. These antigens were attached with a 54 Ångstrom linker to the NCL chips surface. This allows them greater freedom of movement, extending the volume in which the antibody could bind to them, and allowing more contact between their epitopes and antibodies in the experiment.

Two different concentrations of A2 HLA-proteins were immobilised (2.5 and 0.25 umg/ml) for reaction with the SN607D8 and SN230G6 antibodies. These are simply termed high and low density.

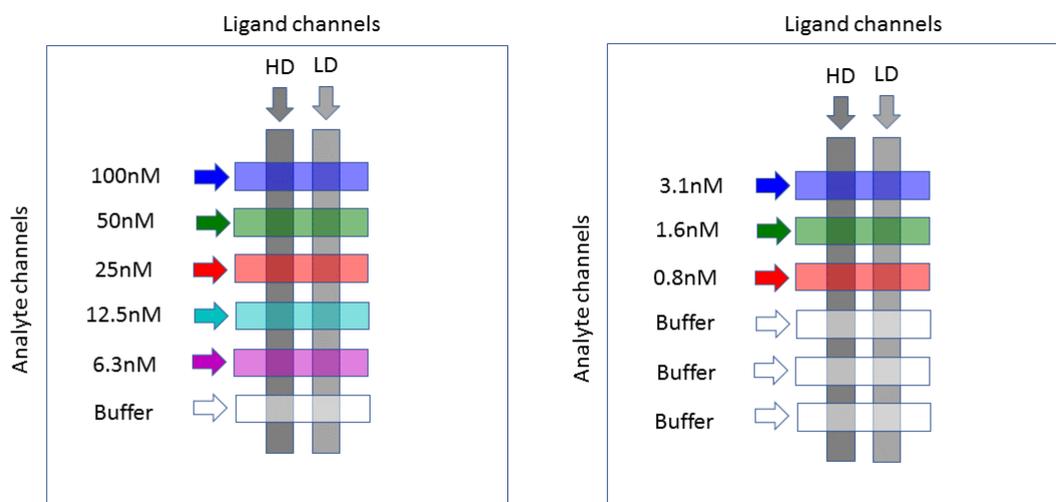
A range of eight concentrations, 100, 50, 25,12.5, 6.3, 3.1, 1.6 and 0.8nM of SN607D8 and SN230G6 were used to study the binding interactions. These concentrations were intended to explore the full range of responses that SPR was expected to measure - from very high to very low.

Due to the number of concentrations and densities in the experiments, each used multiple chips. Diagrams of these chips are shown in Fig. 4.1 and Fig. 4.3. Each of these diagrams shows two chips, although in either case, replicates were made for both chips, meaning four were used in each.

These experiments were performed at 37° and the pH was maintained at 7.4 for running buffer, to allow binding to happen as it would in vivo.

The sensorgrams produced by the SN607D8 vs A2 experiments are shown in Fig. 4.2, and those produced by the SN230G6 vs A2 interactions are shown in Fig. 4.4.

Experiments involving a third antibody, WK1D12, were originally planned as part of a separate dataset that could be used to test the models developed on the first. However, as there were a number of artefacts in the other experiments (see Subsection 4.3.1 and 4.3.2)



**Figure 4.1:** Diagram of interaction spots on the XPR 36 chips for the SN607D8 antibody vs HLA-A2 antigen experiment

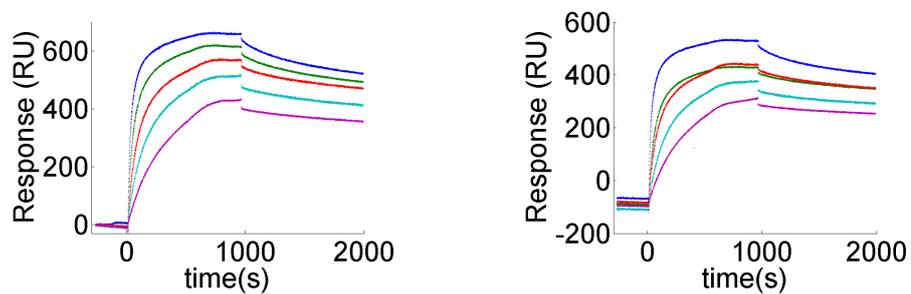
and there was equipment failure meaning the original experiments couldn't be repeated these experiments were used to supplement the original data set.

The antibody WK1D12 was only reacted against antigen immobilised at a single concentration (0.25  $\mu\text{g}/\text{ml}$ ). However, it was reacted against two antigens B7 and B27. The lower density antigen was used because there had been no benefit shown in the model development and fitting process to using the higher density of antigen.

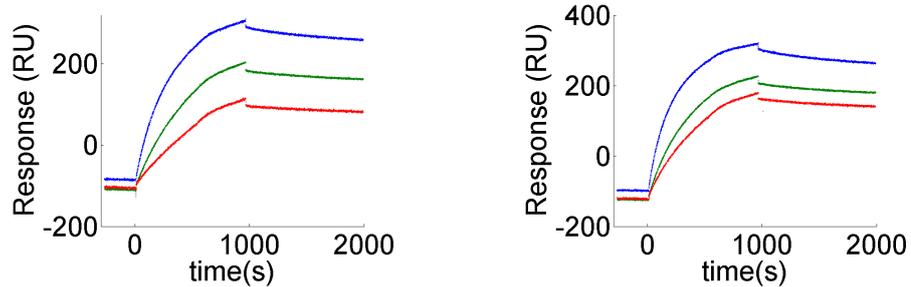
The antigen B7 and B27 share an epitope, 163EW+self73TE, which WK1D12 can bind to. As a result, these experiments could be used to detect the differences in the affinity of a single antibody for a single epitope on two separate antigens. If a model could consistently estimate these affinities so that the estimates varied less between replicates than reactions with separate antigen then this would be a powerful demonstration of the model's reliability.

These reactions were also at 37<sup>0</sup> Celsius with running buffer at a pH of 7.4.

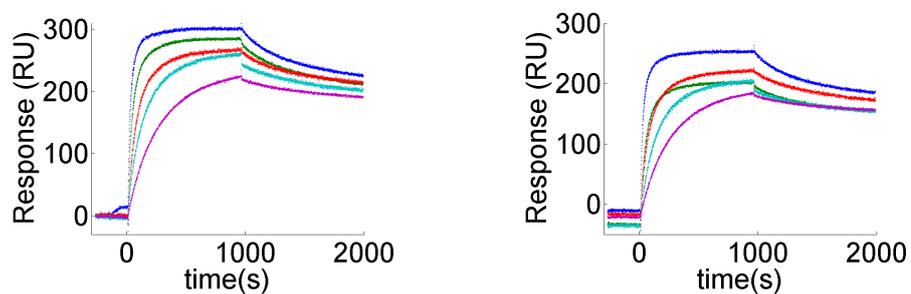
Instead of being conducted against eight concentrations of the WK1D12 antibody the



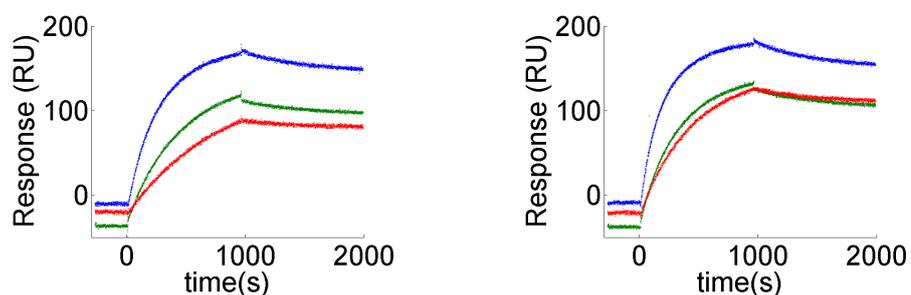
(a) High density antigen, antibody concentrations from 100-6.3nM



(b) High density antigen, antibody concentrations from 3.1-0.8nM



(c) Low density antigen, antibody concentrations from 100-6.3nM



(d) Low density antigen, antibody concentrations from 3.1-0.8nM

**Figure 4.2:** Sensorgrams from SN607D8 antibody vs HLA-A2 antigen experiments

experiments were done with six. This meant that all of the reactions could be done on a single row of the XPR-36 chip. These six concentrations were 25, 12.5, 6.3, 3.1, 1.6 and 0.8nM. The chip design for these experiments is shown in Fig. 4.5.

The sensorgrams produced by these experiments are shown in Fig.4.6.

### 4.3.1 Reacting SN607D8 antibody with HLA-A2 antigen

In Figure 4.2 the sensorgrams produced for SN607D8 antibody reacting against HLA-A2 antigen are shown. The eight concentrations of antibody to be reacted were split across two lanes into five highest concentrations on one lane, and three lowest concentrations on the other, as illustrated in Fig. 4.1. This meant that in either lane one reaction spot was available for reference subtraction.

Looking at all of the graphs, we may note that the curves made with a higher concentration of antibody are initially much steeper in their increase in response to binding at the beginning of the association phase. These curves appear to approach saturation at the end of the phase whereas curves representing lower concentration have much more gradual slopes generally not approaching saturation.

At the end of the association phase, for most time series there is a sudden drop or jump in response. Following this, the response decreases in a much more gradual fashion. In the higher concentration data this gradual decrease is initially faster than in the low concentration data, and generally, it approaches the rate of decrease of the low concentration data at the end of the phase.

Some of these qualitative features are well explained by the Langmuir model (see Example 2.1 and 2.2 Section 2.2) . High concentration of analyte leads to rapid binding at the start of each phase, as the density of free ligand decreases the rate of binding decreases and the output approaches a steady state.

Similarly, in the dissociation phase, the analyte flowing over the chip is replaced with buffer, causing gradual dissociation of bound antibody.

In this data set, four unpredicted features of the sensorgram were categorised. These behaviours are not predicted by either form of the earlier introduced Langmuir model, or any of the models from the literature that will be introduced in Chapter 5. These features are named humps, jumps, convergences and cross-overs.

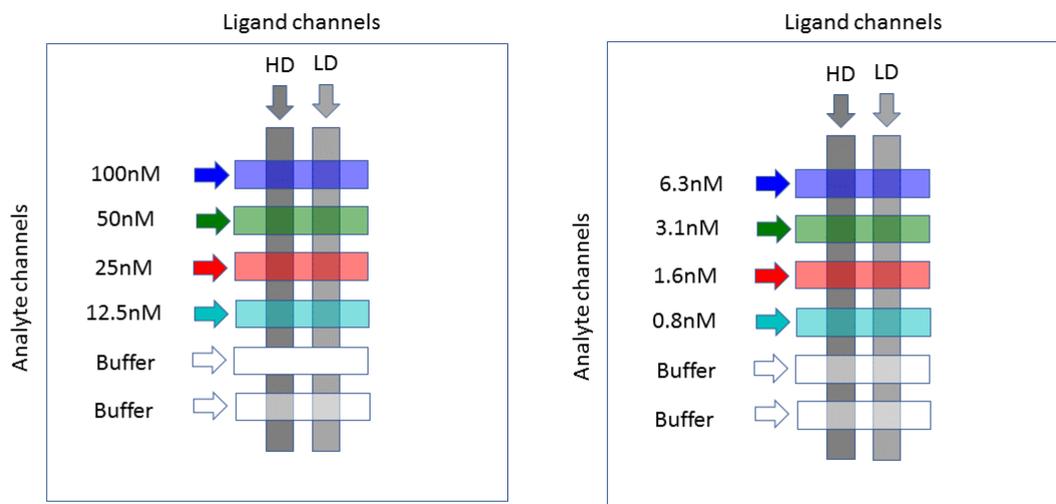
Jumps can be seen in all sensorgrams of Fig. 4.2, although they are most pronounced in the high-density high-concentration experiments, and least pronounced in the low-density antigen experiments. They are abrupt changes in output signal - typically occurring at the transition from association to dissociation phase. These may be caused by changes in the concentration of unbound analyte near the interface, as well as microscopic changes in temperature and pressure resulting in changes in the properties of the metal-dielectric interface of the measurement device (Richard B. M. Schasfoort, 2008).

Humps can be seen in two sensorgrams from Fig. 4.4 (a) and (b), and the first sensorgram in Fig. 4.2. A hump is defined as an abrupt and non-smooth change in the curve of the sensorgrams data. In all four cases listed they happen in the association phase.

Cross-overs are shown in all four sensorgrams of Fig. 4.2 (c) and (d) and one sensorgram from both (a) and (b). A cross over is defined as when a signal from one lane crosses with the signal from another lane.

Convergences appear in the dissociation phase of one sensorgram from each of Fig. 4.4 (a) and (c). Convergence is defined as a when the signal from one lane converges over the signal with another lane, and the output from the two lanes continues to overlap for the rest of the experiment.

When these experiments were initially conducted, it was unclear which of these features of the data were artefacts and which were a direct result of the antibody - antigen binding



**Figure 4.3:** Diagram of interaction spots on the XPR 36 chips for the SN230G6 antibody vs HLA-A2 antigen experiment

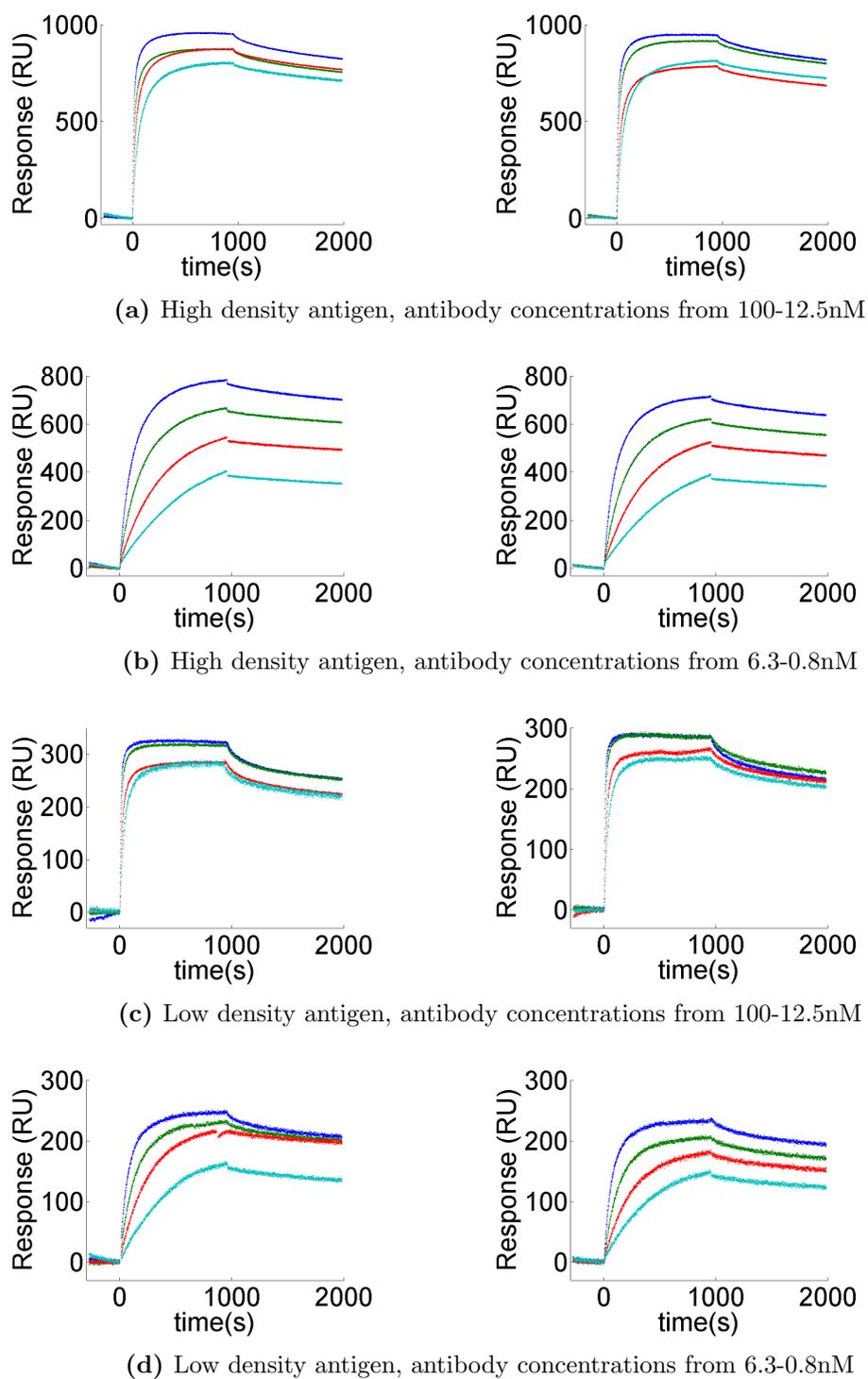
processes and needed to be included in the modelling.

#### 4.3.2 Reacting SN230G6 antibody with HLA-A2 antigen

In Figure 4.4 we see sensorgrams for the interaction of SN230G6 versus A2.

When these experiments were conducted, the parameter fitting discussed in Section 6.4.1 had already begun with the previous dataset. Because of this, the eight concentrations were divided into the four highest and the four lowest (see Fig.4.3). This was to decrease the size of the parameters space used by reducing the number of parameters required to model the high-concentration interactions together. As well as reference subtraction, these curves were adjusted so that the association phase would begin with a response of zero.

Associations are much more rapid than for the SN607D8 curves, with some of the low concentration curves appearing to approach saturation. There are fewer jumps in response when the phases switch, and where there are jumps, they are smaller.



**Figure 4.4:** Sensorgrams from SN230 vs HLA - A2 experiments

However, in five out of eight of the data sets, we see the outputs crossing over each other. We also see the outputs from different interaction spots of the chip converging with each other.

### 4.3.3 Reacting WK1D12 antibody with HLA-B7 and HLA-B27 antigens

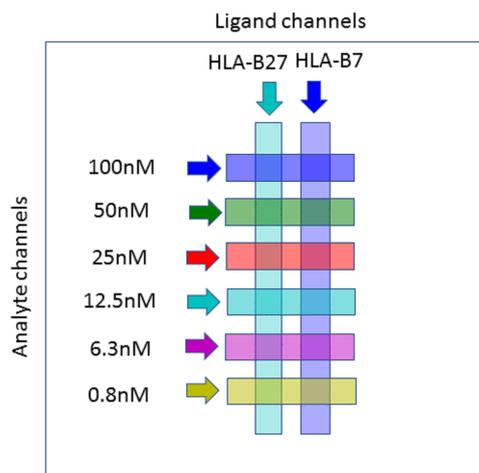
When these experiments were conducted, the parameter fitting discussed in Section 6.4.1 had already begun with the previous two datasets. A significant obstacle to this was the anomolous features of humps, jumps, convergence and cross-overs.

The decision to reduce the number of analyte concentrations in the experiment was made to reduce the number of chips used in each experiment and in anticipation of improvements in model fitting that would allow all interactions between an antibody/antigen pair to be simultaneously fit (this is achieved in Section 6.6). The high ligand density lanes were removed from the experiments because the anomalous features were more common in the sensorgram output from those lanes.

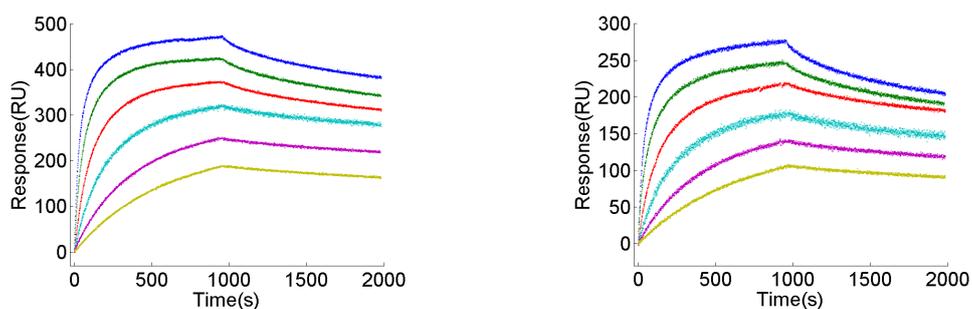
The new chip design is shown in Fig.4.5). Notably, it includes interaction with two antigens.

In Figure 4.6 we see sensorgrams for the interaction of WK1D12 against two antigens, HLA-B27 and HLA-B7. In this data set we see most of the same patterns but no humps in the association phase and no jumps or drops between the phases, although the signal from the two highest concentration lanes crosses over on one sensorgram and converges on its replicate.

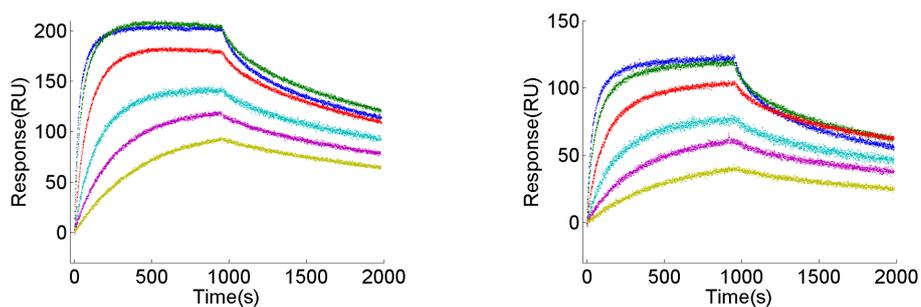
The data from this and the other original monoclonal experiments (Section 4.3) is analysed in Section 6.4.



**Figure 4.5:** Diagram of interaction spots on the XPR 36 chips for the WK1D12 antibody vs HLA-B27 and HLA-B7 antigens experiment



(a) Low density HLA - B27 antigen, antibody concentrations from 100-1.6nM



(b) Low density HLA - B7 antigen, antibody concentrations from 100-1.6nM

**Figure 4.6:** sensorgrams from experiments reacting WK1D12 against HLA - B27 and HLA-B7 antigens

## 4.4 Supplementary Monoclonal Experiments

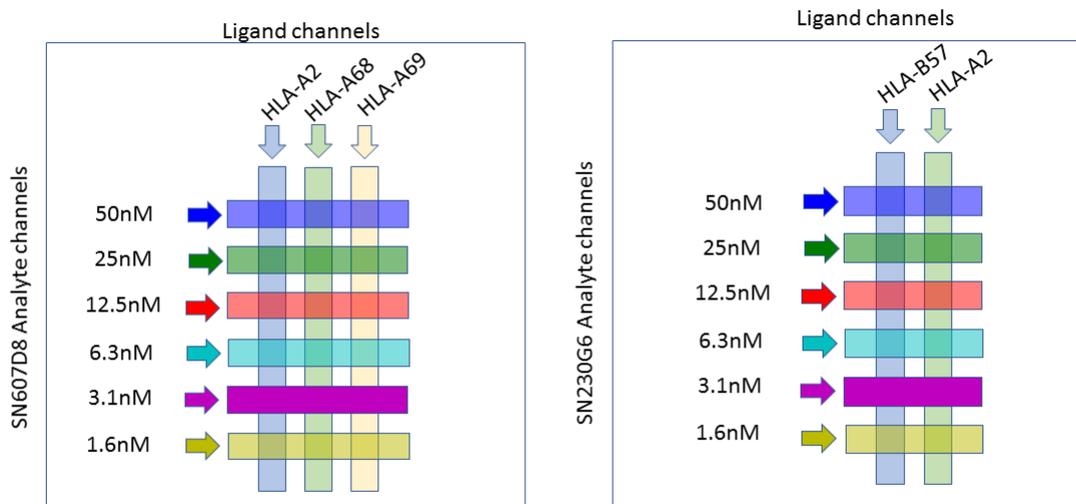
The experiments of the previous section allowed for model development and fitting (Chapters 5 and 6), which eventually resulted in a version of the bivalent ERC model with spatial effects (presented in Section 5.1.6). However, many of the sensorgrams that had been used in this development process included artefacts such as humps, abrupt jumps and lanes crossing over. Further experiments were then conducted to test the resulting models.

These experiments were of similar design to the WK1D12 against HLA-B7 and HLA-B27 experiments. Both SN607D8 and SN230G6 were reacted separately against antigen sharing an epitope they would bind to. The antibody SN607D8 was reacted against A2, A68 and A69; and SN230G6 reacted against A2 and B57.

The sensorgrams produced are shown in Fig. 4.8 and Fig. 4.9. These show convergence between lower concentration lanes in both of the SN607D8 vs. A2 sensorgrams and in the second SN607D8 vs. A68 sensorgram. The SN230G6 sensorgrams feature two lanes converging in one of the B57 sensorgrams and two cross-overs in the other, and two lanes converging in each of the A2 sensorgrams. Notably half of the sensorgrams in Fig. 4.9 have interactions from one lane removed due to problems with the sensor chip.

To compare the fits of the models developed with the previous data to the existing models these sensorgrams were fit with Bio-Rad's ProteOn<sup>TM</sup> Manager Software. This software uses a proprietary algorithm to instantly fit one of a small number of established models to the data (Bio-Rad, 2010).

The model chosen to fit the data was the bivalent Langmuir model (discussed in Section 5.3.4) because it was the only one of the models available that allowed for the analyte to bind both monovalent and bivalently to the analyte.



(a) Diagram of interaction spots on the XPR 36 chips for the SN607D8 antibody vs HLA-A2, HLA-A68 and HLA-A69 antigens experiment.

(b) Diagram of interaction spots on the XPR 36 chips for the SN230G6 antibody vs HLA-B57 and HLA-A2 antigens experiment.

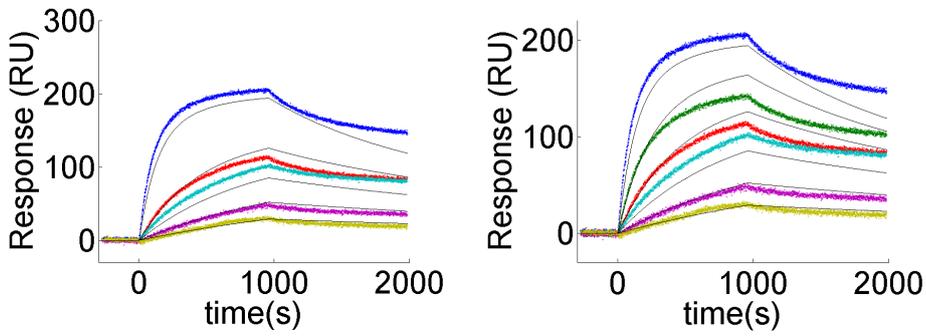
**Figure 4.7:** Diagram of interaction spots on XPR 36 chips

The fits are displayed as black lines on the Figs. 4.8 and 4.9. Typically they are poor, with the black lines remaining mostly outside of the coloured data.

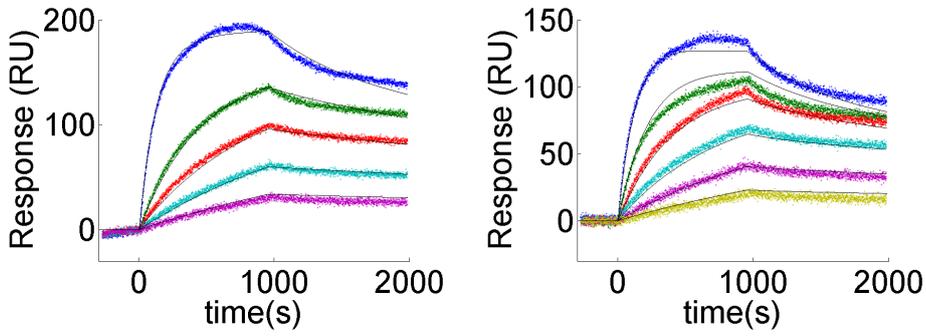
One of the weaknesses of the ProteOn manager software is that it only displays the kinetic rate constants and not the other parameters fitted. The parameters that are shown to the user are displayed in Tables 4.3 and 4.4.

Their analysis with the novel model, the Bivalent ERC model, is shown in Section 6.6.

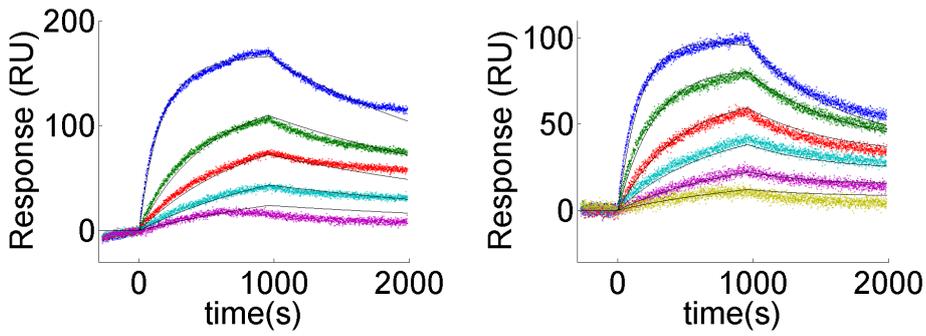
The ProteOn parameter estimates of the kinetic constants appear to show some difference between the two antibodies overall; Sn607D8's interactions have  $K_{DS}$  ranging from  $1.1E-8$  to  $8.7E-8$ , whereas Sn607D8's interactions have  $K_{DS}$  ranging from  $1.5E-9$  to  $2.7E-9$ . However, there is little difference between either antibody and its interactions with specific antigen.



(a) SN607D8 vs HLA-A2



(b) SN607D8 vs HLA-A68



(c) SN607D8 vs HLA-A69

**Figure 4.8:** SN607 vs HLA-A2, HLA-A68 and HLA-A69

Antigen	$k_a$ ( $M^{-1}s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$k_D$ (M)
HLA-A2	4.5E+4	5.1E-4	8.7E-8
	4.7E+4	6.5E-4	1.4E-8
HLA-A68	4.5E+4	1.1E-3	1.9E-8
	5.2E+4	5.9E-4	1.1E-8
HLA-A69	5.4E+4	4.9E-4	2.6E-8
	5.5E+4	1.2E-3	2.2E-8

Table 4.3: XPR estimates of SN607D8 kinetic constants

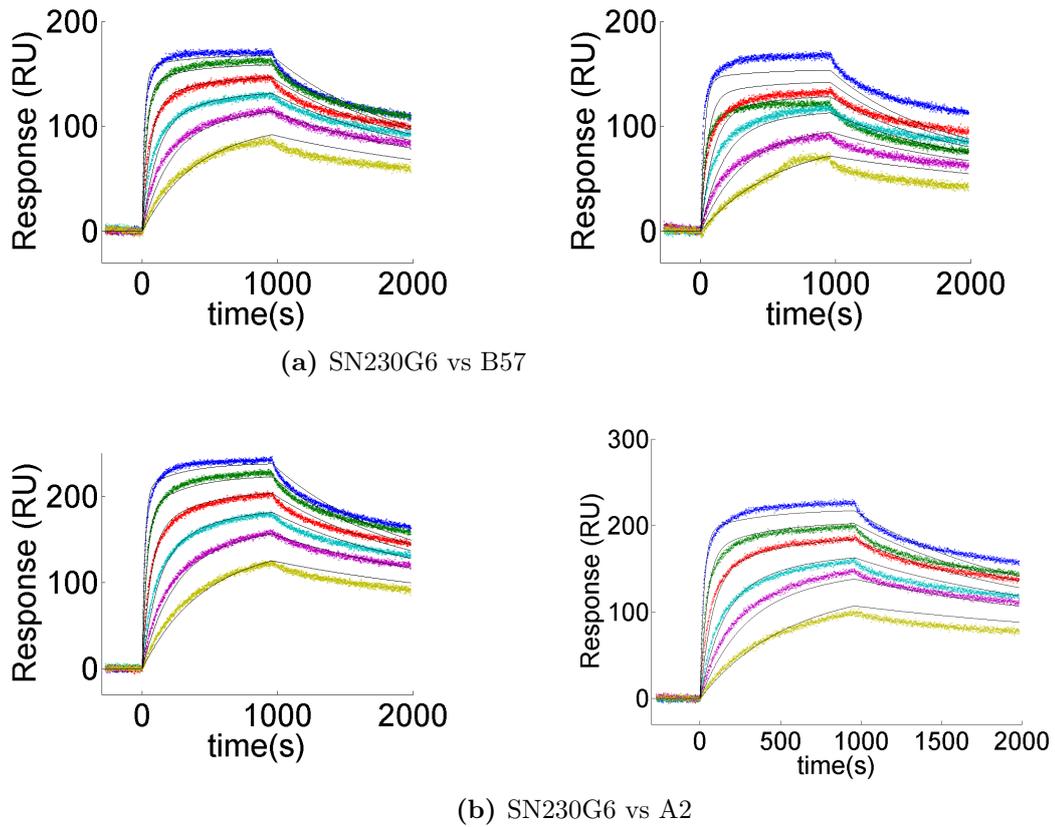
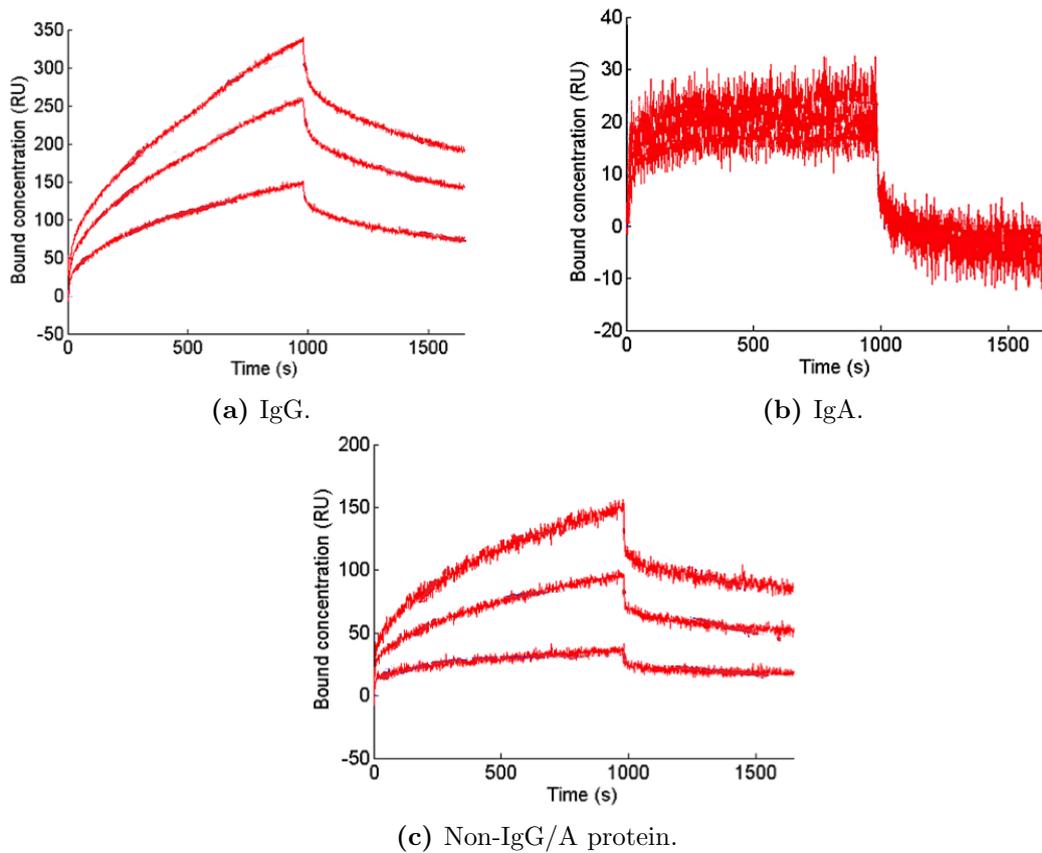


Figure 4.9: SN230G6 vs B57 and A2

Antigen	$k_a$ ( $M^{-1}s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$k_D$ (M)
HLA-B57	3.7E+5	6.5E-4	2.7E-9
	4.1E+5	6.0E-4	1.5E-9
HLA-A2	4.2E+5	6.9E-4	1.6E-9
	3.8E+5	6.9E-4	1.8E-9

Table 4.4: XPR estimates of SN230G6 kinetic constants.



**Figure 4.10:** Sensorgrams from purified patient protein vs A trisaccharide.

## 4.5 Purified IgG/IgA protein samples

The first of the experiments using clinical samples that became a part of this project was performed fully by Dr Andrew Bentall as part of his PhD (Bentall, 2015). Protein samples from a single patient that had been purified into three classes: IgG, IgA, and the remainder, a sample called non-IgG/A protein, which was expected to include IgM antibodies, amongst other protein.

These analytes were simultaneously passed over three ligands of an A trisaccharide with densities 0.3mg/ml, 0.2mg/ml and 0.1mg/ml at room temperature. The association phase began at 7.2s and lasted until 981.0s. The experiment lasted a further 666.0s. This gives

a total of 1647.0s. Sensorgrams from these experiments are shown in Fig. 4.10.

The analysis of these data is shown in Section 8.1 and was published in H.A.J.Moyse *et al.* (2012).

## 4.6 Plasma effluent and serum samples

The experimental methods of Sections 4.3- 4.4 were adapted for use with clinically derived samples. These samples were of two varieties: plasma effluent and serum both extracted from patient blood.

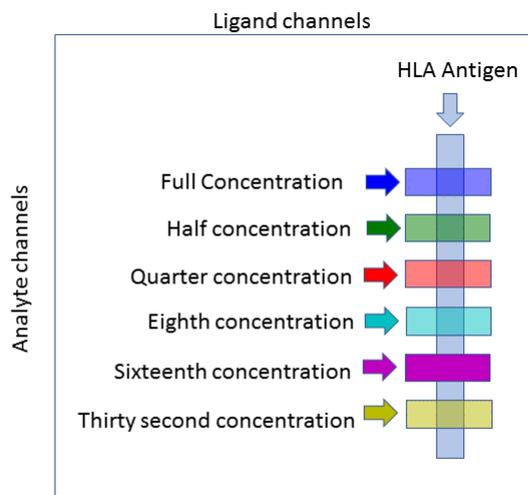
Plasma effluent, the cruder of the two sample types was readily available. This is because it constituted the waste of plasmapheresis therapy for the patients.

Plasmapheresis therapy was used to remove the antibodies of patients who were judged likely to undergo antibody mediated kidney rejection. The therapy consists of removing patient blood and returning the cells to the bloodstream by transfusion. The antibody along with other blood protein is left in its waste product, plasma effluent.

Plasma effluent contains a huge variety of molecules as well as antibodies: proteins, glucose, clotting factors, electrolytes, and hormones dissolved in water, as well as a variety of anticoagulants, depending on the patient it has been taken from.

Serum samples consisted of the liquid that was supernatant when patient blood was allowed to clot. As a result, these include many of the same elements as plasma effluent, except for fibrinogens- the proteins involved with forming the clot.

Six serum samples were obtained by special request from NHS Blood and Transplant (NHSBT) Birmingham. These were typing sera taken from multi-parous women, and they were chosen for having high complement dependent cytotoxicity (CDC) titre for HLA-A2 and HLA-B7.



**Figure 4.11:** Diagram of interaction spots on the XPR 36 chips for the patient samples against HLA antigen experiments

Undesirable binding to the matrix chip from both plasma effluent and serum samples was measured in Daga (2015). In response to this, a variety of techniques for the refinement and enrichment of the samples was developed. Whilst these were being developed numerous SPR experiments were conducted. Those that showed a high signal to noise ratio were selected for modelling, these are shown in Fig. 4.12 and Fig. 4.13.

These experiments were conducted whilst I had been modeling the monoclonal sensorgrams, so are primarily the work of Dr Daga.

These samples have unknown concentration so a chip design diagram like Fig. 4.3 cannot be made for them. Instead with all patient samples, each was diluted so that concentration was halved between lanes (see Fig. 4.11).

Plasma effluent from the cases LT33 and LT79 was processed through HLA-A2 affinity chromatography. The eluted proteins were concentrated before being run through a Protein G-Sepharose column and subjected to size exclusion chromatography.

Serum samples were processed via sequential precipitation and HLA-protein chromatog-

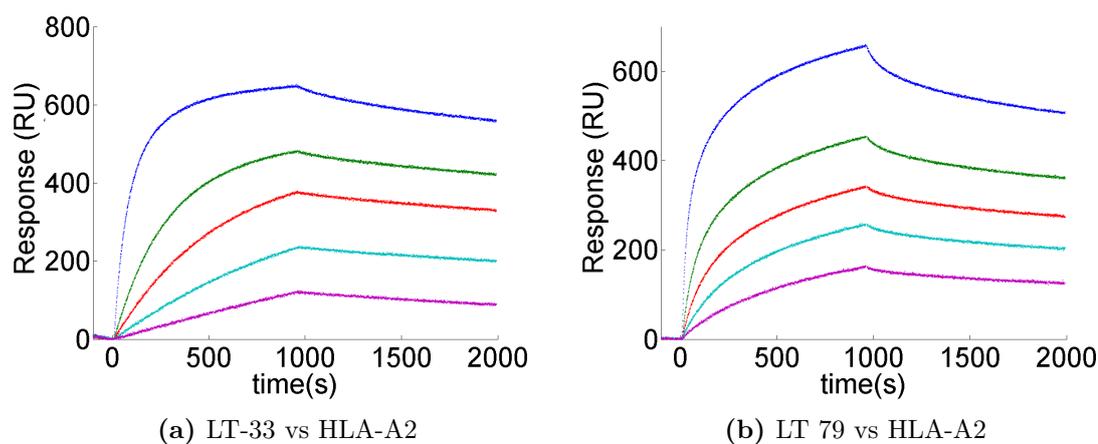


Figure 4.12: Sensograms from processed plasma effluent reacted against HLA antigens

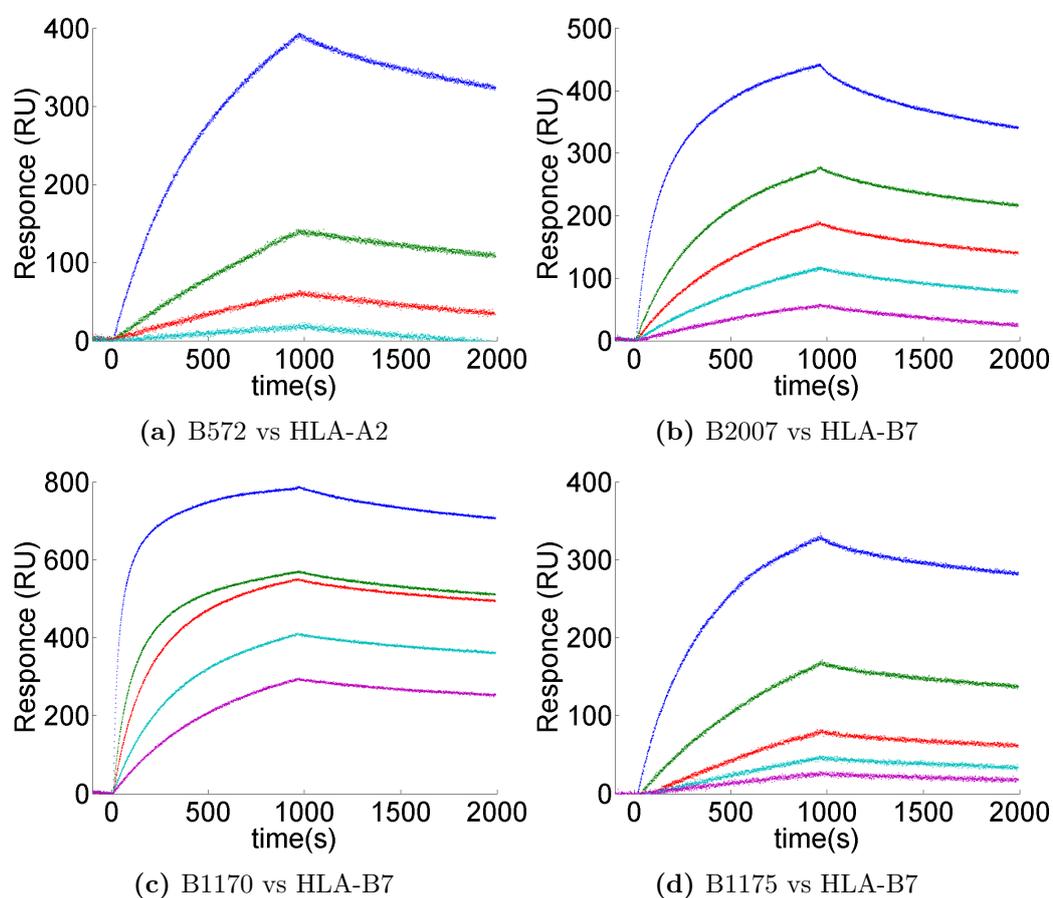


Figure 4.13: Sensograms from processed serum samples reacted against HLA antigens

raphy followed by size exclusion chromatography. The HLA protein chosen for the chromatography was in each case the protein that the sample had a high CDC titre for. In the case of B572 it was HLA- A2, in the cases of B2007, B1170 and B1175 it was HLA - B7.

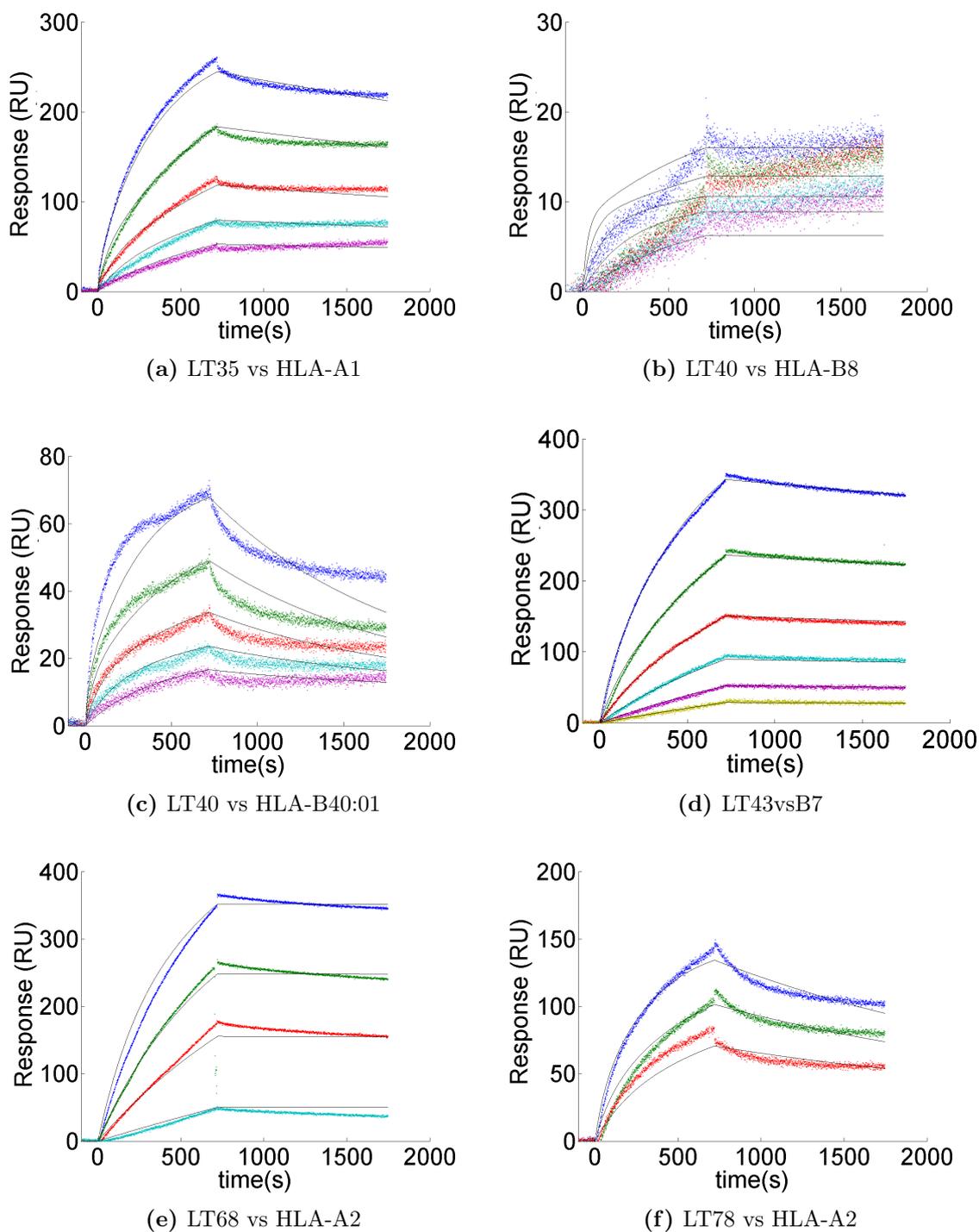
The sensorgrams produced (Fig. 4.12) show large differences from the sensorgrams produced from the monoclonal experiments. The unexpected features of monoclonal sensorgrams, humps, jumps, cross-overs and convergences are not present on any of the sensorgrams produced by the reacting of clinical samples.

## 4.7 High affinity plasma effluent sensorgrams

Whilst the monoclonal sensorgrams were being analysed more sensorgrams featuring patient serum samples reacting against HLA antigen were made available. These were sensorgrams that were deemed anomalous in a study of sensitised patient antibody affinity that was being conducted at the time by Dr Daga. Although they were produced with the same techniques that had been developed in the previous section and had few visual defects, the Proteon Manager modelling software had produced affinity estimates close to or higher than  $5E-12M$ , the maximum affinity of antibody previously estimated (Zahnd *et al.*, 2004; Abcam, 2013a).

The sensorgrams are shown in Fig. 4.14 and the affinity estimates based on these curves are shown in Table 4.5. Qualitatively these differ from the previous curves, particularly the total decrease of response over the whole dissociation phase appears to be very low, making the dissociation curves appear very shallow.

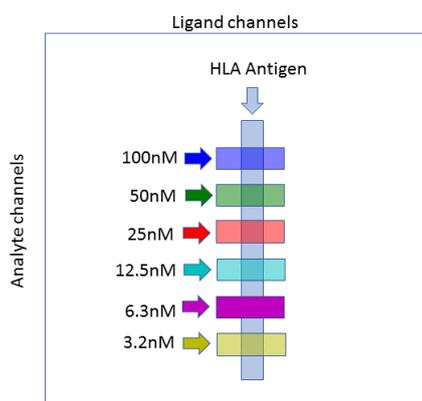
Notably, these sensorgrams are shown fitted with the Langmuir two ligand model by the ProteOn Manager software (this model is discussed in Subsection 7.1.1). The kinetic constants estimated by this model were beyond the bounds of what was considered possible



**Figure 4.14:** Sensorgrams of patient antibody being reacted against HLA antigen with fits

Patient	Antigen	$k_a$ ( $M^{-1}s^{-1}$ )	$k_d$ ( $s^{-1}$ )	$k_D$ (M)
LT68	HLA-A2	2.4E+5	2.1E-6	8.7E-12
LT16	HLA-A1	1.2E+4	1.2E-9	1.0E-13
LT40	HLA-B8	1.2E+6	1.0E-11	8.4E-18
LT35	HLA-A1	6.6E+6	2.4E-11	3.6E-18
LT43	HLA-B7	1.1E+6	3.5E-16	3.3E-22
LT40	HLA-B40:01	5.0E+6	1.9E-17	3.8E-24

**Table 4.5:** XPR estimates of kinetic constants



**Figure 4.15:** Mixed SN607D8 and SN230G6 vs HLA-A2 experiment design.

for antibody binding. This motivated the development of a new model which included the assumptions that had created the best monoclonal models, as well as features that may emerge from numerous antibody binding to numerous epitopes on the antigen (see Subsection 7.2).

## 4.8 Mixed monoclonal experiments

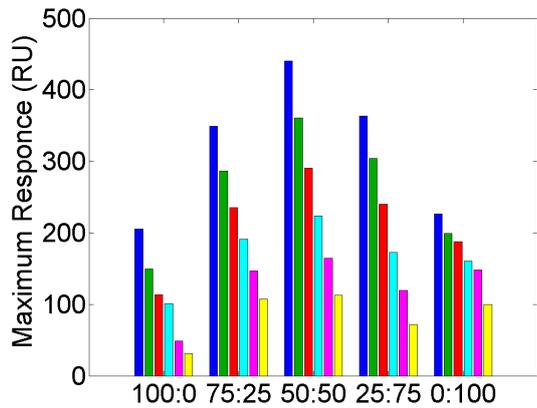
One hypothesis about a physical phenomenon that could produce structures that dissociate very slowly would be complexes containing multiple antibody binding together with multiple epitopes on a single antigen. The mathematics of this interaction is explored in Subsection 7.2.

To test whether mixtures of monoclonal antibody would show similar behaviours to the high-affinity sensorgrams of the previous section a further set of experiments was conducted. These are shown in Fig. 4.16, with a diagram of the sensor chip shown in Fig. 4.15.

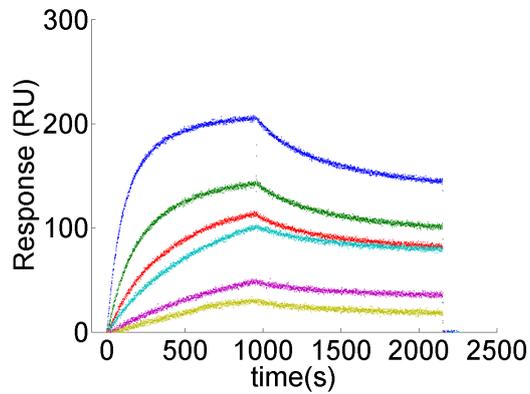
Figures 4.16 (a) and 4.17 (a) shows that the sensorgrams from the mixed experiments were qualitatively different to the monoclonal experiments. That is that mixed antibody binding produces a higher response signal than either monoclonal.

However Fig. 4.16 (b)-(f) 4.17 (b)-(f) do not show dissociation output curves as shallow as those in 4.14.

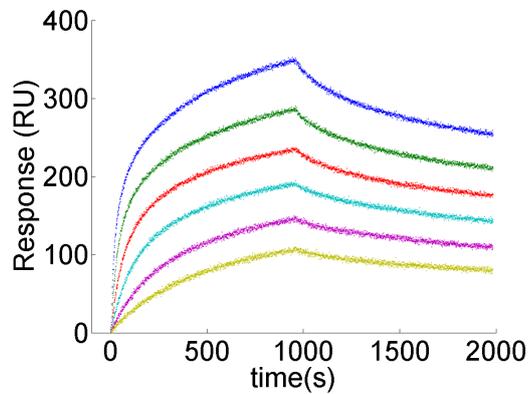
An artefact that can be seen in some of these sensorgrams is particularly clear in Fig. 4.17 (d), when the well at the inlet of the SPR machine runs out of sample, in this case buffer there is a sudden change in the angle of surface plasmon resonance that can be seen here as drops in signal in each lane.



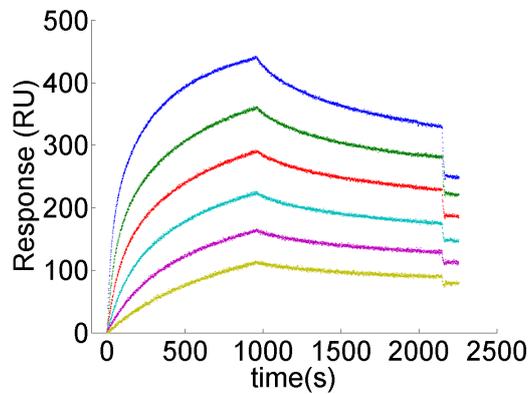
(a) Maximum response for each lane for each antibody mixture



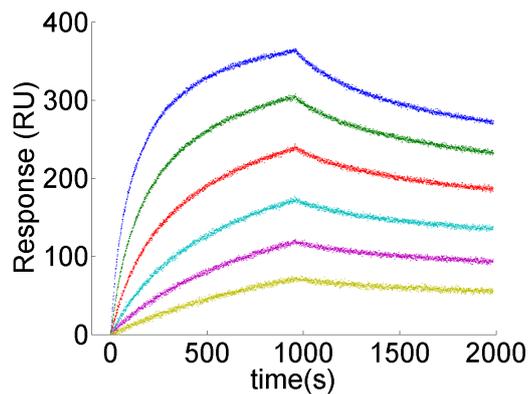
(b) SN607D8 in concentrations 100-3.1nM



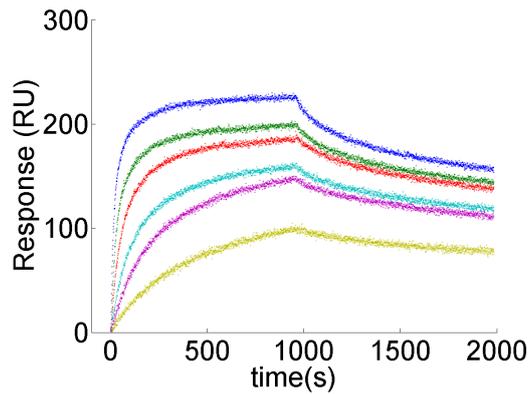
(c) 75:25 mixture of SN607D8 with SN230 in concentrations 100-3.1nM



(d) 50:50 mixture of SN607D8 with SN230 in concentrations 100-3.1nM

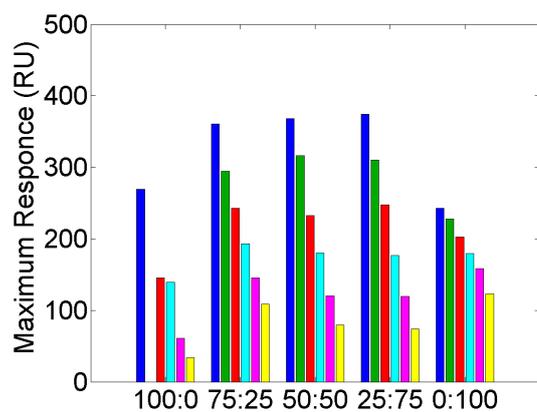


(e) 25:75 mixture of SN607D8 with SN230 in concentrations 100-3.1nM

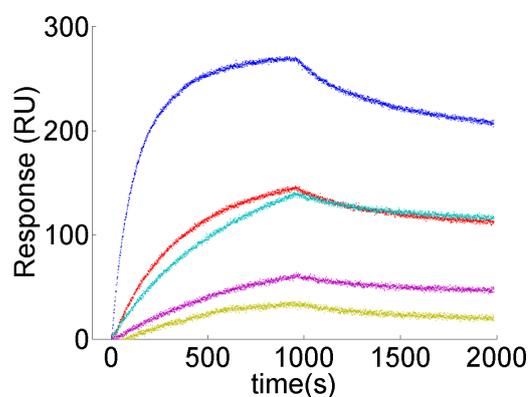


(f) SN230G6 in concentrations 100-3.1nM

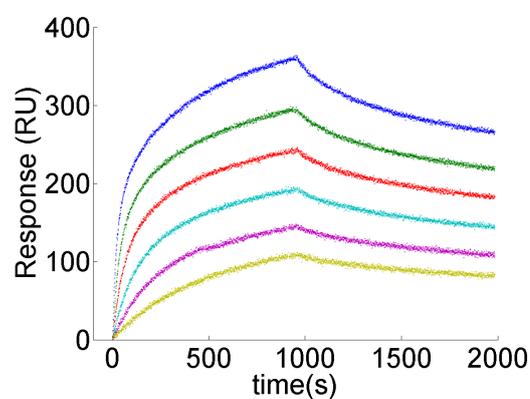
**Figure 4.16:** Sensorgrams from mixtures of monoclonal antibody reacted against HLA-A2 antigen



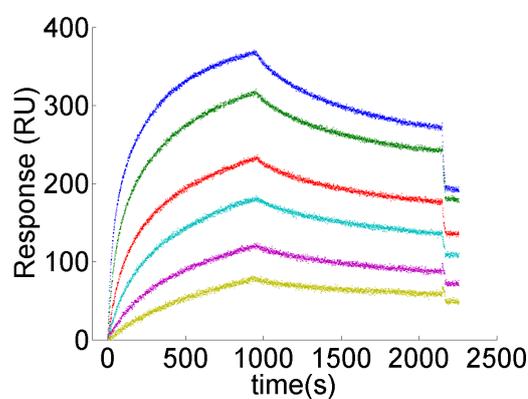
(a) Maximum response for each lane for each antibody mixture



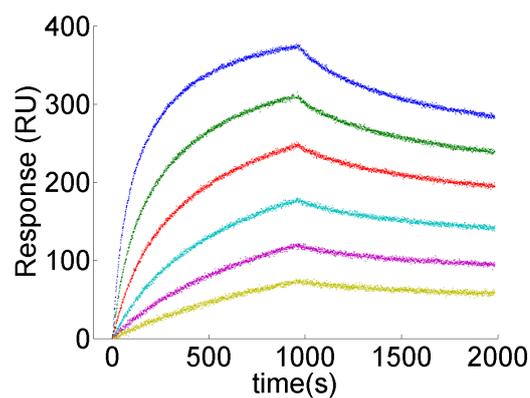
(b) SN607D8 in concentrations 100-3.1nM



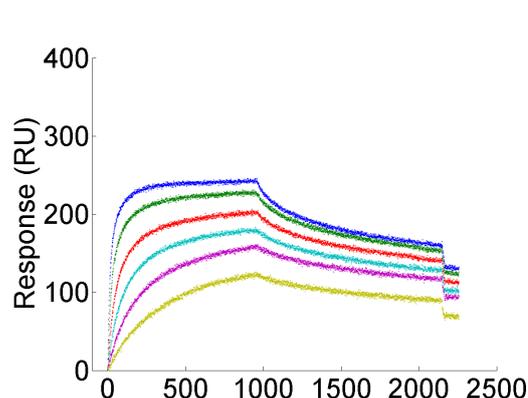
(c) 75:25 mixture of SN607D8 with SN230 in concentrations 100-3.1nM



(d) 50:50 mixture of SN607D8 with SN230 in concentrations 100-3.1nM



(e) 25:75 mixture of SN607D8 with SN230 in concentrations 100-3.1nM



(f) SN230G6 in concentrations 100-3.1nM

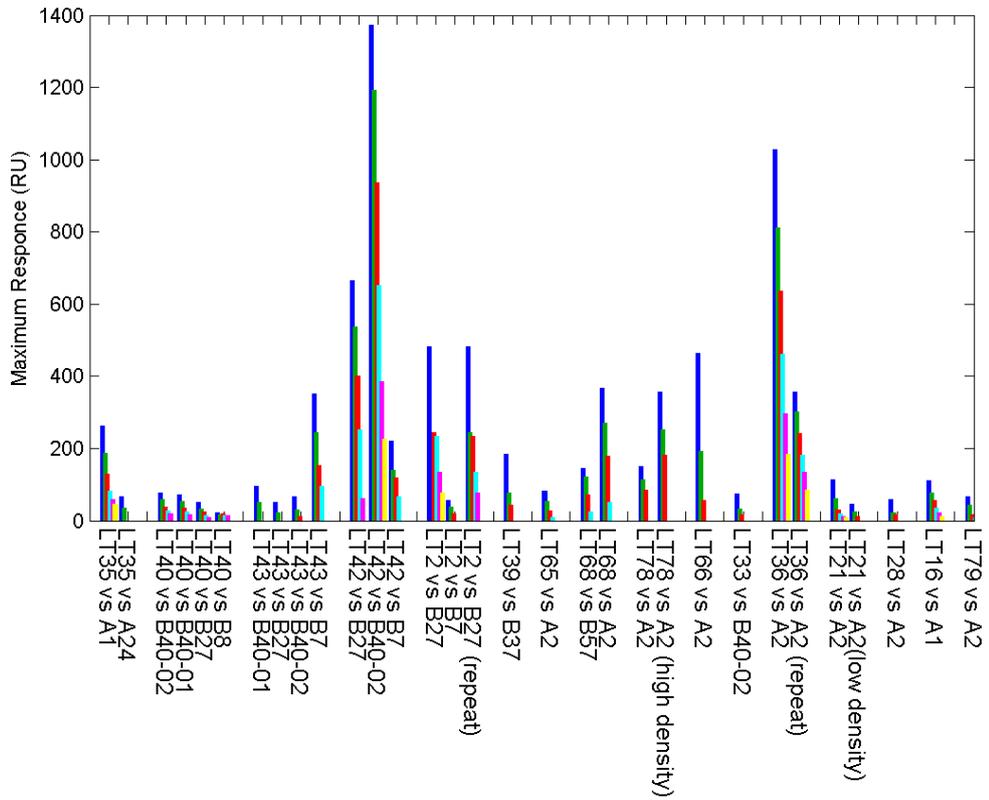
**Figure 4.17:** Sensorgrams from mixtures of monoclonal antibody reacted against HLA-A2 antigen

## 4.9 Sensitised patient study

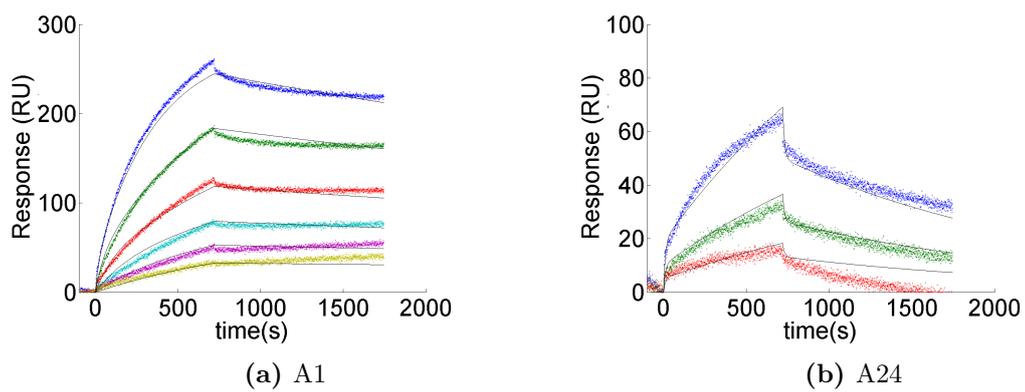
After the mixed monoclonal experiments were conducted the full dataset that the experiments from Section 4.7 were taken from was made available for modelling.

Due to the size of this dataset, it is useful to summarise the data. The maximum values taken by the sensorgram outputs are shown in (Fig. 4.18). The values given by the XPR fits of the model are shown in Fig. 4.6.

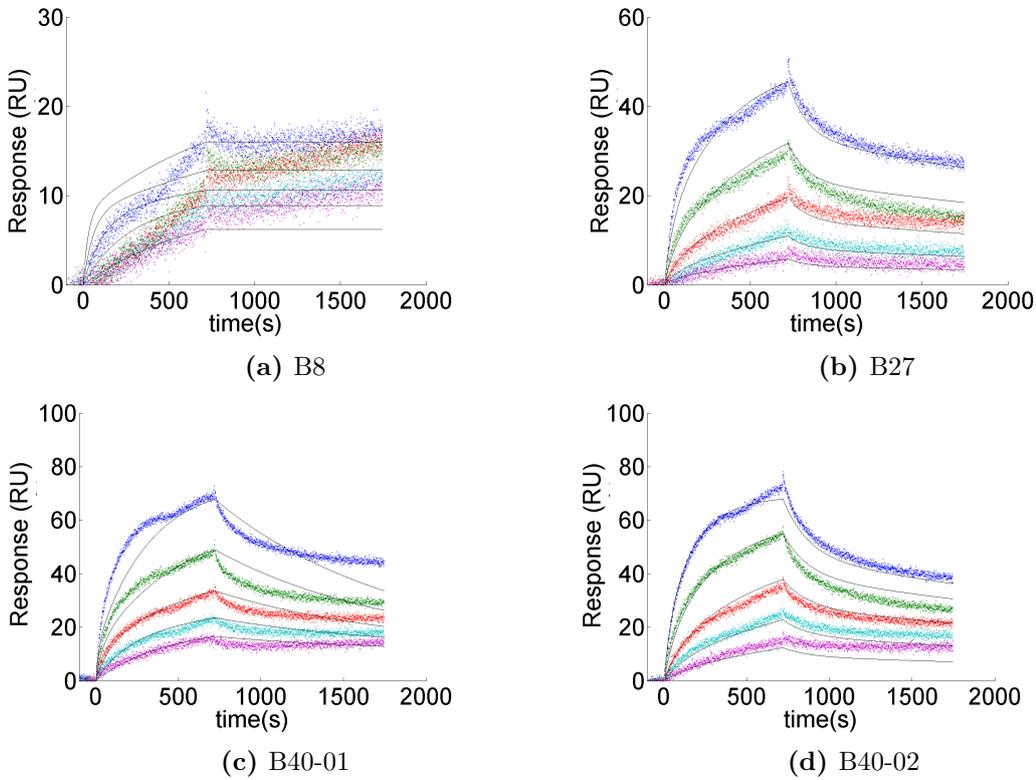
Notably, both the tables and figures include the data previously presented in Subsection 4.7.



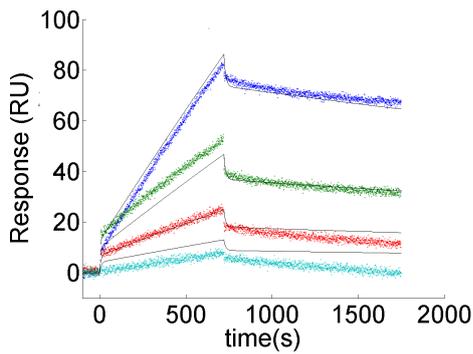
**Figure 4.18:** Maximum response for each lane in each sensorgram from the sensitised patient study.



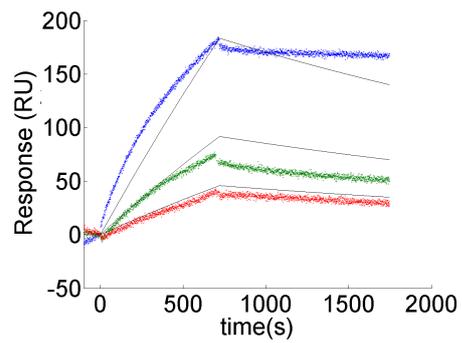
**Figure 4.19:** Serum from LT35 reacted with the multiple antigen



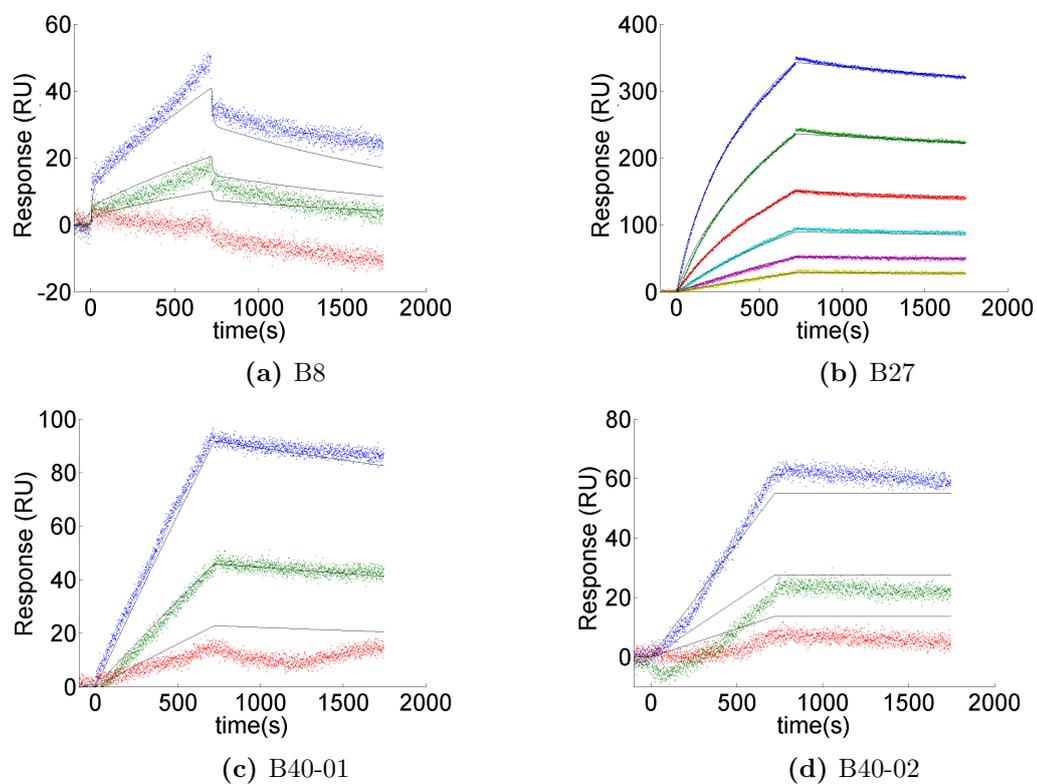
**Figure 4.20:** Serum from LT40 reacted with the multiple antigen



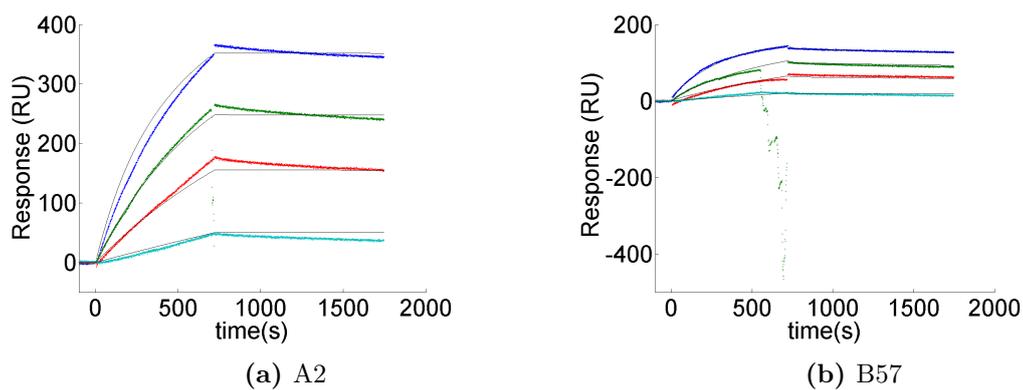
**Figure 4.21:** Serum from LT65 reacted against A2 antigen



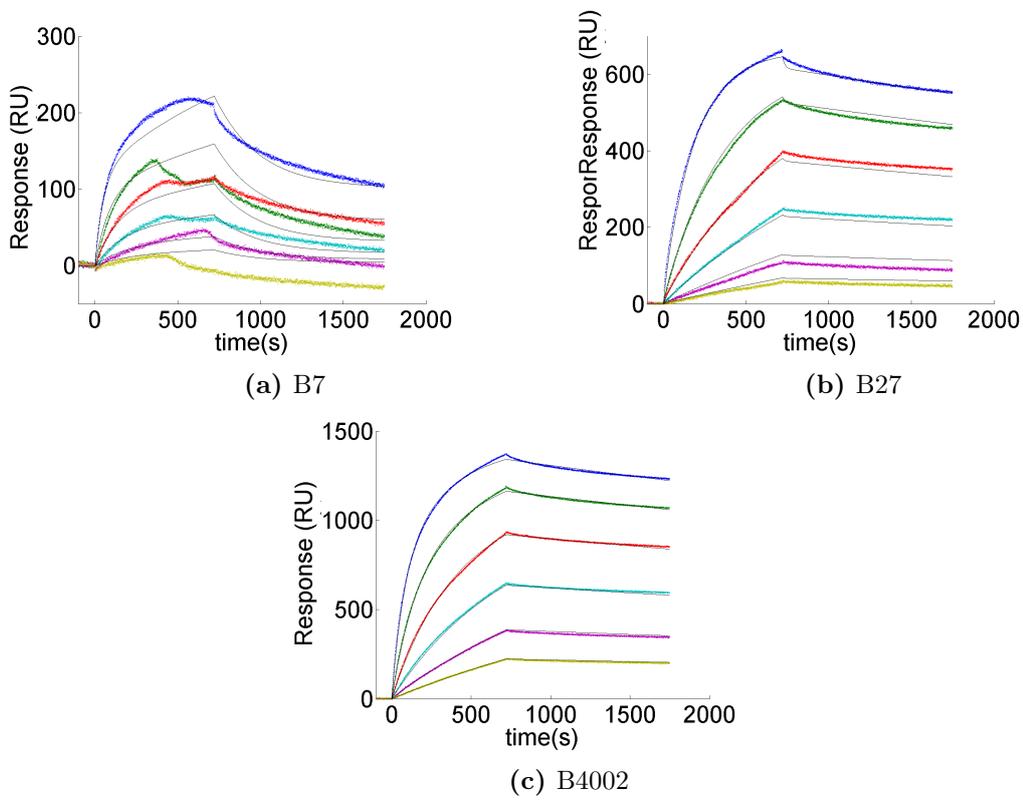
**Figure 4.22:** Serum from LT39 reacted against B37 antigen



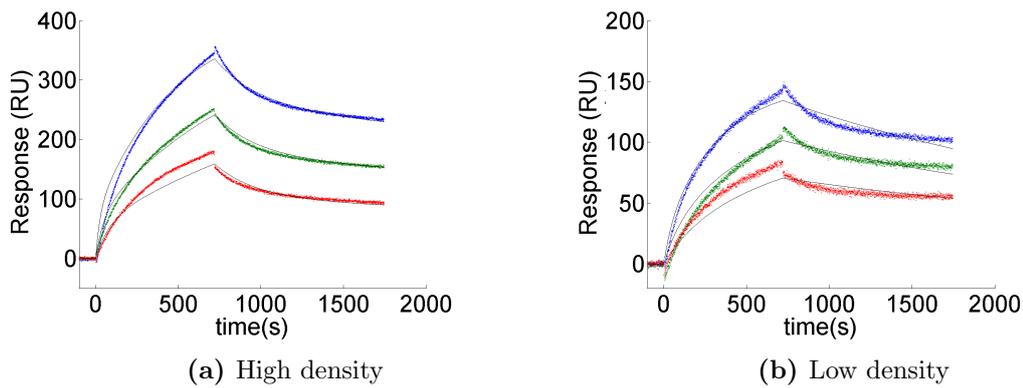
**Figure 4.23:** Serum from LT43 reacted with multiple antigen



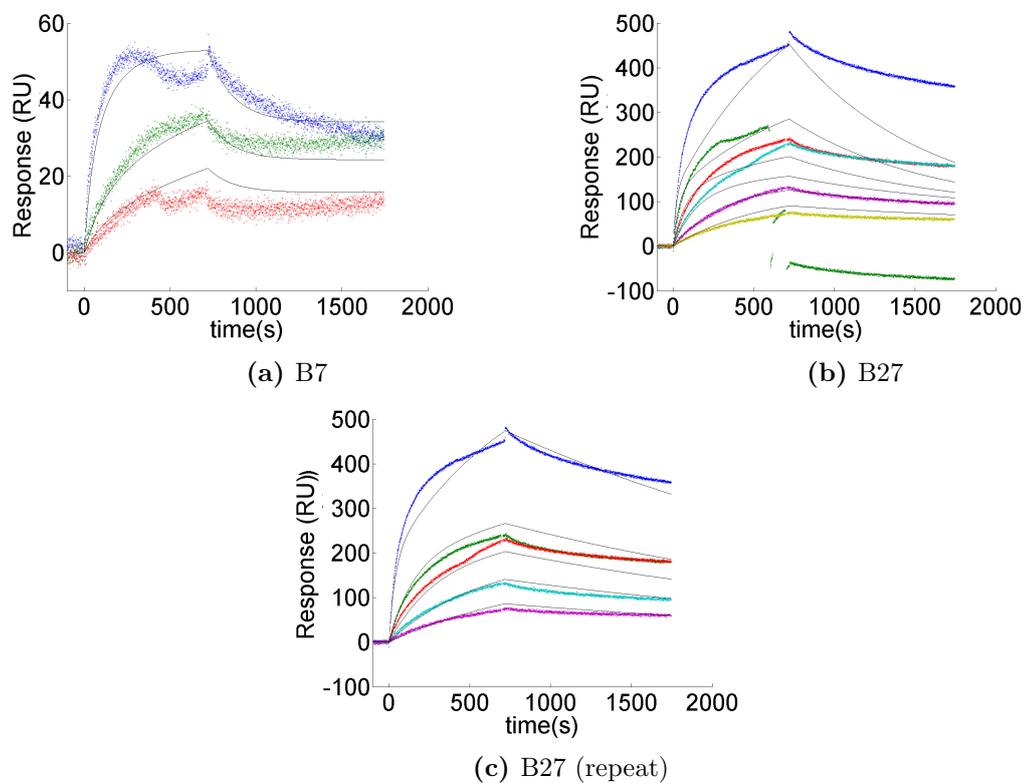
**Figure 4.24:** Serum from LT68 reacted with multiple antigen



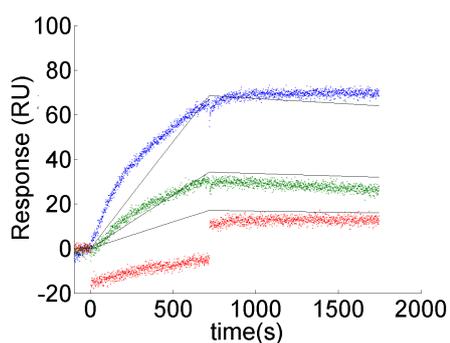
**Figure 4.25:** Serum from LT42 reacted with multiple antigen



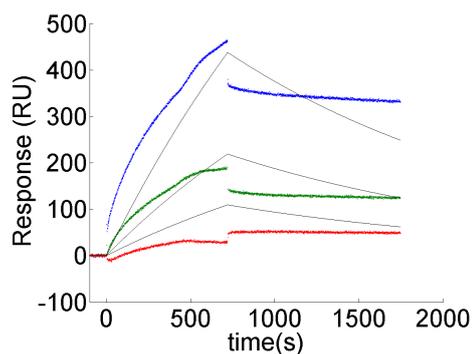
**Figure 4.26:** Serum from LT78 reacted against A2 antigen in various densities



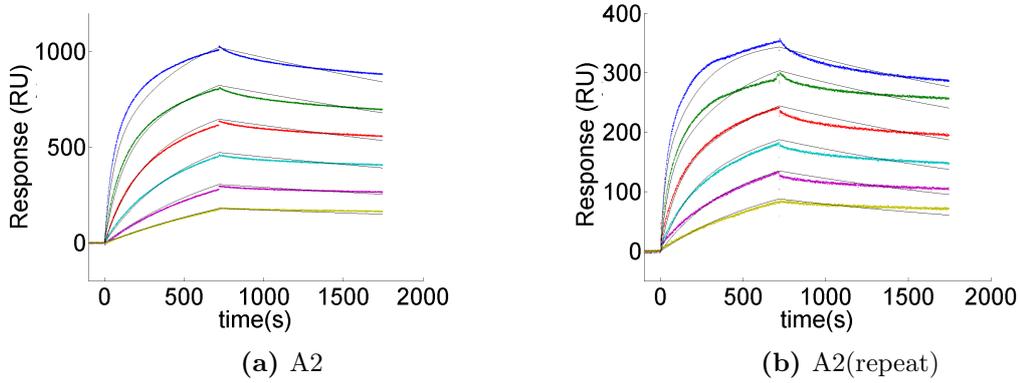
**Figure 4.27:** Serum from LT2 reacted with multiple antigen



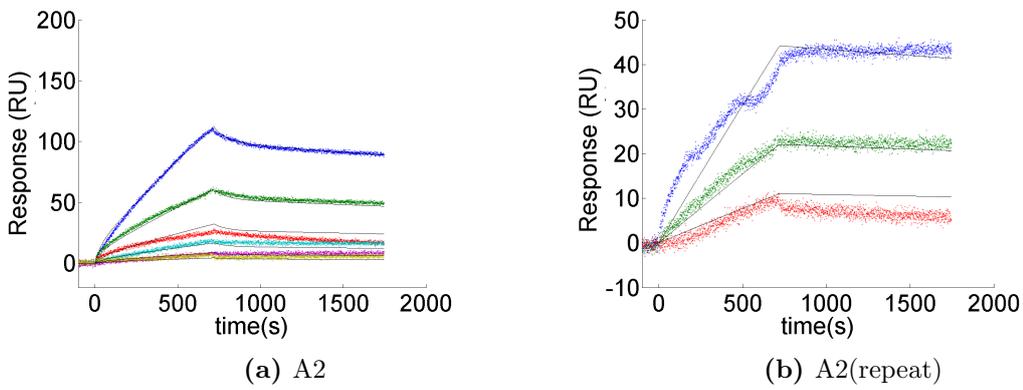
**Figure 4.28:** Serum from LT33 reacted against B4002 antigen



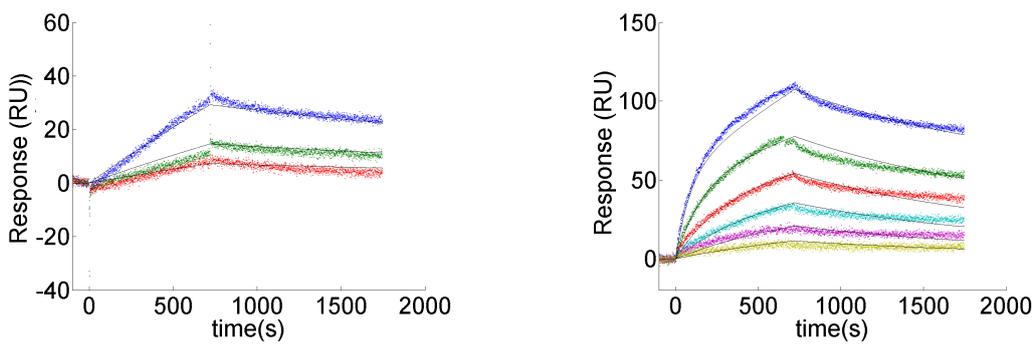
**Figure 4.29:** Serum from LT39 reacted against A2 antigen



**Figure 4.30:** Serum from LT36 reacted against multiple antigen



**Figure 4.31:** Serum from LT21 reacted against multiple antigen



**Figure 4.32:** Serum from LT16 reacted against A1 antigen

**Figure 4.33:** Serum from LT79 reacted against A2 antigen

Patient	Antigen	$k_a$ $M^{-1}s^{-1}$	$k_d$ $s^{-1}$	$k_D$ M
LT42	B27	3.27E+1	7.57E-2	2.31E-03
LT42	B7	4.56E+05	3.72E-03	8.16E-09
LT50	B7	8.45E+00	2.77E-03	3.28E-04
LT28	A2	2,78E+0	2.56E-04	9.20E-05
LT43	B27	1.29E+1	5.43E-04	4.20E-05
LT43	B40:01	3.84E+00	1.04E-04	2.99E-05
LT43	B40:02	2.01E+00	4.88E-08	2.42E-08
LT43	B7	1.06E+06	3.48E-16	3.28E-22
LT39	B37	1.55E+1	2.63E-04	1.69E-05
LT35	A24	4.98E+1	5.75E-04	1.15E-05
LT35	A1	.55E+06	2.36E-11	3.61E-18
LT33	B40:01	8.93E+00	5.26E-05	5.90E-06
LT66	A2	9.21E+1	5.10E-04	5.53E-06
LT2	B27	1.12E+04	3.42E-04	3.04E-08
LT2	B7	2.01E+06	6.98E-03	3.47E-09
LT36	A2	1.83E+04	2.00E-04	1.09E-08
LT36	A69(LD)	5.09E+04	1.52E-04	2.99E-09
LT65	A2	1.37E+03	1.28E-04	9.37E-08
LT40	B27	1.51E+06	9.66E-03	6.37E-09
LT40	B40:02	1.20E+06	6.74E-03	5.61E-09
LT40	B8	1.21E+06	1.01E-11	8.38E-18
LT40	B40:01	5.04E+06	1.91E-17	3.78E-24
LT21	B57	2.71E+04	1.14E-04	4.20E-09
LT79	A2	1.21E+05	2.66E-04	2.20E-09
LT42	B40:02	6.62E+04	8.77E-05	1.33E-09
LT78	A2	1.35E+07	7.39E-03	5.46E-10
LT68	B57	1.97E+05	7.43E-05	3.78E-10
LT68	A2	2.44E+05	2.12E-06	8.70E-12
LT16	A1	1.17E+04	1.19E-09	1.02E-13

Table 4.6: XPR estimates of kinetic constants

## 4.10 Summary

A large amount of data, from multiple studies conducted by multiple investigators has been presented. This experimental work forms the backbone of the main body of this thesis. In Chapters 5-8 every model and fitting procedure is presented or developed with

Experiment	Experamentalists	Analysis
SN607G8 vs HLA-A2	H. Moyse and S. Daga	Section 6.4
SN230G6 vs HLA-A2	H. Moyse and S. Daga	Section 6.4
WK1D12 vs HLA-B7 and HLA-B27	H. Moyse and S. Daga	Section 6.4 and 6.6
SN230G6 vs HLA-A2 and HLA-B57	S. Daga	Section 6.4 and 6.6
SN607G8 vs HLA-A2, HLA-A68 and HLA-A69	S. Daga	Section 6.4 and 6.6
SN607G8 and SN230G6 mixture	H. Moyse and S. Daga	Section 8.3
Purified IgG/IgA protein samples	A. Bentall	Section 8.1 and 6.6
Plasma Effluent	S. Daga	Section 8.2
Serum Samples	S. Daga	Section 8.2
Sensitized patient study	S. Daga	Section 8.5

**Table 4.7:** XPR estimates of kinetic constants

the motivation that it can be used to model some part of this data. Table 4.7 summarises where the data presented in this chapter is analysed, as well as who the key investigators were that worked on it.

# Chapter 5

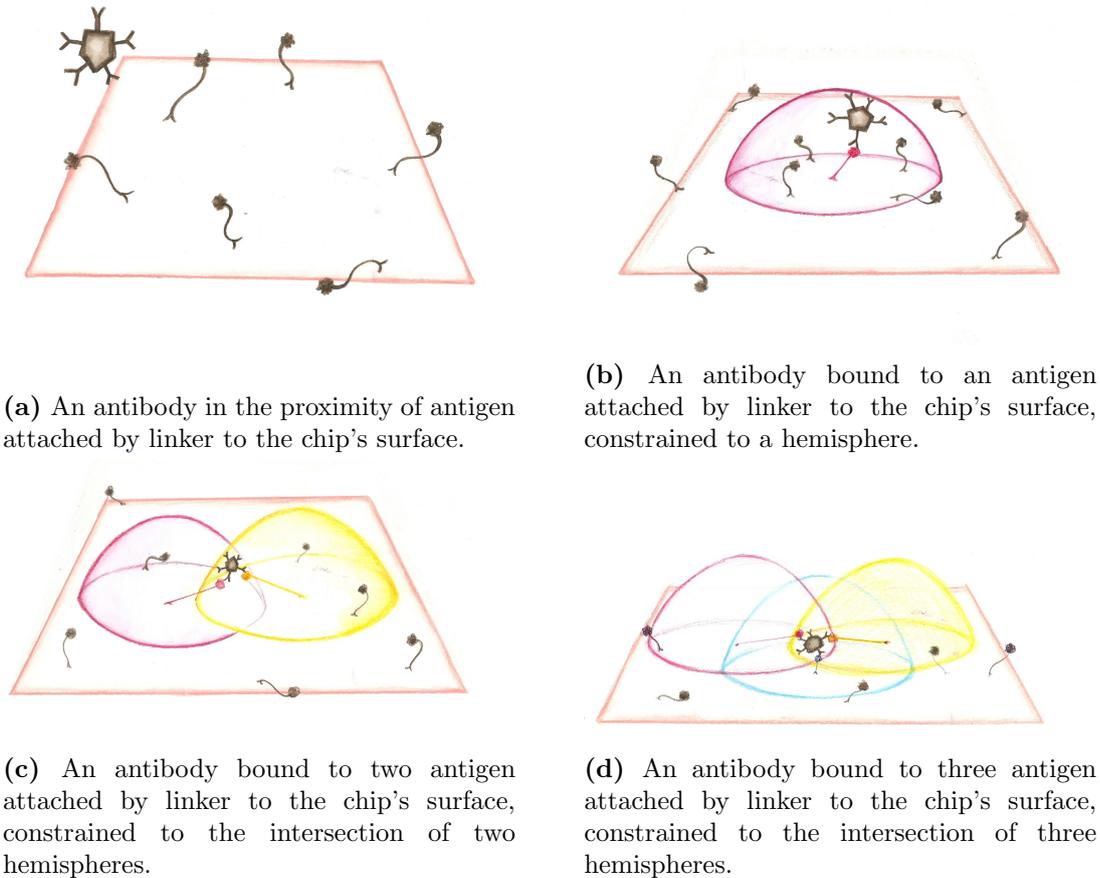
## Models for the reaction kinetics of monoclonal antibody in SPR experiments

### 5.1 Models

There are mathematical models in the literature that would be appropriate for simulating SPR experiments between monoclonal antibody and homogenous antigen, and there are models that with some changes will be appropriate. This chapter will present these models along with newly formulated ones. Identifiability and indistinguishability results for all models will be presented, and all of the models will be discussed and compared.

We may divide models into those that only allow monomers to dimerise, and those that allow the formation of trimers and larger complexes.

The first three models dealt with in this chapter are of the first type: the Langmuir model, the Langmuir with transport model, and the effective rate constant (ERC) approximation model. Previously they were adapted to allow for dimers formed by multiple binding configurations of the same monomers. As a result, these models require some changes to be appropriate for modelling monoclonal antibody.



**Figure 5.1:** Diagrams of an IgM antibody binding to multiple antigen

The last three models of this chapter are of the second type: the bivalent Langmuir model, the bivalent ERC model, and the bivalent ERC model with spatial effects. These models allow monomers to dimerise and dimer-monomer pairs to trimerise. This accurately reflects the binding reactions that monoclonal IgG may undergo when binding with antigen.

IgA and IgM may undergo more complex binding processes in SPR experiments as they have more paratopes, either 10 or 12 respectively (Roitt *et al.*, 2017).

However, as an antibody binds to an antigen, it becomes localised into a hemisphere, with a radius equal to the maximum distance the antibody's paratopes can be from the point at which the linker joins the chip. It is only able to bind to another antigen within that

hemisphere. If it binds to another antigen, its movement is restricted to the intersection of two hemispheres. As a result, the number of antigen that are available for it to bind to, without unbinding from the antigen it is already bound to, is reduced. As it binds with more antigen, the volume it may move within diminishes, as does the number of antigen that can come into that volume. This is illustrated in Fig. 5.1.

As a result, the models considered in this chapter will only consider monovalent and bivalent binding, although they could be expanded to consider more complex multivalent binding processes.

### 5.1.1 Langmuir model

The **single configuration Langmuir (L1) model** is the oldest model applied to SPR experiments. It was introduced in Eq. 2.5 and discussed in Example 2.1 in Section 2.2.

An extension of it, the **two configuration Langmuir (L2) model** was introduced in Eq. 2.10 and discussed in Example 2.2 in Section 2.2, and a phase portrait of it was shown and analysed in Example 2.3 in Section 2.3.

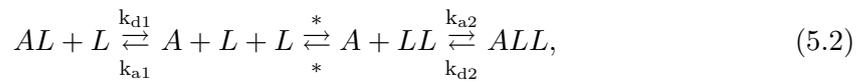
As shown in its chemical equation (Eq. (2.9)), this model allows for two separate species of dimer. However, IgG antibody forms both dimers and trimers with antigen. That is, that one of the configurations that antigen and antibody can form requires two antigen. The **adjusted two configuration Langmuir (LA) model** was derived by changing the capacity for addition antibody to bind in the L2 models state-space equation, Eq. 2.10,

from  $R - x_1 - x_2$  to  $R - x_1 - 2x_2$  giving the LA models definition:

$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = k_{a1}I(R - x_1 - 2x_2) - k_{d1}x_1 \\ \frac{dx_2}{dt} = k_{a2}I(R - x_1 - 2x_2) - k_{d2}x_2 \\ x(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, \mathbf{p}) = \alpha(x_1(t, \mathbf{p}) + x_2(t, \mathbf{p})), \end{cases} \quad (5.1)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R)$ . The parameters  $k_{a1}$  and  $k_{a2}$ ;  $k_{d1}$  and  $k_{d2}$ ; are the association and dissociation rate constants of the two binding species respectively measured in  $M^{-1}s^{-1}$  and  $s^{-1}$  respectively,  $C_T$  is the concentration of analyte flowing over the sensor chip in  $nM$  and  $R$  is the density of ligand attached to the sensor chip in  $pg\ nm^{-2}$ . The input function  $I(t)$  represents concentration of analyte at the machines inlet and  $T_a$  and  $T_d$  are intervals on the real line representing the association and dissociation phase, defined as  $T_a = [0, t_1], T_d = [t_1, t_2]$  respectively, and  $\alpha$  is a known constant  $1\ RU\ pg^{-1}\ nm^2$ .

Whilst this was not derived from a chemical equation, it can be thought of as assuming one. This new chemical equation is:



where  $A$ ,  $L$ ,  $AL$ , and  $ALL$  represent an antibody, an antigen, an antigen-antibody dimer and an antibody-antigen-antibody trimer respectively;  $LL$  represents an antigen pair that an antibody can simultaneously bind to, and the rates of the asterisked reaction are instantaneous, so that each antigen is available both for monovalent binding, and a member of a pair available for bivalent binding.

In terms of this, the elements of the state variable  $(x_1, x_2)^\top = \mathbf{x}(t; \mathbf{p}, I) \in [0, R] \times [0, \frac{1}{2}R]$  can be thought of as representing the concentration of *AL* dimers and *ALL* trimers respectively, in  $\text{pg nm}^{-2}$ .

Notably, even though  $x_2$  now represents an antigen-antibody-antigen trimer, not a dimer, the output equation is still the same as it was for the L2 model, Eq. 2.10. This is because the output represents change in the angle of surface plasmon resonance – and this is assumed to be affected by the set of objects attached to the reaction spot, not their individual arrangement; so the same output is given for the same surface density of antibody bound regardless of whether they are bound bivalently or monovalently (see Section 1.4).

The phase plane for this new system is qualitatively the same as that for the two concentration Langmuir model, so need not be shown or discussed. The major advantage of this adjusted form of the model is that, because it considers dimers and trimers, it better describes the experimental system.

### 5.1.2 The Langmuir with transport model

The single configuration Langmuir model was extended by Myszka *et al.* (1998b) into the **single configuration Langmuir with Transport (LwT1) model**. This model uses the same chemical equation (Eq. (2.7)) but adds an extra variable  $C$  representing the concentration of the analyte at the SPR chip's surface. Like the Langmuir model this model also designed for a generic analyte binding monovalently to a ligand, not antibody antigen interactions.

Parameter	$k_a$ $M^{-1}s^{-1}$	$k_d$ $s^{-1}$	$R$ $pg\ nm^{-2}$	$C_T$ $nM$	$k_m$ $pgnm^{-2}\ M^{-1}s^{-1}$	$h$ $pgnm^{-2}\ M^{-1}$
Value	6E-4	7E-3	500	100	$10^{-4}$	$10^{-6}$

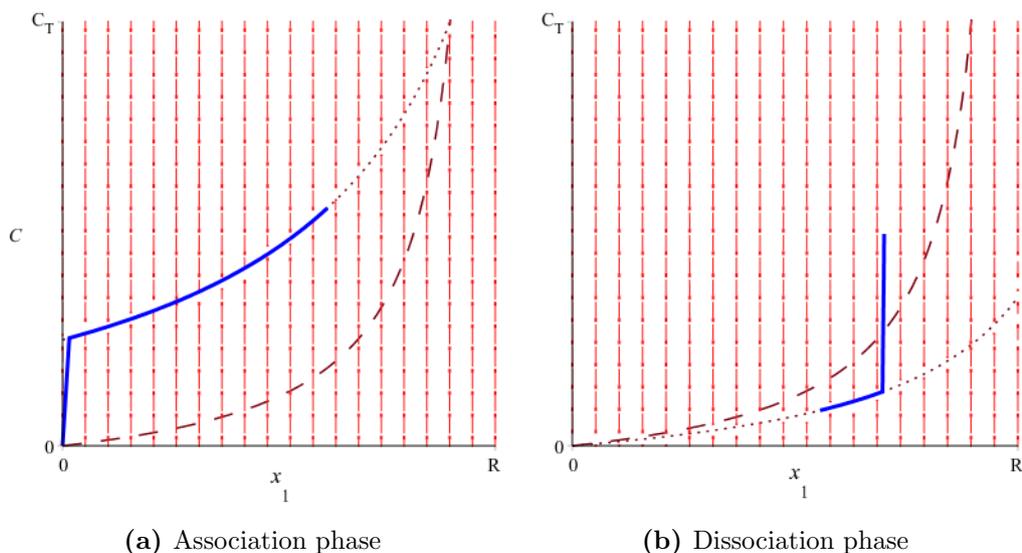
**Table 5.1:** Values taken by parameters in the simulation shown in Fig. 5.2.

Fig. 5.3(a) shows a schematic of this model. The model is defined as:

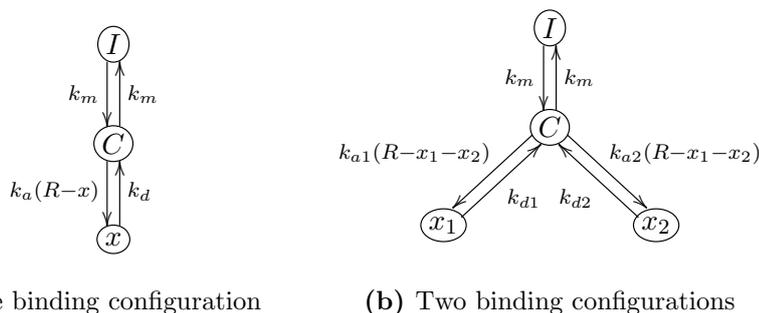
$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ h \frac{dC}{dt} = -k_a C (R - x) - k_d x + k_m (I - C) \\ \frac{dx}{dt} = k_a C (R - x) + k_d x \\ x(0; \mathbf{p}, I) = 0 \\ y(t, \mathbf{p}) = \alpha x(t, \mathbf{p}), \end{cases} \quad (5.3)$$

where the parameter vector is defined as  $\mathbf{p} = (k_a, k_d, C_T, k_m, h, R)$ . The two new parameters  $k_m$  and  $h$  are defined as the transport coefficient describing diffusive movement of analyte between the flow and the volume in contact with the surface and the maximum distance from the chip's surface that antibody can still bind to antigen, and are measured in  $pgnm^{-2}\ M^{-1}s^{-1}$  and  $pgnm^{-2}\ M^{-1}$ , and the other parameters have the meanings and units introduced with the single configuration Langmuir model (Example 2.1 in Section 2.2). The elements of the state variable  $(x, C)^T = \mathbf{x}(t; \mathbf{p}, I) \in [0, R] \times [0, C_T]$  represent the concentration of analyte-ligand dimers, and analyte in contact with the ligand bound to the chip in  $pg\ nm^{-2}$  and  $M$  respectively.

A phase portrait for this model is shown in Fig. 5.2. In the association phase, Fig. 5.2(a) concentration at the surface of the chip rapidly increases until an equilibrium between the analyte coming into contact with the chip and the analyte binding and unbinding with the chip is reached, the dotted nullcline. The system moves along this nullcline as more



**Figure 5.2:** Phase portraits of the single configuration Langmuir with transport model in the association and dissociation phase, with nullclines (dotted and dash lines), and a single solution (blue)



**Figure 5.3:** Compartmental diagram of Langmuir with transport model with inlet analyte concentration  $I$ , concentration of analyte at the reaction chip's surface  $C$  concentrations of bound analyte  $x_1(t)$  and  $x_2(t)$ .

analyte binds for the rest of the phase. In the dissociation phase, the  $C$  nullcline is moved, due to the change in inlet concentration,  $I$ . The concentration of the analyte at the chip's surface rapidly decreases, and the concentration of analyte bound to the chip decreases for the rest of the phase.

Evans *et al.* (2013) similarly extended the two configuration Langmuir model, to create

the **two configuration Langmuir with transport LwT2 model**. This is shown in Fig. 5.3(b). This model is defined as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ h \frac{dC}{dt} = \sum_{i=1}^2 \left[ -k_{ai}C \left( R - \sum_{j=1}^2 x_j \right) + k_{di}x_i \right] + k_m(I - C) \\ \frac{dx_1}{dt} = k_{a1}C \left( R - \sum_{i=1}^2 x_i \right) - k_{d1}x_1 \\ \frac{dx_2}{dt} = k_{a2}C \left( R - \sum_{i=1}^2 x_i \right) - k_{d2}x_2 \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0, 0)^\top \\ y(t, \mathbf{p}) = \alpha(x_1(t, \mathbf{p}) + x_2(t, \mathbf{p})), \end{cases} \quad (5.4)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, k_m, h, R)$ , and  $k_m$  and  $h$  are defined identically and measured with the same units as they were with the single concentration Langmuir with transport model and the other parameters have the definitions and units introduced with the two configuration Langmuir model (Example 2.2 in Section 2.2). The state variable  $(x_1, x_2, C)^\top = \mathbf{x}(t; \mathbf{p}, I) \in \mathcal{X}$ , represents the surface concentrations of two configurations of analyte-ligand binding in  $\text{pg nm}^{-2}$ , and the concentration of the analyte in  $nM$ , and the state-space is defined as  $\mathcal{X} = [0, R] \times [0, R] \times [0, C_T]$

The phase portrait for the two configuration version of the model is shown in Fig. 5.4. It is similar to the one configuration system, presented in Fig. 5.2,f as both concentrations of bound analyte  $x_1$  and  $x_2$  have the same relationship with the concentration  $C$  as  $x_1$  has in the single analyte model. The nullcline planes where  $x_1$  and  $x_2$  are constant include the lines in Fig. 2.5, and have an intersection that includes the same fixed point.

Figure 5.4 also shows the trajectory of the solution curve in cyan. When the system is initialised the concentration of the analyte at the surface of the chip rapidly increases

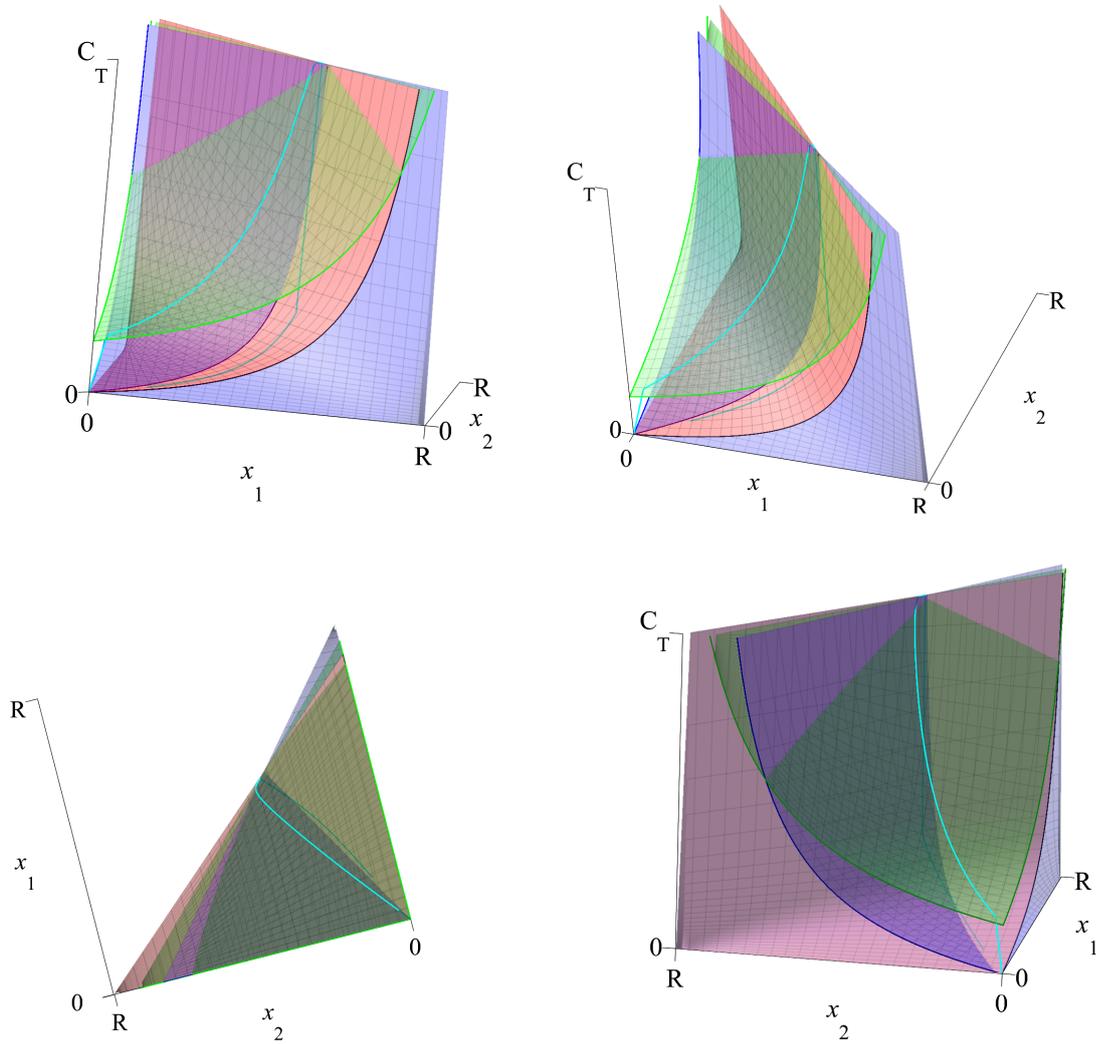
until the solution curve meets the  $C$  nullcline plane. The solution curve then moves along this plane in a similar way to the state-space solution of the two configuration Langmuir model. This movement has its direction parallel to the  $(x_1, x_2)$  plane initially determined by the association rate constants for either configuration's binding and being determined as it approaches the steady state by which of  $\frac{k_{a1}}{k_{d1}}$  and  $\frac{k_{a2}}{k_{d2}}$  is greater.

Figure 5.5 shows the continuing trajectory of the solution curve in cyan. When the concentration of the analyte being fed into the SPR machine drops from  $C_T$  to 0 the  $C$  null-cline plane moves so that it includes the origin. The concentration of antibody at the chip's surface  $C$  rapidly decreases until it rejoins this plane. The system then moves along this plane towards the origin. For the same reasons as the L2 model was adapted into the LA model, the **adjusted two concentration (LwTA)** model was developed from the LwT2 model; that is so that  $x_1$  and  $x_2$  could represent antibody antigen dimers and trimers respectively (see Eq. (5.2)). This was done by changing the capacity for addition antibody to bind in the LwT2 models state-space equation, Eq. 5.4, from  $R - x_1 - x_2$  to  $R - x_1 - 2x_2$ . This gave the LwTA models definition as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ h \frac{dC}{dt} = \sum_{i=1}^2 \left[ -k_{ai}C \left( R - \sum_{j=1}^2 jx_j \right) - k_{di}x_i \right] + k_m(I - C) \\ \frac{dx_1}{dt} = k_{a1}C \left( R - \sum_{i=1}^2 ix_i \right) - k_{d1}x_1 \\ \frac{dx_2}{dt} = k_{a2}C \left( R - \sum_{i=1}^2 ix_i \right) - k_{d2}x_2 \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0)^T \\ y(t, \mathbf{p}) = \alpha(x_1(t, \mathbf{p}) + x_2(t, \mathbf{p})), \end{cases} \quad (5.5)$$

where the parameter vector, parameter definitions and associated units are the same as the two configuration model.

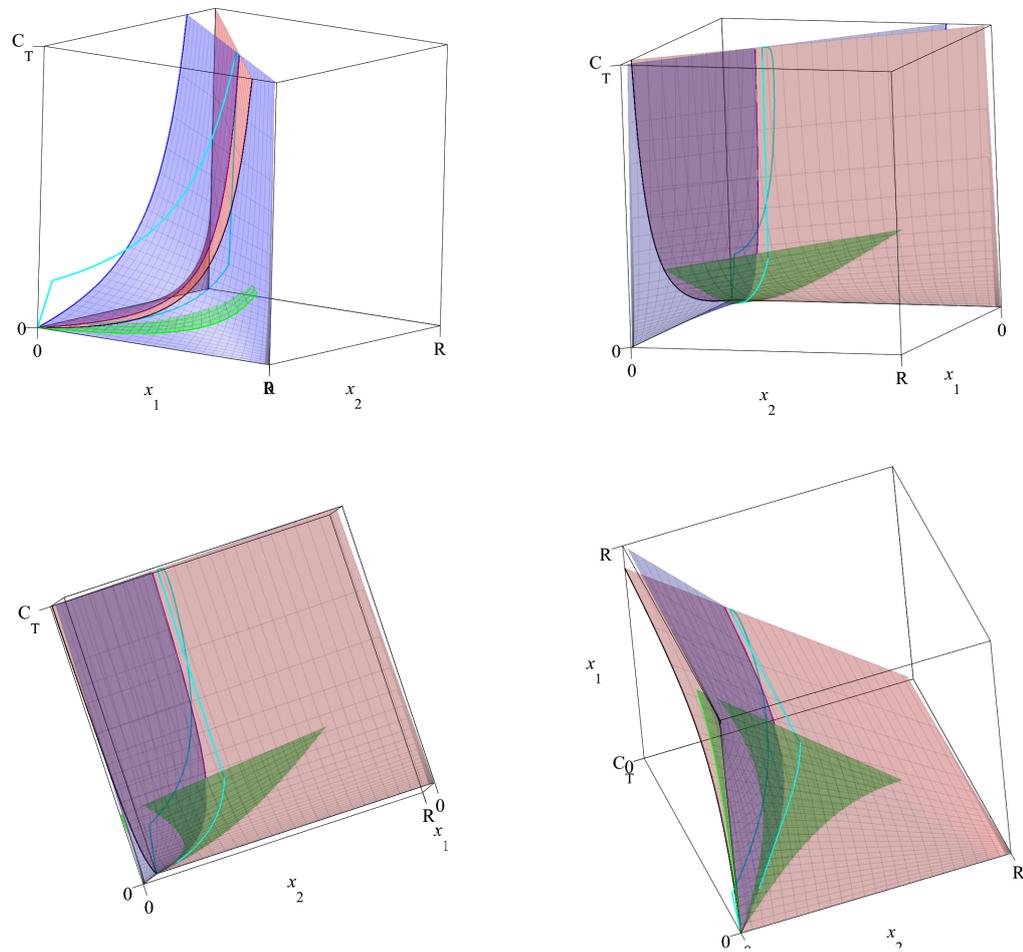
The phase plane for this new system is qualitatively the same as that for the two concen-



**Figure 5.4:** Visualisations of a solution trajectory for the Langmuir two concentration model in the association phase (cyan) alongside nullcline-planes  $x_1$  in blue,  $x_2$  in red,  $C_T$  in green

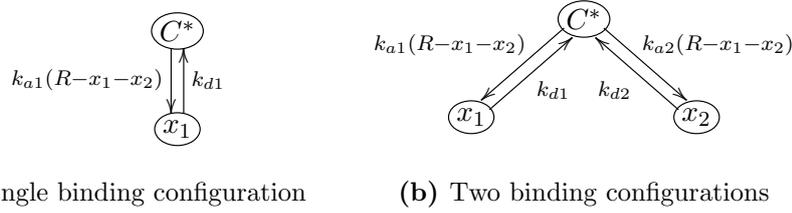
Parameter	$k_{a1}$ $M^{-1}s^{-1}$	$k_{a2}$ $M^{-1}s^{-1}$	$k_{d1}$ $s^{-1}$	$k_{d2}$ $s^{-1}$	$R$ $pg\ nm^{-2}$	$C_T$ $nM$	$k_m$ $pgnm^{-2}\ M^{-1}s^{-1}$	$h$ $pgnm^{-2}\ M^{-1}$
Value	6E-4	5E-4	7E-3	1E-2	500	100	$10^{-4}$	$10^{-6}$

**Table 5.2:** Values taken by parameters in Fig. 5.4 and 5.5



**Figure 5.5:** Visualisations of a solution trajectory for the Langmuir two concentration model in the dissociation phase (cyan) alongside nullcline-planes  $x_1$  in blue,  $x_2$  in red,  $C$  in green

tration Langmuir model. The major advantage of this adjusted form of the model is that, because it considers dimers and trimers, its assumptions better describe the experimental system.



**Figure 5.6:** Compartmental diagram of the ERC model with concentration of analyte at the reaction chip’s surface  $C^*$ , and concentrations of bound analyte  $x_1(t)$  and  $x_2(t)$ .

### 5.1.3 Effective rate constant approximation model

The effective rate constant approximation model for a **single binding configuration (ERC1)** was discussed in the same paper as the LwT1 model (Myszka *et al.*, 1998b), but introduced by Glaser (1993) and derived from fluid dynamics in Edwards *et al.* (1999).

The LwT model predicts that concentration of the analyte at the chip’s surface will quickly equilibrate (as shown in Fig 5.2). This model assumes it to be in quasi-steady state.

In more detail  $\frac{dC}{dt} = 0$  was substituted into Eq 5.3, and the rate of change equation for  $C$  was solved to obtain an expression for  $C$  that can be substituted into the other equations. This simplifies the model, eliminating a state variable  $C$  and the parameter  $h$ . Fig. 5.6(a) shows a schematic of this model.

The model has definition:

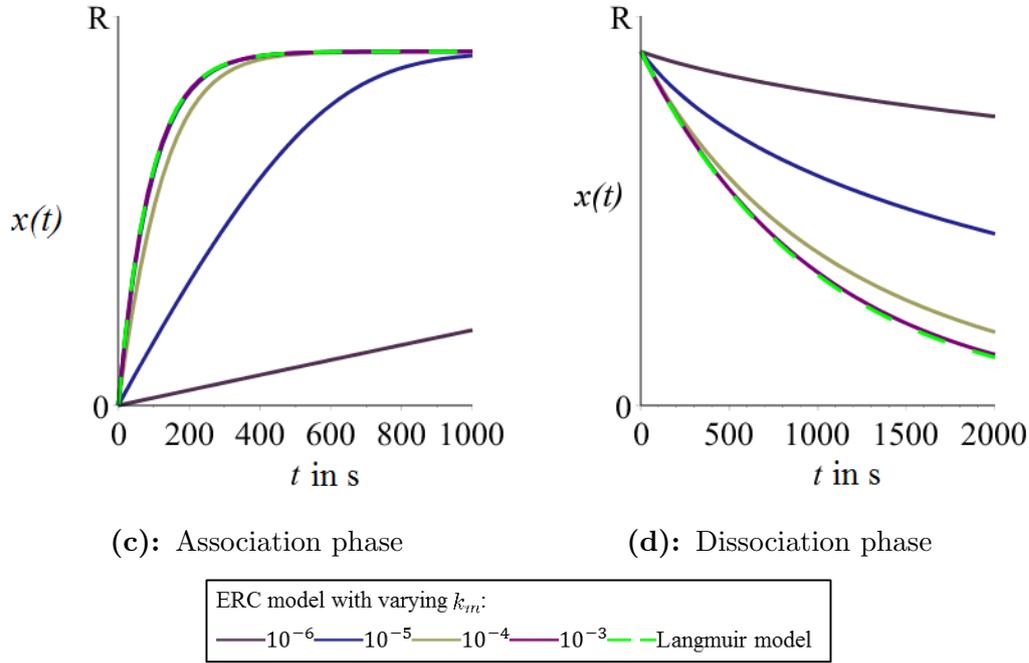
$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx}{dt} = \frac{k_a I (R - x) - k_d x}{1 + (k_a/k_m)(R - x)} \\ x(0; \mathbf{p}, I) = 0 \\ y(t, \mathbf{p}) = \alpha x(t, \mathbf{p}), \end{cases} \quad (5.6)$$

where the parameter vector is defined as  $\mathbf{p} = (k_a, k_d, C_T, k_m, R)$ , and the parameters and state variables have the same definitions and units as the single concentration LwT model (Subsection 5.1.2).

A state-space solution for this model differs from one of the single configuration Langmuir with transport model with equivalent parameters particularly at the beginning of either phase. In the LwT model, the concentration of analyte available to bind is zero at the beginning of the association phase, and in the initial moments of the experiment the concentration is moving towards its quasi-steady-state (see Fig. 5.3). As the ERC model assumes that the concentration of the analyte at the chip is already in a quasi-steady-state, it predicts binding from initialisation of the experiment. Similarly, in the dissociation phase, it assumes that the concentration at the chip's surface is already in quasi-steady state, so antibody begins dissociating as soon as the phase is initialised.

Figure 5.7 shows the effect of varying  $k_m$  on the model. As  $k_m$  becomes large, the denominator of the state-space equation becomes close to 1, and the state-space equation for the ERC model (in 5.6) tends towards that of the Langmuir model. In Fig. 5.7, this may be observed from the lack of distance between the dashed green line depicting a solution of the Langmuir model and the purple line depicting a solution of the ERC model with the same parameter values and a high  $k_m$ . The other curves also share these parameter values – but have differing  $k_m$  values. Varying  $k_m$  does not affect the stationary point. However it limits the rate of association predicted by the model, and as it becomes small makes the curves look straighter and shallower in the early part of the association phase.

Evans *et al.* (2013) extended the effective rate constant model to allow for analyte binding



**Figure 5.7:** Graphs of bound antibody  $x_1$  against time  $t$  for the Langmuir single configuration model (green dashed line) and the ERC model for different values of  $k_m$ .

Parameter	$k_a$ $M^{-1}s^{-1}$	$k_d$ $s^{-1}$	$R$ $pg\ nm^{-2}$	$C_T$ $nM$
Value	6E-4	7E-3	500	100

**Table 5.3:** Values taken by parameters in Fig. 5.7 that are not varied

in **two configurations**, the **ERC2** model. The model has definition:

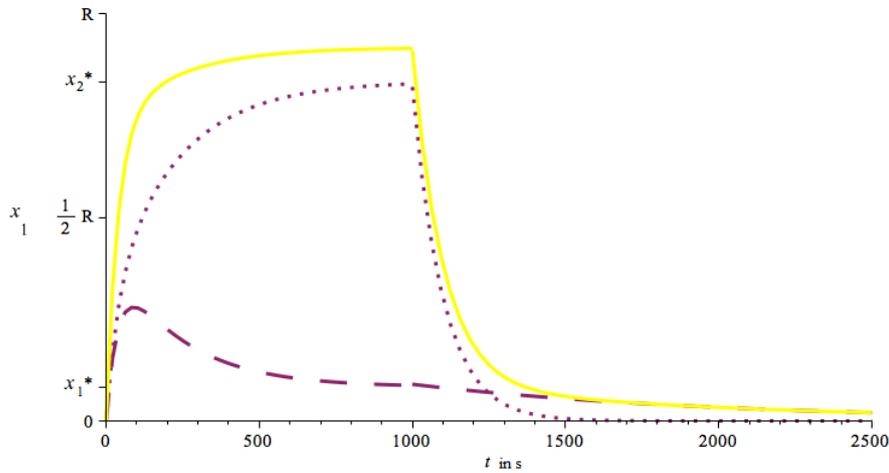
$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = \frac{k_{a1}I \left( R - \sum_{i=1}^2 x_i \right) - k_{d1}x_1}{1 + (k_{a1}/k_m) \left( R - \sum_{i=1}^2 x_i \right)} \\ \frac{dx_2}{dt} = \frac{k_{a2}I \left( R - \sum_{i=1}^2 x_i \right) - k_{d2}x_2}{1 + (k_{a2}/k_m) \left( R - \sum_{i=1}^2 x_i \right)} \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0)^T \\ y(t, \mathbf{p}) = \alpha \sum_{i=1}^2 x_i(t, \mathbf{p}), \end{cases} \quad (5.7)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, k_m, R)$ . The units for the parameters are shown in Fig. 5.4; all parameters are defined identically and measured with the same units as they were in the two configuration Langmuir with transport model (Subsection 5.1.2). The state variable and state-space are defined as they were in the two configuration Langmuir Model (example 2.2, Section 2.2).

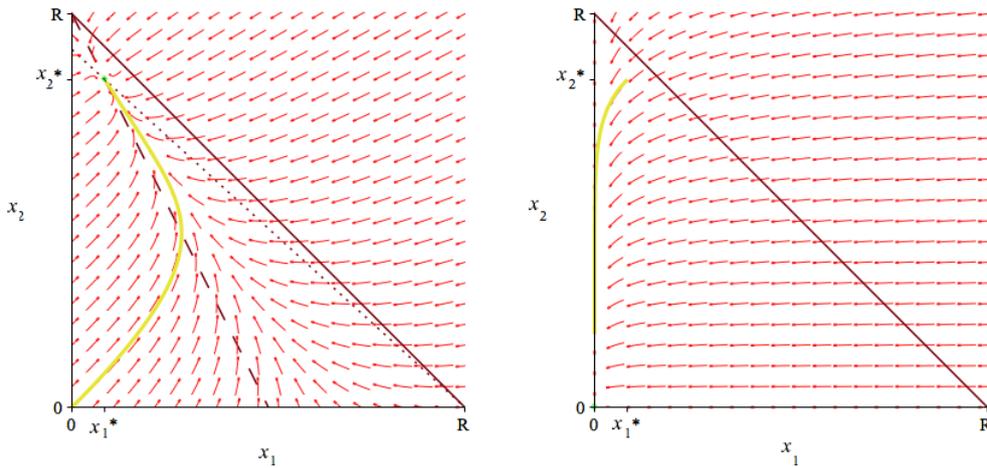
Phase portraits for either phase, as well as a visualisation of the state and output variables, are shown in Fig. 5.8. We may note that the nullclines and stationary points for this system, both in the association and dissociation phase are the same as those for both the Langmuir and Langmuir with transport models. However, the state variables in this system evolve along a separate trajectory to those of the Langmuir model, and with less speed because a constant concentration is not assumed. Their trajectories are qualitatively similar to those of  $x_1$  and  $x_2$  in the Langmuir with transport model.

In Fig. 5.8 (a) we see initially a rapid increase in the concentration of either configuration of dimer as well as the output signal. As the concentration of dimers attached to the chip increases so does the concentration of analyte monomers available to bind. This counterbalances the decrease in ligand monomers, allowing for a nearly constant rate of dimerisation. The output signal begins to slow because of a decrease in the concentration of free antigen, and at this point, the concentration of dimers of the first configuration crosses its nullcline and remains constant before beginning to decrease 5.8 (b). Whilst the concentration of dimers bound in the first configuration is decreasing, the output signal continues to increase because of increase in antigen bound in the second configuration.

With the parameters chosen, the system reaches a steady state just before the end of the association phase. In the dissociation phase, we see a rapid decrease in the number of dimers of either configuration. As the concentration of analyte monomers near the chip



(a): A visualisation of the state and output variables of the two concentration ERC model.  $x_1$  is shown as a dotted line,  $x_2$  as a dashed line and  $y/\alpha$  as a yellow line.



(b): Phase portrait for the association phase, with nullclines for  $x_1$  shown as a dotted line, for  $x_2$  as a dashed line. (c): Phase portrait for the dissociation phase, with nullclines for  $x_1$  shown as a dotted line, for  $x_2$  as a dashed line.

**Figure 5.8:** Phase portraits and visualisation of the two configuration ERC model.

Parameter	$k_{a1}$ $M^{-1}s^{-1}$	$k_{a2}$ $M^{-1}s^{-1}$	$k_{d1}$ $s^{-1}$	$k_{d2}$ $s^{-1}$	$R$ $pg\ nm^{-2}$	$C_T$ $nM$	$k_m$ $pgnm^{-2}\ M^{-1}s^{-1}$
Value	6E-4	5E-4	7E-3	1E-2	500	100	$10^{-4}$

**Table 5.4:** Values taken by parameters in Fig. 5.8.

is non-zero, this decrease is slower than for the Langmuir model (see Fig. 2.5) and non-exponential. Because of the faster rate of dissociation of dimers of the first type, their population quickly becomes negligible, 5.8 (c), and the output becomes dominated by the signal from the second configuration. Although the ERC1 model was derived by putting the LwT1 model in quasi-steady-state, the ERC2 model was derived from a model intended for mixed analytes, which will be discussed alongside polyclonal models in Section 7.1.2. This derivation was done by making the substitution  $C_{T1} = C_{T2} = C_T$  into Eq.7.6.

As a result a third model was developed by assuming the concentration of the analyte in the adjusted two configuration Langmuir with transport model was in quasi-steady-state, the **adjusted two configuration ERC model (ERCA)**. In more detail substituting  $\frac{dC}{dt} = 0$  into Eq. (5.13), and solving the rate of change equation for  $C$  to get an expression for  $C$  that can be substituted into the other equations.

A schematic of this model is shown in Fig. 5.6(b). The model is defined as:

$$\Sigma = \left\{ \begin{array}{l} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = \frac{\left( R - \sum_{i=1}^2 ix_i \right) k_{a1} (k_{d2}x_2 + k_{a1}x_1 + k_m I)}{\left( R - \sum_{i=1}^2 ix_i \right) (k_{a1} + k_{a2}) - k_m} - x_1 k_{d1} \\ \frac{dx_2}{dt} = \frac{\left( R - \sum_{i=1}^2 ix_i \right) k_{a2} (k_{d1}x_1 - k_{d2}x_2 + k_m I)}{\left( R - \sum_{i=1}^2 ix_i \right) (k_{a1} + k_{a2}) - k_m} - x_2 k_{d2} \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, \mathbf{p}) = \alpha \sum_{i=1}^2 x_i(t, \mathbf{p}), \end{array} \right. \quad (5.8)$$

The two configuration model is a simplification of this: where two extra assumptions are made. Firstly, rather than having one concentration of antibody at the surface of the chip (where the rates of association and dissociation to configurations of either type are in quasi-steady state) that both the association and dissociation rates for either configuration are in quasi-steady state. Secondly, that analyte bound in either configuration are only bound to a single antigen.

Plots of this model differ only slightly to the two configuration model. Notably, at the start of the dissociation phase, for the same parameters the rate of change, output signal and molecules of the second configuration, are lower. This is because the concentration of ligand available to bind in the second configuration is determined not just by the concentration of analyte already bound in the second configuration, but the concentration of analyte bound in either configuration.

Whilst this model shares its stationary point with the two configuration Langmuir, Langmuir with transport, and ERC models, it does not share its nullclines. Instead, its  $x_2$  nullcline in either phase is a shallow parabola passing through both the stationary point and  $(R, 0)$ .

Like the LwT2 model, the L2 Model is also a limit for the ERC2 and ERCA models as  $k_m$  tends towards infinity.

#### 5.1.4 Bivalent Langmuir model

The bivalent Langmuir (**BL**) model was first proposed in BIACORE (1998) and its use for antibody-antigen binding was demonstrated in Cooper & Williams (1999). Unlike the previous models it allows analyte-ligand dimers to trimerise and trimers to split into dimer and a single ligand; exactly the way we would expect antibody and antigen to interact. This model assumes the chemical equation:



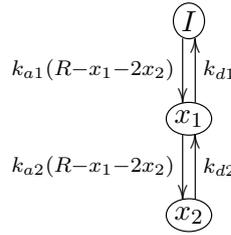
where  $A$ ,  $L$ ,  $AL$ , and  $LAL$  represent a molecule of the analyte of the experiment, a molecule of the ligand of the experiment, and a molecule of analyte bound to a molecule of ligand and a molecule of analyte bound to two molecules of the ligand respectively.

Like the Langmuir model this assumes the concentration of analyte at the chip's surface,  $C$  is fixed. As a result in this thesis it is referred to as the bivalent Langmuir model, to differentiate it from novel bivalent models that make other assumptions about the concentration of analyte at the chip's surface.

This model is defined as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = k_{a1}I \left( R - \sum_{i=1}^2 ix_i \right) - k_{d1}x_1 - k_{a2}x_1 \left( R - \sum_{i=1}^2 ix_i \right) + k_{d2}x_2 \\ \frac{dx_2}{dt} = k_{a2}x_1 \left( R - \sum_{i=1}^2 ix_i \right) - k_{d2}x_2 \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, \mathbf{p}) = \alpha \sum_{i=1}^2 x_i(t, \mathbf{p}), \end{cases} \quad (5.10)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R)$ , and all parameters except  $k_{a2}$  and  $k_{d2}$  are defined identically and measured with the same units as they were in the L2 model (example 2.2 in Section 2.2). The parameters  $k_{a2}$  and  $k_{d2}$  now represent the rates at which analyte - ligand dimers attach to additional ligand and the rate at which trimers split into analyte - ligand dimers and unattached ligand;  $k_{d2}$  continues to



**Figure 5.9:** Compartmental diagram of the bivalent Langmuir model with concentration of analyte at the reaction chip’s surface  $I$ , and concentrations of bound analyte  $x_1(t)$  and  $x_2(t)$ .

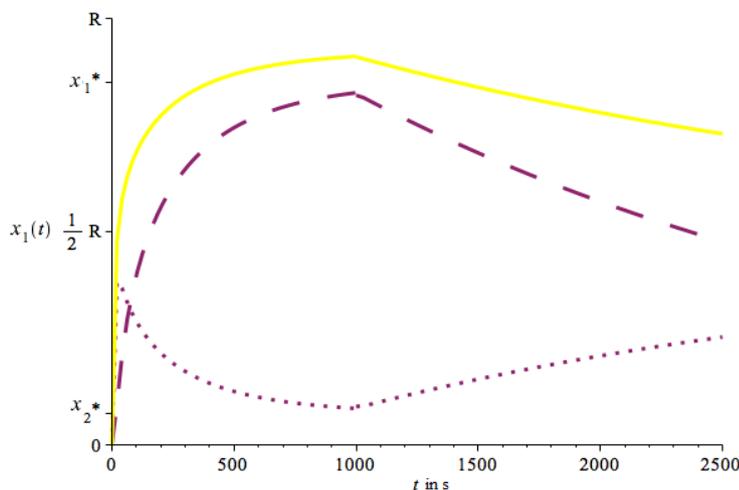
Parameter	$k_{a1}$ $\text{M}^{-1}\text{s}^{-1}$	$k_{a2}$ $\text{nm}^2\text{pg}^{-1}\text{s}^{-1}$	$k_{d1}$ $\text{s}^{-1}$	$k_{d2}$ $\text{s}^{-1}$	$R$ $\text{pg nm}^{-2}$	$C_T$ $\text{nM}$
Value	6E-4	1	7E-3	7E-3	500	100

**Table 5.5:** Values taken by parameters in Fig. 5.10.

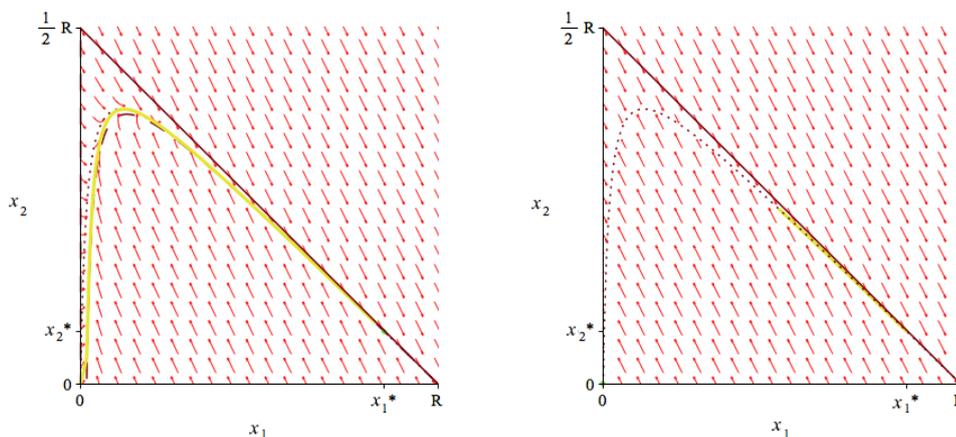
have the unit  $\text{s}^{-1}$ , but  $k_{a2}$  now has the unit  $\text{pg nm}^{-2}\text{s}^{-1}$ . The state variable  $(x_1, x_2)^\top = \mathbf{x}(t; \mathbf{p}, I) \in \mathcal{X}$ , represents the concentrations of monovalently and bivalently bound analyte in  $\text{pg nm}^{-2}$  and the state-space is defined as  $\mathcal{X} = [0, R] \times [0, \frac{1}{2}R]$ .

For experiments where the analyte is antibody and the ligand antigen, if we assume that, on binding to one antigen, the antibody does not go through any conformational changes that alter the affinity of its paratopes for the epitope of the antigen, we may understand  $k_{a2}$  as being  $k_{a1}$  multiplied by a constant representing the difference in available concentration of antigen on the chip’s surface and antigen in the flow. Because antigen is connected to the chip’s surface by linker, the antibody antigen dimer cannot move freely and encounter dimer outside of a hemisphere centering on the linker’s connection to the chip. This is illustrated by Fig. 5.1.

Again, for antibody analyte and antigen ligand, if we assume that the antibody has gone through no conformational changes that make it less likely to disengage from an antigen when it is engaged to two, then we may assume  $k_{d2} = 2k_{d1}$  - because it may disengage from either antigen, allowing for a reduced parameter vector  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, C_T, R)$ .



(a) A visualisation of the state and output variables of the bivalent Langmuir model.  $x_1$  is shown as a dotted line,  $x_2$  as a dashed line and  $y/\alpha$  as a yellow line.



(b) Association phase

(c) Dissociation phase

**Figure 5.10:** Phase portraits of the bivalent Langmuir model in the association and dissociation phase, with nullclines (dotted and dash lines).

A phase portrait and a visualisation of the state and output variables against time is shown in Fig. 5.10. Fig. 5.10(b) shows a phase portrait for the association phase, which is divided into three sections by its nullclines. Outside of the narrow region between the nullclines one of  $x_1$  and  $x_2$  is increasing and the other is decreasing, but between them there is a narrow region, which includes the origin in which both  $x_1$  and  $x_2$  increase.

When the experiment is initialised there is a rapid increase in the output,  $y$ , and concen-

tration of dimerised analyte,  $x_1$ , but no increase in the concentration of trimerised analyte,  $x_2$ . Within the first few seconds the dimers begin to form into trimers, and in Fig 5.10(b) we see the state variable move onto the  $x_1$  nullcline. It moves along the nullcline and  $x_2$  rapidly increases. In this period there is little change in the increase of  $y$ . As the increase in  $x_2$  slows, the system moves onto the  $x_1$  nullcline.  $x_2$  becomes constant and then begins to decrease. As  $x_2$  decreases,  $x_1$  increases and  $y$  continues to increase but less rapidly. The system pictured reaches a steady state just before the end of the association phase.

In the dissociation phase, the concentration of analyte being fed into the system,  $I$ , becomes zero. This moves the  $x_1$  nullcline, removing the region in which both variables can increase – this nullcline is now on the other side of the  $x_1$  nullcline – Fig. 5.10(c) but because of the scale on the figure, it is obscured by the  $x_2$  nullcline. At the start of the phase dimerically bound antibody dissociates. This leaves free antigen which is available for trimeric binding. As a result, the concentration of trimers increases and the system remains on the  $x_2$  nullcline.

If the system had not been at a steady state before the dissociation phase began, the concentration of monovalently bound analyte would rapidly decrease until the system was on this nullcline. This increase would involve analyte dissociating, leaving behind free ligand which in turn would increase the concentration of ligand available for analyte-ligand dimers to trimerise with. As time passes the concentration of trimers will increase as shown in Fig. 5.10, until it peaks, before finally beginning to decrease.

In Fig. 5.10(b) the state variable,  $x_2$  increases, at least initially, in the dissociation phase, as more ligand becomes available for binding. This is because a high value for the secondary binding constant  $k_{a2}$  was chosen to illustrate possible features of this model. It is unclear what range of values  $k_{a2}$  should be expected to take because it is not output by the software used with the machines (Bio-Rad, 2010) and as a result it is generally absent from papers in which the model is used (Jayaraman, 2008; Nielsen *et al.*, 2000; Huang *et al.*, 2006).

Additionally we would expect it to be influenced by factors other than the chemistry of the analyte and ligand, such as the density of the ligand and the radius that the analyte once bound to the ligand can move in (see Fig.5.1 ).

Figure 5.10(a) shows the concentration of bound antibody decreasing much slower in the dissociation phase than previous models. Previously Ong & Mattes (1993) established the problems of attempting to measure the affinity of bivalent antibody with non-SPR methods with models that assume monovalent binding. Investigating this model's asymptotic behaviour gives us a way to quantify the difference between this model and the previous ones, which will be useful for comparing parameter estimates generated by different models.

The concentration of monovalently bound analyte,  $x_1$  rapidly decreases, whereas the concentration of bivalently bound analyte,  $x_2$ , changes much slower, after a sufficiently large period of time has passed  $x_1$  may be treated as being in quasi steady state. Assuming  $\frac{dx_1}{dt} = 0$  allows for a simplification of Eq. (5.10) where an equation for  $x_1$  in terms of the parameters and  $x_2$  is found, and as a result eliminate  $x_1$  from the equation for  $x_2$ .

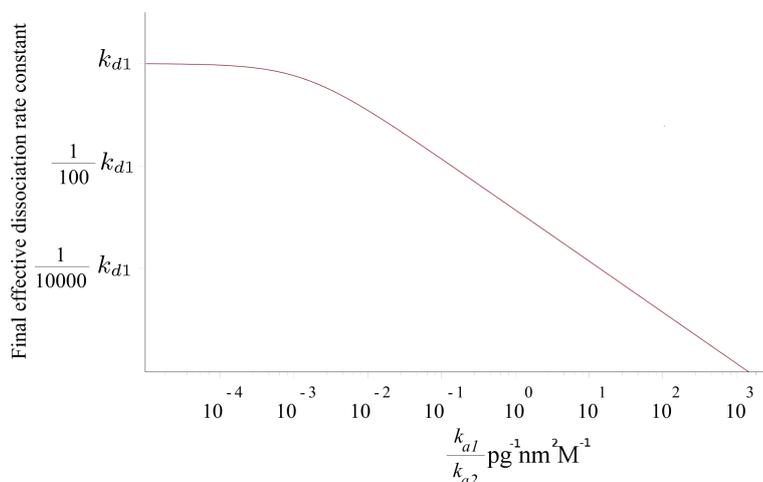
The equation for  $x_2$  can be then expanded as a Taylor series at the origin to give:

$$\frac{dx_2}{dt} = - \left( \frac{k_{d2}k_{d1}}{Rk_{a2} + k_{d1}} \right) x_2 + O(x_2^2) \quad (5.11)$$

As the quadratic and higher order terms tend towards zero faster than  $x_2$  itself, as  $x_2$  tends towards 0, and  $x_1$  is small, the model behaves increasingly like that given by the single configuration Langmuir model, but with a modified dissociation constant.

A graph of this modified dissociation constant as  $k_{a2}$  varies is in Fig. 5.11. It shows that if the rate monovalently bound antibody bind to additional antigen is slow enough, their dissociation is like that of molecules with only one binding domain, but that as this secondary binding rate increases, its relationship with the effective dissociation constant

becomes almost linear.



**Figure 5.11:** A logarithmically scaled graph of the effective dissociation rate constant as  $k_{a2}$  varies in proportion to  $k_{a1}$ .

Parameter	$k_{a1}$	$k_{a2}$	$k_{d1}$	$k_{d2}$	$R$	$C_T$
	$\text{M}^{-1}\text{s}^{-1}$	$\text{nm}^2\text{pg}^{-1}\text{s}^{-1}$	$\text{s}^{-1}$	$\text{s}^{-1}$	$\text{pg nm}^{-2}$	$\text{nM}$
Value	6E-4	-	7E-3	7E-3	500	100

**Table 5.6:** Values taken by parameters in Fig. 5.11.

### 5.1.5 Bivalent ERC model

Previously it was noted that the Langmuir model, when used for parameter estimation can introduce systematic errors in affinity estimation (Chaiken *et al.*, 1992), because it has been demonstrated that for monovalent experiments the rate constants estimated are influenced by the availability of both reactants and the effects of transport processes (Myszka *et al.*, 1998b). Additionally models that included the ERC approximation had been demonstrated to improve fits for both monoclonal and polyclonal antibody (see Subsections 6.4.1 and 8.1, and the paper H.A.J.Moyse *et al.* (2012)). This observation motivated the development of the **bivalent effective rate constant approximation (BERC)** model, that combined features of the ERC and BL models. It uses the chemical

equation 5.9, but assumes the concentration of analyte at the chip's surface,  $C$  varies, but is assumed to be in a quasi steady state.

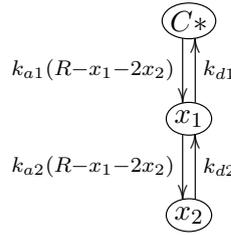
The ERC approximation was used rather than simply adding a transport equation because it removes the parameter  $h$ . This reduces the size of the parameter space, but more significantly fits made with the LwT were shown to be insensitive to changes in the parameter  $h$ , which was shown to be unidentifiable Evans *et al.* (2013).

This model is defined as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = \frac{k_{a1}I \left( R - \sum_{i=1}^2 ix_i \right) - k_{d1}x_1}{1 + \frac{k_{a1}}{k_m} \left( R - \sum_{i=1}^2 ix_i \right)} - k_{a2}x_1 \left( R - \sum_{i=1}^2 ix_i \right) + k_{d2}x_2 \\ \frac{dx_2}{dt} = k_{a2}x_1 \left( R - \sum_{i=1}^2 ix_i \right) - k_{d2}x_2 \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, \mathbf{p}) = \alpha \sum_{i=1}^2 x_i(t, \mathbf{p}), \end{cases} \quad (5.12)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, k_m, R)$ , and all parameters except  $k_{a2}$  and  $k_{d2}$  are defined identically and measured with the same units as they were in the LwT2 model (Subsection 5.1.2), and  $k_{a2}$  and  $k_{d2}$  are defined identically and measured with the same units as they were in the BL model (Subsection 5.1.4). The state variable and the state-space are also defined and measured with the same units as they were in in the BL model (Subsection 5.1.4).

Again, for antibody analyte and antigen ligand, if we assume that the antibody has gone through no conformational changes that make it less likely to disengage from an antigen



**Figure 5.12:** Compartmental diagram of the BERC model with concentration of analyte at the reaction chip's surface  $C^*$ , and concentrations of bound analyte  $x_1(t)$  and  $x_2(t)$ .

Parameter	$k_{a1}$ $M^{-1}s^{-1}$	$k_{a2}$ $nm^2pg^{-1}s^{-1}$	$k_{d1}$ $s^{-1}s^{-1}$	$k_{d2}$ $s^{-1}$	$R$ $pg\ nm^{-2}$	$C_T$ $nM$	$k_m$ $pgnm^{-2}\ M^{-1}s^{-1}$
Value	6E-4	1	7E-3	7E-3	500	100	$10^{-4}$

**Table 5.7:** Values taken by parameters in Fig. 5.13.

when it is engaged to two, than when it is engaged to one, then we may assume  $k_{d2} = 2k_{d1}$  - because it may disengage from either antigen, allowing for a reduced parameter vector  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, C_T, k_m, R)$ .

This model was derived by taking the state and output equations of the bivalent model, Eq. (5.10), and replacing the input variable  $I$  with a variable  $C$  that is defined by the differential equation:

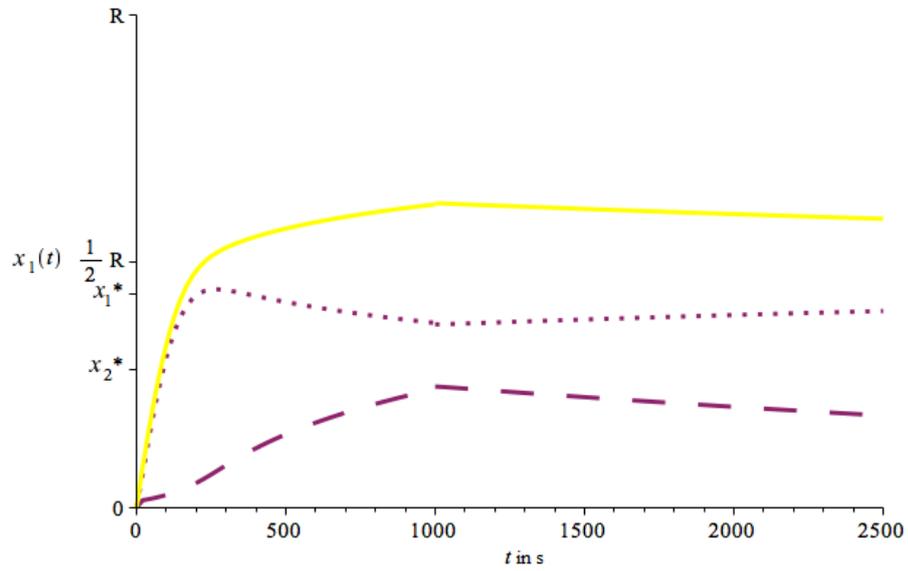
$$h \frac{dC}{dt} = -k_{a1}C \left( R - \sum_{i=1}^2 ix_i \right) + k_{d1}x_1 + k_m(I - C). \quad (5.13)$$

A nullcline for this variable was found:

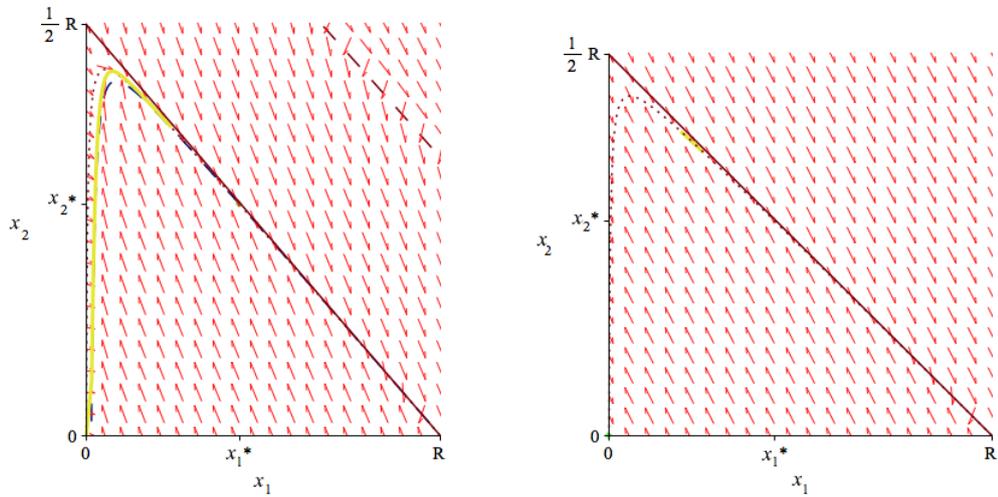
$$C^* = \frac{k_{d1}x_1 + k_m I}{k_{a1} \left( R - \sum_{i=1}^2 ix_i \right) + k_m}, \quad (5.14)$$

and  $C = C^*$  was substituted into the remaining equations, to get Eq. (5.12).

In Fig. 5.13 we see a plot of the output and state variables as well as the phase plane for this model in either phase. Here the nullclines behave near identically to those of the



(a) A visualisation of the state and output variables of the bivalent ERC model.  $x_1$  is shown as a dotted line,  $x_2$  as a dashed line and  $y/\alpha$  as a yellow line.



(b) Association phase

(c) Dissociation phase

**Figure 5.13:** Phase portraits and visualisation of the bivalent ERC model.

bivalent Langmuir model (Fig. 5.10). In the association phase they enclose an area in which both components of the state variable can increase; and in the dissociation phase the  $x_1$  nullcline moves to the other side of the  $x_2$  nullcline, so that the area they enclose is much smaller, and in it both state variables can decrease. Just like Fig. 5.10, this enclosed area is so small that the plot of the  $x_1$  nullcline is obscured by the  $x_2$  nullcline.

When the experiment is initialised there is a rapid increase in output and concentration of trimerised antibody, but much less increase in dimerised antibody. Similar to the bivalent model, the state variables remain on the nullclines. In this model the concentration of antibody able to bind, shown in Eq. 5.14, is always lower than  $C_T$  in the association phase, so for the same parameter values as the bivalent model, the system moves slower along the nullclines and does not reach its steady state in the association phase.

In the dissociation phase, there is little change in output signal. This is because the majority of antibody bound to the chip is trimerised, this system continues to move along the  $x_1$  nullcline towards the origin, very slowly.

### 5.1.6 Bivalent ERC model with spatial effects

In SPR experiments the ligand is attached to the surface of the chip via a series of reactions (Karlsson *et al.*, 1991; Myszkka, 1999) and as a result the ligand molecules are not placed homogeneously on the reaction chip's surface. Furthermore the reaction chip rather than being flat is fibrous (Richard B. M. Schasfoort, 2008; Marquart, 2013). For a model that includes bivalent binding this is significant because some ligand may be too far from other ligand for the analyte to bind to it bivalently.

Fits of the BL and BEREC model were able to produce significantly better fits than models that only allowed monovalent binding (see Fig.6.8(a) and 6.32). However they still produced estimates of dissociation rate constants that varied widely between experimental

replicates (see Fig. 6.10), and typically the model's best fits to the data performed worst at the end of the dissociation phase (see Fig. 6.9).

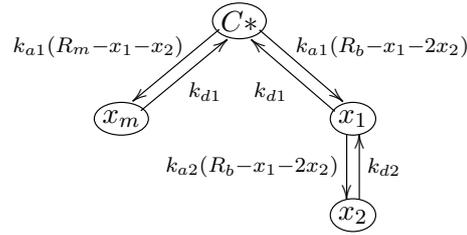
As a result the Bivalent ERC model with spatial effects (**BERCs**) was developed to include the most successful features of previous models, the effective rate constant approximation to the transport process and bivalent analyte binding as well as the effects of the spatial distribution of ligand.

Like the bivalent ERC model, it assumes that antibody may be bound in dimers and trimers, and that the concentration of analyte at the chip's surface,  $C$ , is in a quasi steady state, but also that some antigen are isolated. That is, that not all of the antigen are positioned on the chips surface so that when an antibody-antigen dimer is formed it will be able to trimerise.

This new model is defined as:

$$\Sigma = \left\{ \begin{array}{l} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = \frac{k_{a1} \left( R_b - \sum_{i=1}^2 ix_i \right) (Ik_m + k_{d1} (x_m + x_1))}{k_{a1} (R_m - x_m) + k_{a1} \left( R_b - \sum_{i=1}^2 ix_i \right) + k_m} - k_{d1} x_1 \\ -k_{a2} x_1 \left( R_m - \sum_{i=1}^2 ix_i \right) + k_{d2} x_2 \\ \frac{dx_2}{dt} = k_{a2} x_1 \left( R_b - \sum_{i=1}^2 ix_i \right) - k_{d2} x_2 \\ \frac{dx_m}{dt} = \frac{k_{a1} (R_m - x_m) (Ik_m + k_{d1} (x_m + x_1))}{k_{a1} (R_m - x_m) + k_{a1} \left( R_b - \sum_{i=1}^2 ix_i \right) + k_m} - k_{d1} x_m \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0, 0)^T \\ y(t, \mathbf{p}) = \alpha(x_1 + x_2 + x_m), \end{array} \right. \quad (5.15)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, k_m, R_m, R_b)$ , and all



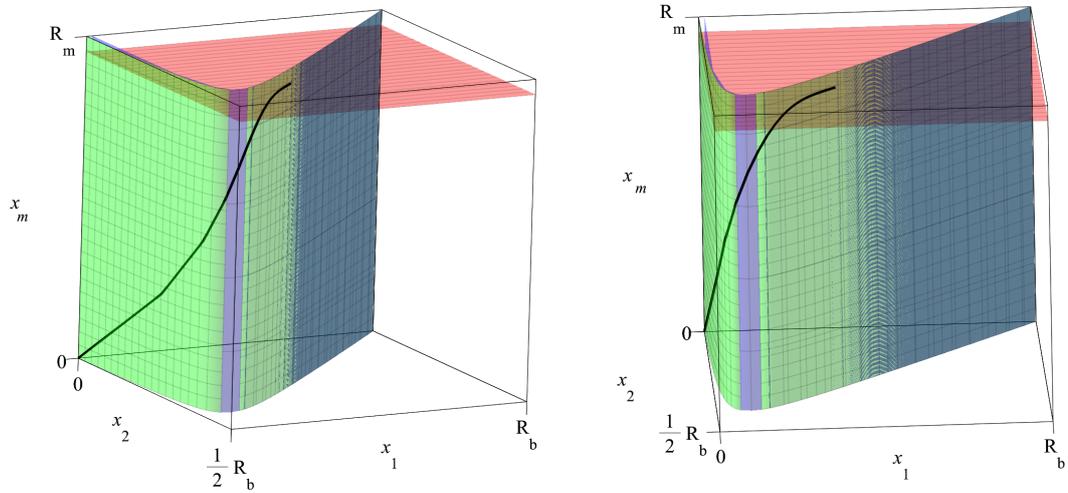
**Figure 5.14:** Compartmental diagram of the BERCs model with concentration of analyte at the reaction chip's surface  $C^*$ , and concentrations of bound analyte  $x_1(t)$ ,  $x_2(t)$  and  $x_m(t)$ .

parameters except  $R_m$  and  $R_b$  are defined identically and measured with the same units as they were in the BERC model (Subsection 5.1.5).  $R_m$  and  $R_b$  are the maximum density of analyte that can attach to the ligand to the sensor chip in  $\text{pg}^{-1} \text{nm}^2$ , on isolated and non-isolated ligand respectively. The state variable  $(x_1, x_2, x_m)^T = \mathbf{x}(t; \mathbf{p}, I) \in \mathcal{X}$ , represents the concentrations of monovalently and bivalently bound analyte attached to ligand located so that bivalent binding would be possible, and monovalently bound analyte attached to ligand located where it would be impossible in  $\text{pg} \text{nm}^{-2}$ , and state-space is defined as  $\mathcal{X} = [0, R_b] \times [0, \frac{1}{2}R_b] \times [0, R_m]$ .

Similar to the previous two models, for antibody-analyte antigen-ligand experiments, the parameter vector may be reduced. If we assume that the antibody has gone through no conformational changes that make it less likely to disengage from an antigen when it is engaged to two, than when it is engaged to one, then we may assume  $k_{d2} = 2k_{d1}$  - because it may disengage from either antigen, allowing for a reduced parameter vector  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, C_T, k_m, R_m, R_b)$ , this is referred to as the BERCs simplified Model.

The derivation of this model begins with the bivalent model (Eq. (5.10)). The parameter  $R$  was relabeled  $R_b$  and the variable  $I(t)$  was replaced with  $C(t)$ . A new variable  $x_m$  was added, with the statespace equation:

$$\frac{dx_m}{dt} = k_{a1}I(R_m - x_m) - k_{d1}x_m, \quad (5.16)$$



**Figure 5.15:** Visualisations of a solution trajectory for the Langmuir two concentration model in the association phase (black) alongside nullcline-planes  $x_1$  in blue,  $x_2$  in green,  $x_m$  in red.

Parameter	$k_{a1}$ $M^{-1}s^{-1}$	$k_{a2}$ $nm^2pg^{-1}s^{-1}$	$k_{d1}$ $s^{-1}$	$k_{d2}$ $s^{-1}$	$R_b$ $pg\ nm^{-2}$	$R_m$ $pg\ nm^{-2}$	$C_T$ nM	$k_m$ $pgnm^{-2}\ M^{-1}s^{-1}$
Value	6E-4	1	7E-3	7E-3	400	100	100	$10^{-4}$

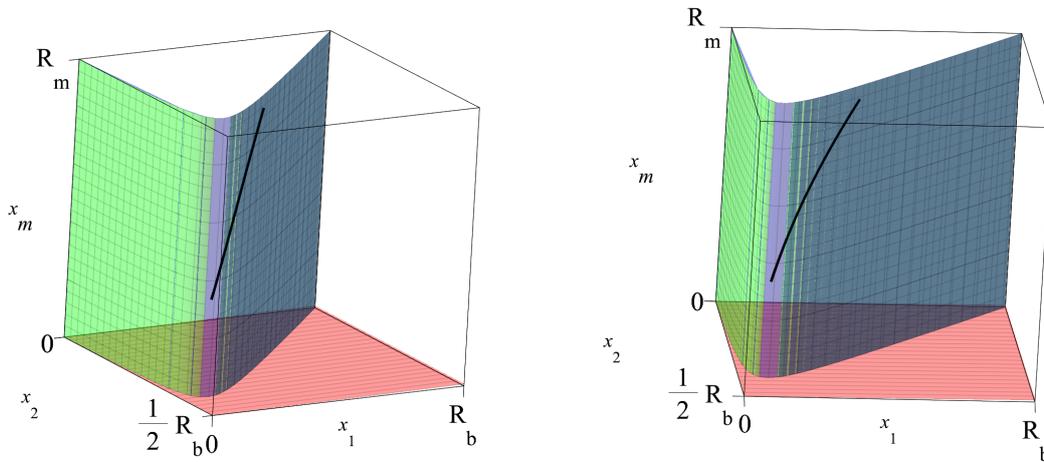
**Table 5.8:** Values taken by parameters in Fig. 5.15-5.17.

and the output equation was changed to  $y = \alpha(x_1 + x_2 + x_m)$ . In the resulting statespace equations the input variable  $I$  was replaced with the variable  $C$ , defined by the differential equation:

$$h \frac{dC}{dt} = -k_{a1}C \left( R_b - \sum_{i=1}^2 ix_i \right) + k_{d1}x_1 - k_{a1}C(R_m - x_m) + k_{d1}x_m + k_m(I - C) \quad (5.17)$$

A nullcline for this variable was found,  $C^*$ , and  $C = C^*$  was substituted into the remaining equations. The resulting system of equations is that given in Eq. (5.15)

Figures 5.15 and 5.16 show visualisations of solution trajectories and nullclines for both phases. These are best understood alongside figures giving similar diagrams for similar

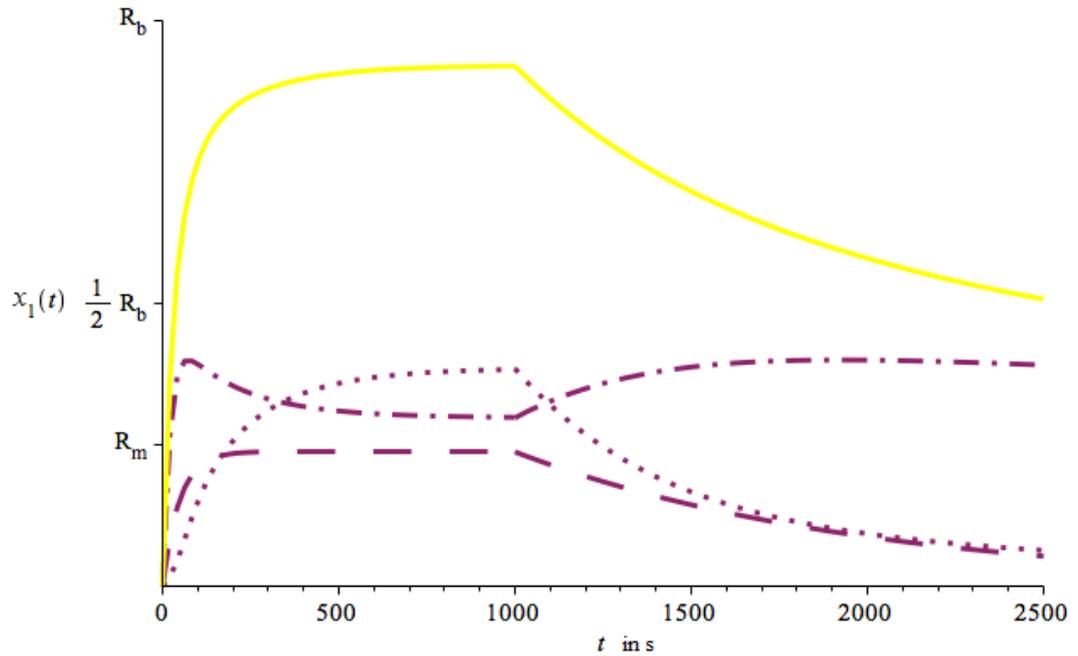


**Figure 5.16:** Visualisations of a solution trajectory for the Langmuir two concentration model in the dissociation phase (black) alongside nullcline-planes  $x_1$  in blue,  $x_2$  in green,  $x_m$  in red.

models like Fig. 5.13. Previously the association phase was divided into three areas by two curved nullclines. The bivalent ERC spatial model has two nullcline planes of similar shape also dividing phase space vertically into three equivalent regions – a central one, including the origin in which both  $x_1$  and  $x_2$  increase – which our solution will travel through; as well as regions on either side of the nullclines where the solution curve cannot travel. In addition to this there is a third nullcline- a nearly horizontal plane above which  $x_m$  decreases, and below which  $x_m$  increases. This narrow region is best seen by looking at the top left corner of both graphs in Fig. 5.15, where it is seen that there is a narrow gap between the corners of the green and blue planes.

Figure 5.16 shows that in the dissociation phase the  $x_m$  nullcline is translated to include the origin. The  $x_1$  nullcline now also includes the origin. The narrow region where both  $x_1$  and  $x_2$  could increase is eliminated. This is best seen by looking at the top left corner of both graphs in Fig. 5.16, there is no narrow gap between the corners of the green and blue planes.

In both Fig. 5.15 and Fig. 5.16 a black solution trajectory is illustrated. However it is



**Figure 5.17:** A visualisation of the state and output variables of the bivalent BERCs model.  $x_1$  is shown as a dash-dotted line,  $x_2$  as a dotted line,  $x_m$  as a dashed line and  $y/\alpha$  as a yellow line.

also useful to display it with a separate graph, see Fig. 5.17.

In Fig. 5.17 we see a corresponding graph of the state and output variables of the model against time. When the experiment is initialised there are rapid changes in the output variable and the concentration of bivalently bound antibody and antibody bound monovalently to isolated ligands. The concentration of bivalently bound antibody peaks and then begins to decrease, at the same time the concentration of antibody monovalently bound to isolated ligand approaches equilibrium. Output signal continues to increase as does the concentration of antibody monovalently bound to non-isolated ligand. As time continues, they all approach their steady states.

In the dissociation phase, rapid decreases in both the concentration of antibody bound to isolated and non-isolated ligand result in a rapid change in output. Whilst the overall concentration of antibody bound to non-isolated ligand is decreasing the concentration of

antibody bound bivalently increases.

This model gave the best fits of any to the monoclonal antibody experiments, except for datasets from the SN230G6 vs HLA -A2 experiments (see Subsection 6.4.3), which had systematic errors shown in Fig. 6.16. Typically these fits featured much more rapid dissociations in the early part of the dissociation phase than the data showed, and estimates of  $k_{d1}$  which were much higher than for other antibody, and were much more scattered, ranging across 5 orders of magnitude, compared to two for other antibody (see Fig. 6.14).

The two explanations for this that were considered were experimental error and a more complex binding process. In the models discussed so far antibody binds instantaneously to antigen, and the strength of this bond never changes. However, the assumption that the strength of the bond is time invariant may be unhelpful. Antibody-antigen binding affinity is largely determined by electrostatic forces and hydrogen bonds (Murphy, 2001). As the antibody and antigen may go through conformational changes throughout the binding process the dissociation rate constants could vary with time.

Creating a new model using delay equations for these datasets was considered, but as a stepping stone two variations of the BERCs model were developed and used. Unlike previous models, these were not intended to give a reductive explanation for features of the data but to explore the features of the data that were causing the systematic errors of the BERCs model.

The **BERCs extended** model is defined identically except with the parameter vector  $\mathbf{p} = (k_{a1}, k_{d11}, k_{d12}, k_{a2}, k_{d21}, k_{d22}, C_T, k_m, R_m, R_b)$ , and with the following two substitutions

made into Eq. (5.15):

$$\begin{aligned}
 k_{d1} &= \begin{cases} k_{d11} & : t \in T_a \\ k_{d12} & : t \in T_d \end{cases} \\
 k_{a2} &= \begin{cases} k_{d21} & : t \in T_a \\ k_{d22} & : t \in T_d. \end{cases}
 \end{aligned} \tag{5.18}$$

This model was expected to produce dissociation rate constant estimates in the association phase that were affected by the fact that many antibody-antigen and antigen-antibody-antigen structures would not have undergone conformational changes in the early part of the phase and have undergone them in the later part of the phase. In the dissociation phase, it was expected to produce dissociation rate constants that were affected by nearly all bonds being older.

This model saw improvements in both the fit for the SN230G6 vs HLA -A2 datasets (see Fig. 6.18), and range of dissociation rate constant estimates (see Fig. 6.17(a)).

The **second extended BERCS** model was developed to further investigate the possibility of these features of the data arising from slow conformational changes in the antibody and antigen occurring upon binding. As well as letting the dissociation rate constants between phases vary it let the variables  $x_1$  and  $x_2$  vary discontinuously at the transition between the phases ( $t_1$ ). This was because if the dissociation rate constants were changing in time, we would expect that the ratio of monovalently to bivalently bound antibody to be incorrect at the start of the dissociation phase. In the dissociation phase as nearly all complexes would have undergone whichever conformational changes they were going through, if at the start of the phase if the variables took the right values fits were expected to be good and parameter estimates consistent.

This model is defined as:

$$\Sigma = \left\{ \begin{array}{l} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = \frac{k_{a1} \left( R_b - \sum_{i=1}^2 ix_i \right) (Ik_m + k_{d1}(t) (x_m + x_1))}{k_{a1} (R_m - x_m) + k_{a1} \left( R_b - \sum_{i=1}^2 ix_i \right) + k_m} - k_{d1}(t)x_1 \\ \frac{dx_2}{dt} = \frac{-k_{a2}x_1 \left( R_m - \sum_{i=1}^2 ix_i \right) + k_{d2}(t)x_2}{k_{a2}x_1 \left( R_b - \sum_{i=1}^2 ix_i \right) - k_{d2}(t)x_2} \\ \frac{dx_m}{dt} = \frac{k_{a1} (R_m - x_m) (Ik_m + k_{d1}(t) (x_m + x_1))}{k_{a1} (R_m - x_m) + k_{a1} \left( R_b - \sum_{i=1}^2 ix_i \right) + k_m} - k_{d1}(t)x_m \\ k_{d1}(t) = \begin{cases} k_{d11} & : t \in T_a \\ k_{d12} & : t \in T_d \end{cases} \\ k_{a2}(t) = \begin{cases} k_{d21} & : t \in T_a \\ k_{d22} & : t \in T_d, \end{cases} \\ \mathbf{x}(0; \mathbf{p}, I) = (0, 0, 0)^\top \\ y(t, \mathbf{p}) = \alpha(x_1 + x_2 + x_m), \end{array} \right. \quad (5.19)$$

where the parameter vector is  $\mathbf{p} = (k_{a1}, k_{d11}, k_{d12}, k_{a2}, k_{d21}, k_{d22}, C_T, k_m, R_m, R_b, b_1, b_2)$ , and all parameters are defined as they were with the **BERCs extended** model, except  $b_1$  and  $b_2$  which are the initial values of  $x_1$  and  $x_2$  in the dissociation phase.

This model resulted in continued improvements in fits (See Fig. 6.18) and distribution of dissociation rate constant estimates (See Fig. 6.17(b)). However, when more experiments were conducted on the same interaction (see Fig. 4.9), and the BERCs simplified model was fitted to them their fits had none of the systematic errors that existed in the sensogram fits that motivated these models (see 6.6). As a result a model including delay equations

was never developed.

## 5.2 Equilibrium analysis and comparisons between the behaviour of models as parameters vary

In the SPR literature equilibrium analysis is often presented as an alternative to the use of a model (Myszka *et al.*, 1998a; BIACORE, 1997; Hadzhieva *et al.*, 2017a). In effect, equilibrium analysis assumes the L1 model and uses a single time point rather than a time series as the basis for parameter estimation.

This may be extended for other models, and this discussion allows us to compare their equilibria.

The output at steady state in the association phase is the same for the single configuration Langmuir, Langmuir with transport, and ERC models. This is given by the equation:

$$y = \alpha \frac{k_{a1} C_T R}{k_{a1} C_T + k_{d1}}. \quad (5.20)$$

A similar equation can be derived for the bivalent and bivalent ERC models

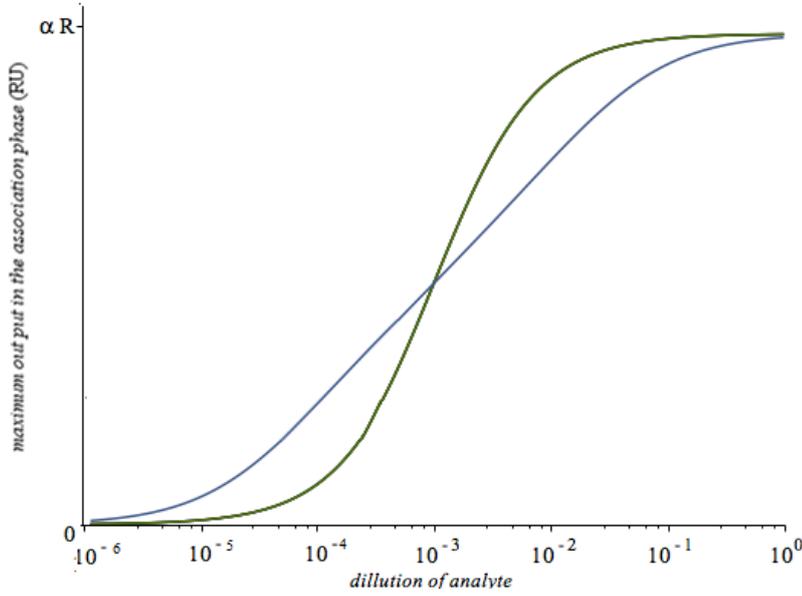
$$y = \frac{1}{2} \alpha R + \frac{1}{8} \alpha \frac{C_T^2 k_{a1}^2 k_{d2} + Q_1 C_T k_{a1} + k_{d1}^2 k_{d2} - Q_1 k_{d1}}{C_T k_{a1} k_{a2} k_{d1}}, \quad (5.21)$$

where

$$Q_1 = \sqrt{k_{d2}} \sqrt{8 R C_T k_{a1} k_{a2} k_{d1} + C_T^2 k_{a1}^2 k_{d2} + 2 C_T k_{a1} k_{d1} k_{d2} + k_{d1}^2 k_{d2}}. \quad (5.22)$$

A further equation can be derived for the bivalent ERC model with spatial effects

$$y = \alpha \frac{2 k_{a1} C_T R}{2 k_{a1} C_T + k_{d1}} + \frac{1}{2} \alpha \frac{Q_2}{k_{a2} k_{d1}} - \frac{1}{4} \alpha \frac{Q_2 (Q_2 - 2 k_{a2} R_b k_{d1})}{k_{a2} k_{d1} (Q_2 + 2 k_{d1} k_{d2})} \quad (5.23)$$



**Figure 5.18:** Output at steady state in association phase against analyte concentration, green for Langmuir, Langmuir with transport and ERC models, and blue for the bivalent and bivalent ERC model.

Parameter	$k_{a1}$ $M^{-1}s^{-1}$	$k_{a2}$ $nm^2pg^{-1}s^{-1}$	$k_{d1}$ $s^{-1}$	$k_{d2}$ $s^{-1}$	$R$ $pg\ nm^{-2}$	$C_T$ $nM$	$k_m$ $pgnm^{-2}\ M^{-1}s^{-1}$	$h$ $pgnm^{-2}\ M^{-1}$
Value	6E-4	1	7E-3	7E-3	500	1000	$10^{-4}$	$10^{-6}$

**Table 5.9:** Values taken by parameters in Fig. 5.10.

where

$$Q_2 = -2C_T k_{a1} k_{d2} - k_{d1} k_{d2} + \sqrt{k_{d2} \sqrt{8R_b C_T k_{a1} k_{a2} k_{d1} + 4C_T^2 k_{a1}^2 k_{d2} + 4C_T k_{a1} k_{d1} k_{d2} + k_{d1}^2 k_{d2}}}. \quad (5.24)$$

To compare the steady states different models predict in the association phase at different dilutions of analyte a new unitless parameter is introduced,  $d$ , representing the dilution of the sample concentration, and  $C_T$  is replaced by  $dC_T$  in Eq. (5.20)-5.24.

In Fig. 5.10 we see a graph of the output at the steady state for the Langmuir, Langmuir with transport, ERC, bivalent, and bivalent ERC models. These curves are sigmoidal,

having asymptotes at 0 and  $\alpha R$ . Notably these curves are identical for each of the Langmuir, Langmuir with transport and ERC models, and identical for the bivalent, and bivalent ERC models.

This analysis has two primary results. Firstly, it demonstrates that if the output function of our models is replaced by the output they had at their steady states then the Langmuir, ERC and Langmuir with transport are all indistinguishable; as are the bivalent and bivalent ERC models. This could be significant for research being done that uses equilibrium analysis solely to investigate antibody binding within SPR experiments.

Secondly it indicates what we are to expect in terms of the effects of doing experiments with multiple dilutions of the same analyte and the sensitivity of the output at steady state to dilution.

### **5.2.1 Extending models for parallel experiments with multiple antibody concentrations and antigen densities**

In many cases SPR platforms allow for multiple experiments to be conducted in parallel. A common practice is to conduct them with varying analyte concentrations and ligand density. Chapter 4 showed sensorgrams from 75 SPR experiments using this protocol.

We have already discussed varying concentration in part in the previous section. However, equilibrium analysis only involves the fixed points of the association phase. By discussing how the output solutions of the models change as the parameters relating to concentration of analyte  $C_T$  and density of ligand  $R$  change allows us to better understand these sensorgrams.

The effect of varying  $C_T$  and  $R$  may be illustrated by considering the dependency of the rate of change of output at the beginning of the association phase and the dissociation phase and the steady state on these parameters. The rate of change of the output at the

initialisation of the association phase can be obtained for each model by substituting the initial conditions of the models into their system equations, and substituting the result into their output equations. These rates of change of output are given by:

$$\frac{dy}{dt} \Big|_{t=0} = \begin{cases} 0 & \text{for the Langmuir with transport model} \\ \alpha k_{a1} C_T R & \text{for the Langmuir and bivalent Langmuir model} \\ \alpha \frac{k_{a1} C_T R}{k_{a1} R + k_m} & \text{for the ERC and bivalent ERC model} \end{cases} \quad (5.25)$$

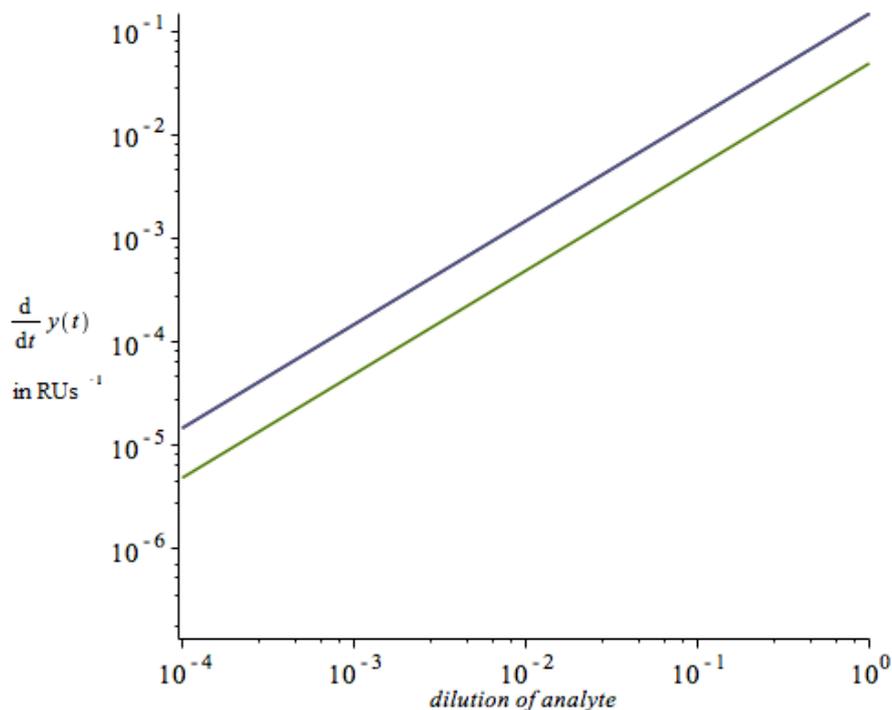
For the sake of brevity, in the above only the single configuration versions of the models are considered. The values taken by these equations are shown in Fig. 5.19. The rate of change of the output at the initialisation of the association phase is proportional to  $C_T$  for all models except for the Langmuir with transport model. It is proportional to  $R$  for both the Langmuir and bivalent Langmuir models and determined by the same quotient of linear functions of  $R$  for the ERC and bivalent ERC models. This is illustrated in Fig. 5.19.

The dependency of the output at a steady state may be viewed in Eq. (5.20), and in Fig. 5.10. Viewing these together with Eq. (5.25) and Eq. (5.19) gives a more complete picture of how the output of the models varies as these parameters vary.

The rate of change of output at the initialisation of the association phase is proportional to  $R$  for the Langmuir and bivalent Langmuir models. It is proportional to a quotient of linear functions of  $R$  for the ERC and bivalent ERC models.

Whilst the initial rate of change of output in the association phase is unaffected by whether a model is bivalent, its output at steady state is, as we have seen in Fig. 5.10.

In Fig. 5.22 we see the rates of change of the output of the models as their dissociation phases begin. These curves are also sigmoidal and they share asymptotes at 0 and  $-\alpha k_{d1} R$ , but they cross over each other. Their equations are omitted for the sake of brevity, but



**Figure 5.19:** Initial rate of change of output variable in the association phase against analyte concentration, blue for the Langmuir and bivalent Langmuir model, green for the ERC and bivalent ERC model.

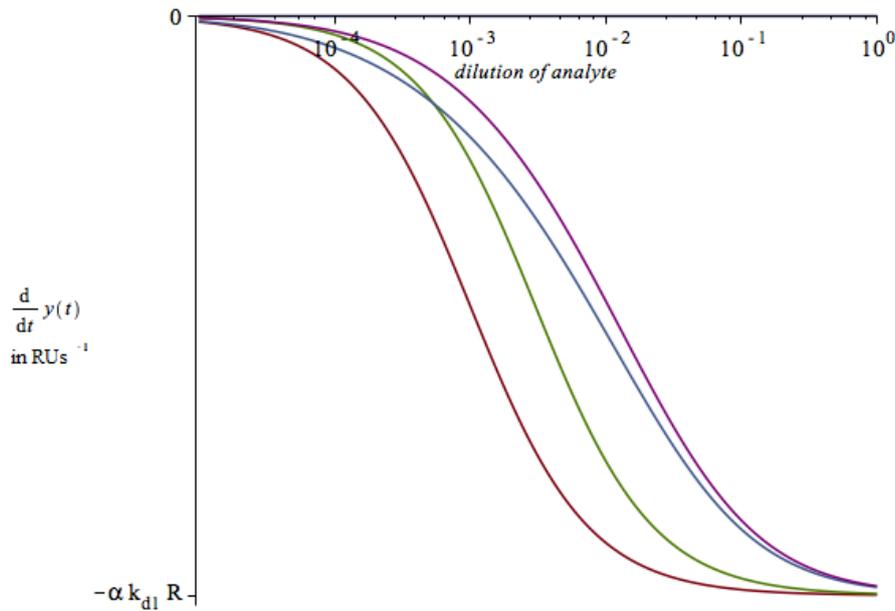
Parameter	$k_{a1}$ $M^{-1}s^{-1}$	$k_{a2}$ $nm^2pg^{-1}s^{-1}$	$k_{d1}$ $s^{-1}$	$k_{d2}$ $s^{-1}$	$R$ $pg\ nm^{-2}$	$C_T$ $nM$	$k_m$ $pgnm^{-2}\ M^{-1}s^{-1}$	$h$ $pgnm^{-2}\ M^{-1}$
Value	6E-4	1	7E-3	7E-3	500	1000	$10^{-4}$	$10^{-6}$

**Table 5.10:** Values taken by parameters in Fig. 5.19 and 5.22 .

can be derived with maple code presented in Section A.4.

The Langmuir model predicts the greatest rate of change of output, of any of the models at each concentration, the bivalent ERC model predicts the lowest rate of change of output of any of the models at any concentration. The bivalent Langmuir for low dilutions offers a higher rate of change of output than the ERC model, but as dilutions increase the ERC model overtakes it.

Variation in  $y$  at the steady state of the association phase and  $\frac{dy}{dt}$  evaluated at the start of



**Figure 5.20:** Initial rate of change of output variable in the dissociation phase where state variables had been in a steady state against analyte concentration, red for the Langmuir, blue for the bivalent Langmuir, green for the ERC and magenta bivalent ERC.

either phase in response to variation in  $R$  is approximately linear. However, having noted the above, we can understand how the output from the models varies when the parameters  $C$  and  $R$  vary. These observations give us theoretical insight into why we see variation in between multiple curves within the same sensorgram in Chapter 4.

### 5.3 Identifiability and Indistinguishability

The concepts of structural identifiability and indistinguishability were introduced in sections 3.1 and 3.3. Due to the nature of SPR experiments the models we are interested in analysing have complex non linearities and are employed in parallel, as a result the methods found in the literature generally had to be adapted. In section 3.1 new concepts relating to identifiability were developed, and more new tools were introduced for their analysis. These tools and concepts will allow us to build a full picture of which parameters can be

identified by which model and will due to the new concepts and tools help in designing experiments with improved identifiability.

For some of the models analysed here an analysis was already presented in Evans *et al.* (2013). In this section the analysis of these models is extended with the same definitions and techniques that allow for the analysis of the more complex models and will later lead to improved experiment design.

### 5.3.1 Langmuir model

The **single configuration Langmuir (L1) model** was used to illustrate the generic Taylor series method for identifiability analyses in Example 3.1. It was found that of its four parameters  $k_{a1}$ ,  $k_{d1}$ ,  $C_T$  and  $R$  the parameters  $R$  and  $k_{d1}$  are identifiable, as is the product  $k_{a1}C_T$ - although  $k_{a1}$  and  $C_T$  components are simply structurally unidentifiable. As a result in the experiments the Langmuir model could be applied to where  $C_T$  is known (Sections 4.3 - 4.4) , the model is structurally globally identifiable. In experiments where  $C_T$  was unknown (Sections 4.5 - 4.9) the model is simply structurally unidentifiable, as neither  $k_{a1}$  and  $C_T$  can be estimated.

This identifiability equivalence class for this function was shown in Fig. 3.2, and the properties of the mappings that this arises from is shown in Fig. 3.1.

The identifiability of the **two concentration Langmuir (L2) model** was analysed in Examples 3.2, 3.4, 3.6 and 3.9. The model has six parameters:  $k_{a1}$ ,  $k_{d1}$ ,  $k_{a2}$ ,  $k_{d2}$ ,  $C_T$  and  $R$ . These analyses showed that for two models to give the same output for a generic input, their parameters must be related by either  $(k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R) = (\bar{k}_{a1}/\gamma, \bar{k}_{d1}, \bar{k}_{a2}/\gamma, \bar{k}_{d2}, \bar{C}_T\gamma, \bar{R})$  or  $(k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R) = (\bar{k}_{a2}/\gamma, \bar{k}_{d2}, \bar{k}_{a1}/\gamma, \bar{k}_{d1}, \bar{C}_T\gamma, \bar{R})$  for  $\gamma = \bar{C}_T/C_T$ . In summary, the parameter  $R$  is globally identifiable; the products  $k_{a1}C_T$ ,  $k_{a2}C_T$  and the parameters  $k_{d1}$  and  $k_{d2}$  are locally identifiable; and the parameters  $k_{a1}$ ,

$k_{a2}$ ,  $C_T$  are unidentifiable. As a result in the experiments it where  $C_T$  is known, the model is structurally locally identifiable, although if it is defined that  $k_{d1}|k_{d2}$  then the locally identifiable parameters become globally identifiable and the model becomes structurally globally identifiable. In experiments where  $C_T$  was unknown it was simply structurally unidentifiable.

Examples 3.8 and 3.4 showed that multiple parameter vectors with the same input give the same state-space solution – so the models system mapping is  $X_{\Pi}$  non-injective. Example 3.4 showed that there were multiple state-space solutions that would give the same output solution – so the models output mapping  $Y_{\Sigma}$  is non-injective.

This means that when  $C_T$  is unknown the model is simply structurally unidentifiable.

The identifiability of the **adjusted two concentration Langmuir (LA) model** was analysed with the output equation method. This had the same result as considering the two configuration model with  $k_{d1} \leq k_{d2}$ ; i.e. for two models to give the same output for a generic input, their parameters must be related by either  $(k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R) = (\bar{k}_{a1}/\gamma, \bar{k}_{d1}, \bar{k}_{a2}/\gamma, \bar{k}_{d2}, \bar{C}_T\gamma, \bar{R})$ . In the experiments it where  $C_T$  is known, the model is structurally globally identifiable. In experiments where  $C_T$  was unknown it was simply structurally unidentifiable.

In Example 3.11, it was shown that the L1 and L2 models are structurally distinguishable. An analysis with the same techniques shows that the LA and L2 models are also structurally distinguishable, code for this is shown in ??.

### 5.3.2 Langmuir with transport model

Again, Evans *et al.* (2013) analysed the identifiability of both models, but we have made new definitions which allow us to expand this analysis.

For the **single configuration Langmuir with transport (LwT1) model** (Eq. 5.3),

there are six parameters  $k_a$ ,  $k_d$ ,  $C_T$ ,  $k_m$ ,  $h$  and  $R$ . Their identifiability was analysed with the output equation method in Evans *et al.* (2013). This was repeated with the BVDP and TSA with limits around the discontinuity, full code for this analysis is shown in the appendix, Subsection A.5.1.

This analysis showed that the parameters  $R$  and  $k_d$  as well as the product  $C_T h$ , and quotients  $k_a/h$  and  $k_m/h$  are all globally identifiable; although the parameters  $k_a$ ,  $C_T$  and  $k_m$  and  $h$  are not.

This analysis also gave the relationship between the values the parameters take at the change of the phases:

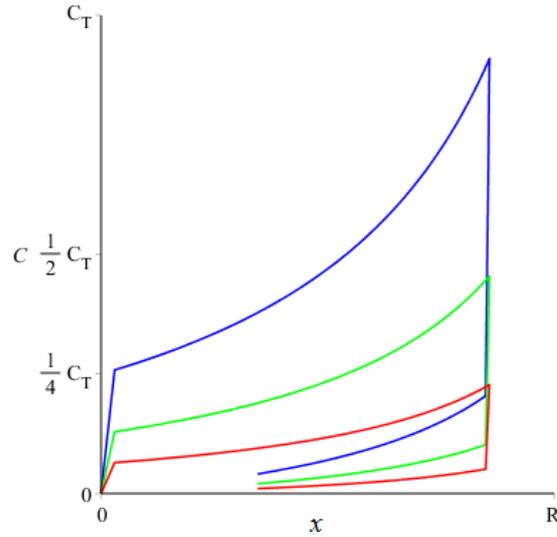
$$\begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix} \mathbf{x}(t, \mathbf{p}). \quad (5.26)$$

where  $\gamma = \frac{\bar{h}}{h}$ .

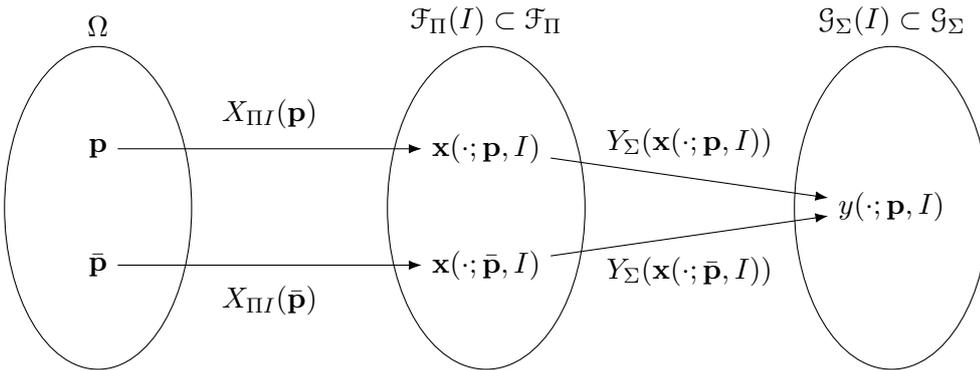
As the user of an SPR or BLI machine determines when  $t_1$  is. So this must also hold throughout the association phase. Additionally because of Eq. 3.23 these relationships will generally hold for the dissociation phase too. As a result we get a mapping  $\lambda(\mathbf{x}(t, \mathbf{p}))$  between indistinguishable state-space solutions.

This tells us that each parameter vector is mapped onto a unique state-space solution. That is the models parameter state-space solution mapping  $X_{III}$  is bijective. Fig. 5.21 shows the trajectories of state-space solutions that correspond to the same output solution. Notably each of these trajectories is an anisotropic scaling of the other trajectories.

This model has a bijective state-space solution mapping, and a non-injective output solution mapping. This is illustrated in Fig. 5.22. The structure of this mapping is a contrast to that of the single configuration Langmuir model which had a non-injective state-space solution mapping and a bijective output solution mapping. Whilst both models are structurally unidentifiable their unidentifiability arises from different aspects of the



**Figure 5.21:** Graph showing three indistinguishable state-space trajectories (red, blue and green), corresponding to parameter vectors indistinguishable to those in Table 5.3.



**Figure 5.22:** In the L2 model indistinguishable parameters are indistinguishable because both  $X_{\Pi}(\cdot, I)$  and  $Y_{\Sigma}(\cdot)$  are surjective, for a generic input  $I$ .

model. As a result there are conceivable simple outputs where the system contained in the Langmuir with transport model would be globally identifiable, whereas there are not for the Langmuir model. One such simple output is the vector  $\mathbf{y} = (x_1, C)$ .

The **two configuration (LwT2) model** has eight parameters  $k_{a1}, k_{a2}, k_{d1}, k_{d2}, C_T, k_m, h$  and  $R$ . The same method was used to determine the relationship between these parameters.

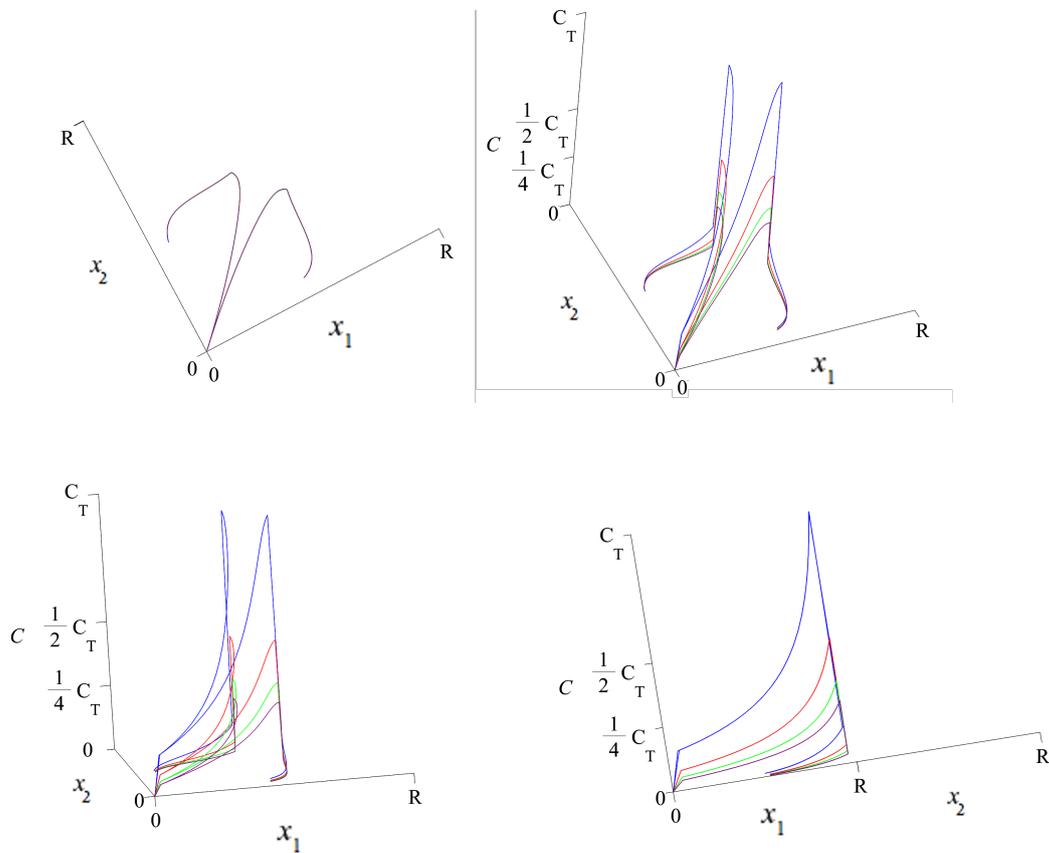
The parameter  $R$  and the functions  $hC_T$  and  $k_m/h$  are globally identifiable and the functions  $k_{a1}/h$  and  $k_{a2}/h$  as well as the parameters  $k_{d1}$ , and  $k_{d2}$  are locally identifiable. As a result the model is structurally unidentifiable, unless  $C_T$  is known, in which case it is structurally locally identifiable.

Two forms for the mapping between solutions with equal outputs were found:

$$\lambda(\mathbf{x}(t, \mathbf{p})) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \gamma \end{bmatrix} \mathbf{x}(t, \mathbf{p}), \text{ or } \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & \gamma \end{bmatrix} \mathbf{x}(t, \mathbf{p}), \quad (5.27)$$

where  $\gamma = \frac{\bar{h}}{h}$  and the state variable is defined as  $\mathbf{x} = (x_1, x_2, C)$ . Finding these mappings allow for the visualisation of state-space solutions that are mapped to the same output solution. This is shown in Fig. 5.23. Similar to the L2 model we may define that  $k_{d1} \leq k_{d2}$ , this further improves the identifiability of the system, making the products  $k_{a1}C_T$  and  $k_{a2}C_T$  as well and the parameters  $k_{d1}$ , and  $k_{d2}$  are globally identifiable. In this case the model is structurally unidentifiable, unless  $C_T$  is known, in which case it is structurally globally identifiable.

Both versions of this model are distinguishable from the previous models because due to their initial conditions they predict  $\frac{dy}{dt} = 0$  initially. The two forms of this model are also structurally distinguishable from each other. But there are parameter vectors where LwT2 and LwT1 are output indistinguishable, these are the same as those for L2 and L1, that is where one or more of the parameters of the two configuration model are set to 0, or where the kinetic constants of the two configuration model are the same and the products of  $k_{a1}I$  are conserved between the models.



**Figure 5.23:** Visualisations of a solution trajectories for the Langmuir two concentration model with identical outputs.

### 5.3.3 Effective rate constant model

For the **single configuration** model there are five unknown parameters:  $k_a$ ,  $k_d$ ,  $C$ ,  $R$  and  $k_M$ . This model was analysed with the output equation method and there was no reason to investigate the relationship between state-space solutions that are mapped to the same output solution, because the model's output function  $y = \alpha x_1$  is bijective. This analysis showed that the parameters  $R$  and  $k_d$  as well as the product  $Ck_a$ , and the quotient  $k_a/k_M$  are globally identifiable, although the parameters in these expressions are simply unidentifiable. The model is therefore simply structurally unidentifiable, unless one of  $C$ ,

$k_a$  and  $k_M$  are known, in which case it is structurally globally identifiable.

The **two configuration** was analysed with the BVDP method and the TSA method with limits around the discontinuity.

The models the parameter  $R$  is globally identifiable, the parameters  $k_{d1}$  and  $k_{d2}$ ; as well as the products  $Ck_{a1}$  and  $Ck_{a2}$ , and the quotients  $k_{a1}/k_m$  and  $k_{a2}/k_m$  are locally identifiable. Notably the parameters  $C$ ,  $k_{a1}$ ,  $k_{a2}$ , and  $k_m$  are simply unidentifiable.

This model shares its  $\lambda(\mathbf{x}(t, \mathbf{p}))$  function with the L2 model:

$$\lambda(\mathbf{x}(t, \mathbf{p})) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ or } \mathbf{x}(t, \mathbf{p}) \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t, \mathbf{p}) \quad (5.28)$$

Correspondingly when their solution trajectories are visualised they are qualitatively similar to those for the two configuration Langmuir model, visualised in Fig. 3.6. If it is defined that  $k_{a1} > k_{a2}$  this makes the locally identifiable parameters, products and quotients globally identifiable similar to Example 3.2.

As a result in these experiments, where  $C$  is known, the model is structurally globally identifiable. In experiments where  $C$  was unknown it was simply structurally unidentifiable.

The **adjusted two configuration** version of the model shares its identifiability properties with the two configuration model where  $k_{a1} > k_{a2}$  is assumed.

As a result in these experiments, where  $C$  is known, the model is structurally globally identifiable. In experiments where  $C$  was unknown it was simply structurally unidentifiable.

All three versions of this model are distinguishable from the previous models. However as the two Langmuir models are limiting cases of these three models (as  $k_m$  grows un-

boundedly large), outputs can be produced by this model that are arbitrarily close to the outputs of those models. Like the previous models the three forms of this model are also structurally distinguishable from each other, but there are parameter vectors where each of the two concentration models are output indistinguishable from the single configuration model and as a result each other. Namely when one or more of the parameters of either the the two configuration or adjusted two configuration model are set to 0, or where the kinetic constants of the two configuration models are the same and the products of  $k_{a1}$  are conserved between the models.

### 5.3.4 Bivalent Langmuir model

This model has six unknown parameters:  $k_{a1}$ ,  $k_{d1}$ ,  $k_{a2}$ ,  $k_{d2}$ ,  $C$  and  $R$ . The output equation method was used to determine the relationship between these parameters, and the ORC method was used to determine the relationship between state-space solutions corresponding to each indistinguishable parameter vector.

The parameters  $k_{d1}$ ,  $k_{a2}$ ,  $k_{d2}$  and  $R$  as well as the product  $Ck_{a1}$ , are globally identifiable; and  $C$  is unidentifiable. The model is therefore structurally unidentifiable, unless  $C$  is known, in which case it is globally identifiable.

For two output solutions to give the same output solution their curves have to be related by:

$$\lambda(\mathbf{x}(t, \mathbf{p})) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}(t, \mathbf{p}). \quad (5.29)$$

This model is structurally distinguishable from all previously mentioned models. If  $k_{a2} = 0$  then it becomes output indistinguishable to the single configuration Langmuir model.

### 5.3.5 Bivalent ERC model

The analysis of this model was problematic. Because of the complexity of its state-space equations the pre existing methods for structural identifiability analysis were unable to process this model. Attempts to implement them resulted in computers over-heating and crashing.

As a result this one of the models that necessitated the use of the limits of the outputs at the start and end of phases (subsection 3.2.3). The BVDP and TSA methods were used.

This model has seven unknown parameters:  $k_{a1}$ ,  $k_{d1}$ ,  $k_{a2}$ ,  $k_{d2}$ ,  $k_m$ ,  $C$  and  $R$ . The parameters  $k_{d1}$ ,  $k_{a2}$ ,  $k_{d2}$  and  $R$  as well as the product  $Ck_{a1}$  and the quotient  $k_{a1}/k_m$  are globally identifiable; and  $C$  is simply unidentifiable. The model is therefore structurally unidentifiable, unless  $C$  is known, in which case it is globally identifiable.

Like the bivalent Langmuir model it was found that if multiple state-space solutions are mapped to the same output solution they are the same state-space solution, i.e. Eq. (5.29).

This model is structurally distinguishable from all previously mentioned models. However the bivalent Langmuir model is the limiting cases of this model (as  $k_m$  grows unboundedly large), outputs can be produced by this model that are arbitrarily close to the outputs of those models. Additionally if  $k_{a2} = 0$  then it becomes output indistinguishable to the single configuration ERC model.

### 5.3.6 Bivalent ERC model with spatial effects

As this model is even more complex than the previous one, the analysis of this model was also problematic. However the method using the limits of the outputs at the start and

end of phases (subsection 3.2.3), BVDP and TSA were used to obtain a partial solution. This was used for the basis of an analysis with an augmented version of the ORC model. Code for this analysis can be found in the appendix, in Subsection A.5.2.

This model has eight unknown parameters:  $k_{a1}$ ,  $k_{d1}$ ,  $k_{a2}$ ,  $k_{d2}$ ,  $k_m$ ,  $C$ ,  $R_1$  and  $R_2$ . The parameters  $k_{d1}$ ,  $k_{a2}$ ,  $k_{d2}$ ,  $R_1$  and  $R_2$  as well as the product  $Ck_{a1}$  and the quotient  $k_{a1}/k_m$  are globally identifiable; although  $C$ ,  $k_{a1}$  and  $k_m$  are simply unidentifiable. The model is therefore structurally unidentifiable, unless  $C$  is known, in which case it is globally identifiable.

The ORC method was used to analyse relationship between state-space solutions that are mapped to the same output solution. Like the bivalent Langmuir model it was found that if multiple state-space solutions are mapped to the same output solution they are the same state-space solution.

$$\lambda(\mathbf{x}(t, \mathbf{p})) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}(t, \mathbf{p}). \quad (5.30)$$

This model is structurally distinguishable from all previously mentioned models. However, if  $k_{a2} = 0$  then it becomes output indistinguishable to the single configuration ERC model

## 5.4 Discussion

There were models in the literature appropriate for estimating the kinetic rates of antibody-antigen binding in SPR experiments, however for models that allowed multiple binding types often required some adaptation, to allow for the fact that bivalently bound antibody takes up more antigen than monovalently bound antibody. New models were developed to include combinations of the effects of non homogeneity of the distribution of both antibody

Model	$X_\pi$	$Y_\sigma$
Single configuration Langmuir	non-injective	bijjective
Two configuration Langmuir	non-injective	non-injective
Adjusted two configuration Langmuir	non-injective	bijjective
Single configuration Langmuir with transport	bijjective	non-injective
Two configuration Langmuir with transport	bijjective	non-injective
Single configuration ERC	non-injective	bijjective
Two configuration ERC	non-injective	non-injective
Adjusted two configuration ERC	non-injective	bijjective
Bivalent Langmuir	non-injective	bijjective
Bivalent ERC	non-injective	bijjective
Bivalent ERC with spatial effects	non-injective	bijjective

**Table 5.11:** Structure of mappings for models.

and antigen as well as the the mechanisms of monovalent and bivalent binding.

By analysing the structural identifiability of these models it has been established that analyte concentration is required for affinity to be accurately estimated. All models were structurally unidentifiable, unless concentration,  $C_T$  was known beforehand. Definitions were made for each model so that if  $C_T$  was known then the model would be structurally globally identifiable.

Whilst all models were structurally unidentifiable in one case and globally identifiable in another these properties arose from separate properties of the mappings associated with the model. These are shown in Table 5.11.

The new models, and the adjusted models as a consequence of the way they dealt with bivalent binding had bijjective output solution functions. This is an improvement over the original models which needed additional definitions to be made to have bijjective output solution functions.

These improvements are useful- they mean that in experiments with monoclonal antibody in known concentration accurate parameter estimation can be done. This is the circumstance that should be expected with artificially produced antibody- so these models would

be appropriate for use in the development of therapeutic antibodies.

However the fact that the models state-space solution functions are generally non-injective suggests that when models are developed from these for use with polyclonal antibody from a clinical setting- where the number of concentrations will be unknown- there will be problems with accurate parameter identification.

In the case of the Langmuir with transport models these functions are bijective. However, whilst there are conceivable outputs that could allow for full parameter identification- there is no practical way of implementing them. The SPR machine doesn't allow for measuring the concentration of the antibody except at the chip's surface- and even there it does it indirectly by measuring the amount bound, and therefore the rate at which this amount changes.

In nearly all cases at least two of the parameters we are most interested in estimating with each model, the association rate constants, and the concentration of analyte are simply unidentifiable. In the case of the Langmuir with transport model, these parameters are still not identifiable and it is unclear how the experiments could be adapted to allow for measurements that would make the parameters identifiable.

This suggests that experiments with a mechanistically dissimilar experimental system or a technique that combines the SPR experiments with such a system would be required to estimate these parameters. The possibility of using reactions in a closed system (i.e. with no flow of analyte) will be discussed in a later chapter.

## Chapter 6

# Affinity estimation for monoclonal antibody

One necessary step for developing methods of measuring antibody-antigen affinities for pharmaceutical and medical use is to repeatably measure these affinities for samples of monoclonal antibody with known concentration. This chapter focuses on fitting the models of Chapter 5 to experiments in 4.3 and 4.4 for this purpose.

To do this, two groups of SPR experiments were conducted. The protocol used was discussed in Section 4.3 and Section 4.4. These two groups of experiments were originally conducted with separate aims, in the first case to investigate the effect of high levels of variation of antigen density and antibody concentration on the experiments, and in the second case to investigate the difference in binding reactions of a single antibody against a variety of antigen.

However, partway through the second group of experiments, the SPR machine broke down. As a result, some of the experiments of the second set were available much sooner than the rest, so were modelled along with the first group.

In practice, the first group of these experiments was used to develop and compare models, until models were developed that fitted them well that also gave consistent parameter estimations. These models were applied to the second set of experimental data, which they

also fit well and provided consistent parameter estimates for. Notably, these parameter estimates were so consistent that the variation in affinity caused by the antibody binding to the same epitopes on separate antigen was greater than the variation between experiments done with the same antigen. This allowed for comparisons between the reaction rates of each antibody against multiple antigens.

In more detail, the first group of experiments, those of Section 4.3, eight separate concentrations of antibody were reacted against two separate densities of antigen. This was to test whether models could give consistent affinity estimates at different analyte concentrations and ligand densities and to test whether affinity estimates between models would be consistent. Particularly as the Langmuir model is a limit of the ERC model, it was wondered if there were antigen densities and antibody concentrations where the two models would produce the same results.

In the second group of experiments, those of Section 4.4, six concentrations of antibody were reacted against varying ligands. This was to test whether models would allow for affinities to be estimated so consistently that ligands could be differentiated by their affinity estimates. This was both to establish the affinities of these reactions and to evaluate the models.

In practice, as the first data set was available long before most of the second, and it featured characteristic artefacts and anomalies, the data available from the second group of experiments was used with the first to develop models and fitting techniques. In contrast, the remainder of the second group of experiments when finally conducted, was used to test these new models and techniques against the fits produced by the default software that is used on the SPR machine: ProteOn Manager.

## 6.1 Error in SPR experiments

There are two different kinds of error that affect SPR. Errors that come through the models and fitting processes necessary to get affinity estimates and measurement error. As this chapter will centre on minimising the first and reducing the residuals of fits to being as close as possible to the second a discussion of both is necessary.

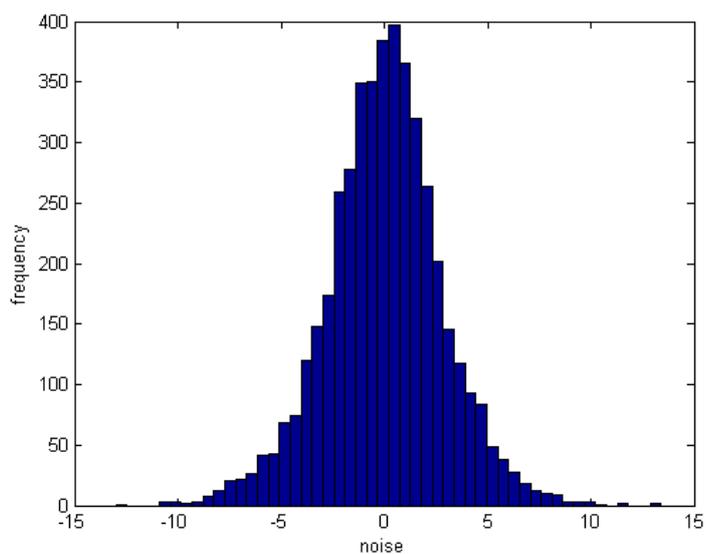
Figure 1.15 illustrates that in the benchmark studies estimates of antibody-antigen affinity have been hugely inconsistent, whereas Fig. 1.16 showed estimates of other kinds of affinity have been very consistent. Section 1.4 suggested this was due to systematic flaws in the models used by the benchmark studies.

Chapter 5 centred on model development which was conducted simultaneously with the fitting shown in this chapter, with the goal of developing models that would give more consistent parameter estimates, and better fits.

However, even when no error is introduced by modelling, the signal generated by the SPR machine is noisy. Chapter 4 shows numerous sensorgrams with measurements taken from before the start of the association phase.

In these periods buffer was running over a chip, and no interactions were happening. To study the noise that effects measurements, these were combined into a dataset with 4500 measurements so that the distribution of noise on the signal could be measured.

As SPR only measures the change in surface plasmon resonance only, the mean signal from each pre-association portion of a sensorgram was calculated and taken away from each measurement from that sensorgram.



**Figure 6.1:** Histogram of SPR noise

The resulting SPR measurements were then plotted in a histogram to produce Fig. 6.1 and the standard deviation of the data was calculated at 2.87. To determine if the noise was distributed normally the Kolmogorov-Smirnov test was used. The test gave a p-value of 1.6E-172 for the hypothesis that the data was normally distributed. As a result this hypothesis was accepted.

## 6.2 Fitting methods

In Section 3.4 a fitting method involving two existing algorithms were discussed, FACSIMILE and Differential Evolution. However before these can be implemented, the data needs to be prepared.

## 6.3 Preparing the data for fitting

The data was initially processed using the BioRad ProteOn Manager software, where reference subtractions were performed to remove noise. Time-series corresponding to each interaction spot were then put into FACSIMILE programs to be fitted.

First, the data had to be cleaned. Whilst the SPR machine regularly measures the concentration of bound antibody; the outputs were missing some data points. Because the fitting process is reliant on correctly identifying the beginnings and ends of the association and dissociation phases, it was necessary to give values to missing data points. These were obtained with linear interpolation from non-missing data points. Linear interpolation was chosen because of its comparative simplicity, and the fact that the points for which measurements were available were very dense.

Whilst reference subtractions removed much of the noise, data included sudden changes of signal, at the beginning and end of phases, that are too abrupt to be explained by antibody association and dissociation processes. One explanation for these is that as buffer is replaced with analyte, and analyte with buffer, there are small changes in the pressure on the reaction spot as well as the optical properties of the substance in flow.

To cope with these effects a step function  $s(t)$  was added to the output of each of the models which takes a different value in each phase of the model.

$$s(t) = \begin{cases} c_0 & t < 0 \\ c_1 & t \in [t_0, t_1) \\ c_2 & t \in [t_1, t_2) \end{cases} \quad (6.1)$$

This effectively adds three extra parameters to each of the models but does not affect the identifiability of the models, so these parameters were not considered in the identifiability analyses (Section 5.3). This is because all the models of Chapter 5 were time continuous, whereas this addition is discontinuous at the points  $t_0$  and  $t_1$  as a result theoretically the values can be derived from one-sided limits:

$$c_1 - c_0 = \lim_{t \rightarrow 0^+} y - \lim_{t \rightarrow 0^-} y \quad (6.2)$$

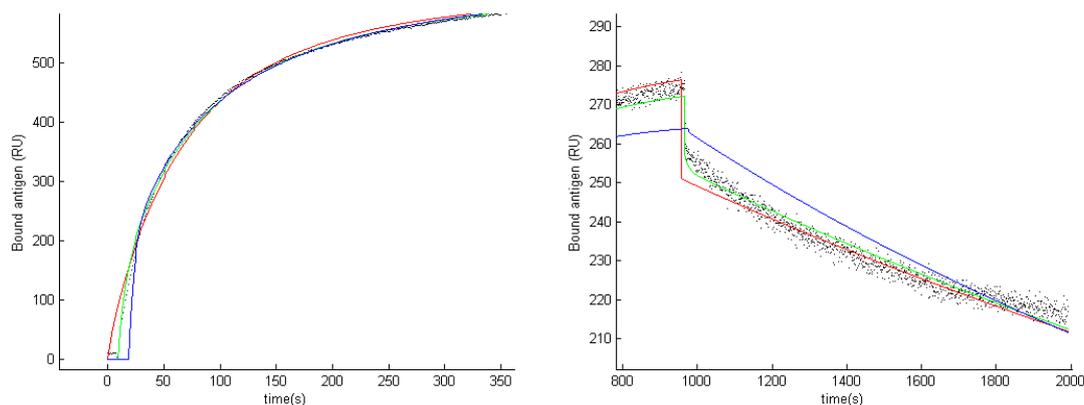
$$c_2 - c_1 = \lim_{t \rightarrow t_1^+} y - \lim_{t \rightarrow t_1^-} y \quad (6.3)$$

and setting  $c_0$  so that the output signal is at 0 at the start of the association phase.

Practically establishing these values is more complex, because the data contains noisy measurements made every 0.9 seconds. So whilst sudden changes in output signal can be observed, it is not known to what degree they are due to measurement error, rapid antibody association and dissociation, or the effects of analyte changing at the start and end of phases.

It is necessary to accurately determine the values of  $c_0$ ,  $c_1$  and  $c_2$ , as well as fit the models to the data, because determining these values incorrectly will alter the parameter values for which RSS is minimised.

A complication to judging these is that it can be unclear when the association phase begins. Although the machine is programmed to start the phase at a specific moment, it



(a) Association phase data fit with three curves:  $c_1$  overestimated by 20RU and  $t_1$  underestimated by 9 seconds (red),  $c_1$  and  $t_1$  estimated correctly (green),  $c_1$  underestimated by 20RU and  $t_1$  overestimated by 9 seconds (blue).

(b) Dissociation phase data fit with three curves:  $c_1$  overestimated by 10RU and  $t_1$  underestimated by 9 seconds (red),  $c_1$  and  $t_1$  estimated correctly (green),  $c_1$  underestimated by 10RU and  $t_1$  overestimated by 9 seconds (blue).

**Figure 6.2:** Two configuration bivalent fits.

takes time for the analyte in the lane to reach the concentration with which it is stored. The ProteOn manager has features that align the time measurements from the experiment so that association on each lane starts and finishes at the same time. It was unclear what effect this feature would have on fitting the data to differential equations, so it was not used.

The problems caused by wrong estimation of  $c_1$ ,  $c_2$  and  $c_3$ ;  $t_1$ ,  $t_2$  and  $t_3$  are illustrated in Fig 6.2. With the BL model (Eq. 5.10) being fit to the 3.1nM curve from the SN607D8 vs HLA-A2 high density antigen data set (taken from Section 4.3).

In Fig. 6.2(a) we see three different fits to the data in the association phase. Although all three fits look similar, and the model in all three cases quickly converges on the data, they have very different affinity estimates. This is because, for the model to converge on the data when the step function is wrongly set, it initially needs a much steeper or shallower gradient. The rate of change where no antibody is bound for all of the proposed models

is proportional to either  $k_{a1}R$  or  $(k_{a1} + k_{a2})R$ , where  $k_{a1}$  and  $k_{a2}$  are the association rate constants, and  $R$  is the maximum density of antibody that could bind at the interaction spot. These are independent of many of the parameters. So, different values of  $c_1$  and  $t_1$  lead to very different values for  $k_{a1}$ . In the case pictured the three estimates varied from  $1.5 \times 10^{-3}$  to  $9.3 \times 10^{-5} \text{ nM}^{-1}\text{s}^{-1}$ .

In 6.2(b) we see three different fits to the data in the dissociation phase. Because the step function puts the green model in the correct place, the model is able to find parameter values for which the model stays within the data until close to the end of the phase where it diverges from the data. Both the red and blue models converge on the data later in the phase but pass through it. In this case, we see that a badly estimated step function can make it impossible for the model to be well fitted to the data. Again these fits give three very different affinity estimates. This is because, for the model to converge on the data, when the step function is wrongly set, it initially needs a much steeper gradient. As the gradient is determined primarily by the dissociation rate constants,  $k_{d1}$  and  $k_{d2}$ , incorrectly set step functions at the start of the dissociation phase lead to strange estimates of these parameters. Often when  $c_2$  or  $t_2$  are incorrect the fit with the smallest RSS found will be one where  $k_{d1}$  or  $k_{d2}$  is given an extreme value. In this case the incorrect estimates of  $k_{d1}$  vary from  $3.1 \times 10^{-9}\text{s}^{-1}$  to  $3.1\text{s}^{-1}$ ; whereas the fit with the correct step function was  $8.1 \times 10^{-2}\text{s}^{-1}$ .

The problems relating to incorrect values chosen for  $c_1$ ,  $c_2$  and  $c_3$ ;  $t_1$ ,  $t_2$  and  $t_3$  became apparent early on in the fitting process. To eliminate these problems, four approaches were used for estimating the parameters of the step function.

1. Assuming that  $s(t)$  takes a constant zero value.
2. The parameters of the step function were fit to the data in FACSIMILE.
3. The parameters of the step function were estimated in Matlab.

4. The parameters of the step function were fit visually.

Before the significance of the step function was realised, it was not fit, which was equivalent to **assuming it takes a constant zero value**. This worked well for Wk1d12 vs A2 and B57 datasets where the output does not rapidly change abruptly at either the start or end of the association phase. However, this worked badly for Sn607d8 vs A2 and Sn230G6 vs A2. This caused errors for these data sets, these which were similar to those with a blue line shown in Fig 6.2.

An intuitive way of dealing with this problem is to create the step function, and **allow FACSIMILE to fit the parameters of the step function simultaneously with the parameters of the model**. This was problematic, as the RSS was more sensitive to change in the parameters of the step function than the parameters of the model. Typically  $c_1 - c_2$  was set much too high, allowing the output of the model to miss the data in the early part of the dissociation phase, but to stay within the data for the majority of the phase. Visually these fits were similar to the red line in Fig. 6.2(b), and typically they gave very low estimates of  $k_{d1}$  and  $k_{d2}$ .

To avoid the problems of fitting it simultaneously with the data, the parameters of the step function were **fit in Matlab, before FACSIMILE was run**. A number of methods for this were developed.

Consider the time series data to which the model is fitted, as a mapping  $\mathbf{D}(t) : c\mathbb{N} \rightarrow \mathbb{R}^n$ , where  $c \in R$  is the period of the measurements and  $n$  the number of channels. Then define the difference between one point and the next as  $\Delta\mathbf{D}(t) = \mathbf{D}(t + c) - \mathbf{D}(t)$ . The value of  $t$  for which  $\sum_1^n \Delta D(t)_i$  is maximised would be intuitively a good place to start the association phase, because the models predict the greatest rate of change of signal in any phase at its beginning. If this value of  $t$  is chosen as the start of the phase then the values  $D_i(t)$  would be good values for  $c_2$  for each of the channels.

This approach has some problems, firstly as the measurements  $\mathbf{D}(t)$  are noisy and for the expected parameters  $\Delta D(t)_i$  would remain within measurement error for several measurements at the beginning of the association phase, so various methods were used to weight the comparison between sums so as to prioritise earlier points.

A more complicated problem is that the concentration of antibody in the inlet is likely to reach the sample concentration between measurements. Change over the interval in which this happens is unlikely to be as great as change in the next interval.

We also had to estimate  $c_3$ . This is a more complex challenge. For some time series there is a rapid change of 20RU or more at the beginning of the association phase; for some time series, there is no change greater than the error of measurement. That being said, the models predict that at the beginning of the dissociation phase we will see the biggest decrease in output, so the value of  $t$  for which  $-\sum_1^n \Delta D(t)_i$  is maximised is used as the start to the dissociation phase, and for the  $i^{th}$  channel  $c_3$  is set so that  $c_2 - c_3 = \Delta D_i(t - c)$

Unfortunately, all the problems for estimating the step values of the association phase also apply here. The values in the data series  $\mathbf{D}(t)$  are still noisy. This means that the value of  $t$  for which  $-\sum_1^n D(t)_i'$  is maximised may not be the beginning of the dissociation phase. Unfortunately, this comes in the middle of the data series rather than at the beginning so we are not simply able to give a higher weighting to early values of  $t$ .

An alternative approach is to fit the values of the **step function visually**. Whilst the data is noisy, trends can generally be observed.

The final approach taken was to let Matlab do the initial estimates, using the method described in this section and then to adjust them visually where appropriate.

## 6.4 Developing and comparing models

Initially the data presented in Subsections 4.3.1 - 4.3.3 was used to test existing models against new models. These fits in some cases showed systematic error. So gradually more complex models were developed to deal with these errors until models that generally fitted the data well in each case were developed. The following section parallels Section 5.1. In the same order as the models were presented the models are used to fit the data and perform parameter estimation.

Initially, this fitting process was very slow, due to the numbers of parameters needed to be fitted in each model and the large range that parameters were allowed to vary across. As a result, it was chosen to limit the number of reaction spots modelled at a single time to 3 in the case of the SN607G8 vs A2 dataset and the WK1D12 dataset. By the time the SN230G6 was being modelled, the range parameters could vary over was better established, so 4 reaction spots were fitted simultaneously.

### 6.4.1 Testing the ERC2 model against the L2 model

The L2 model (defined in Example 2.2 in Section 2.2) was initially selected as an established model against which to test other models. This is because it is relatively simple and it is one of the common models available to anyone using a Bio-Rad SPR machine (Bio-Rad, 2011).

The ERC2 model (defined in Eq. 5.6) was the first model tested against it. This was because of two key features, namely that it does not assume the analyte in contact with the antigen attached to the chip was at the same concentration as the analyte in flow, and that it was structurally globally identifiable when the concentration of analyte was known.

	Minimum	Maximum	Step size	Initial estimate
$k_a(\text{M}^{-1}\text{s}^{-1})$	1E+4	1E+11	1E+0.1	1E+8
$k_d(\text{s}^{-1})$	1E-5	1E+2	1E+0.1	1E-2
$R(\text{pgnm}^{-2})$	1E-3	1E+3	1E+0.1	1E+0
$k_m(\text{pgnm}^{-2}\text{M}^{-1}\text{s}^{-1})$	1E-6	1E+6	1E+0.2	1E+0

**Table 6.1:** Details of parameter space as considered by Differential Evolution.

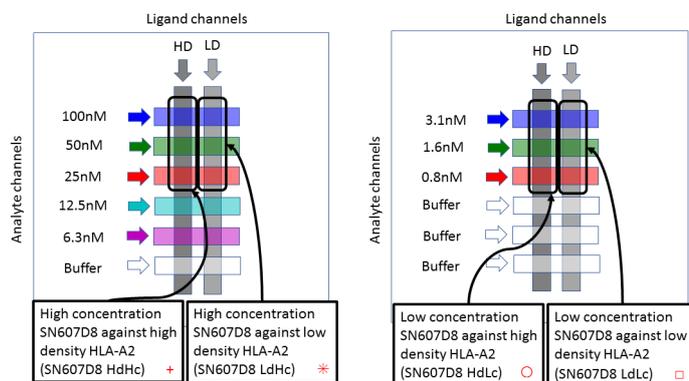
The data presented in Subsections 4.3.1 - 4.3.3 was divided into 12 data sets and their replicates. The contents of each of the datasets are illustrated in Fig. 6.3.

Facsimile programs were written to use these models to fit each dataset. A sample Facsimile program is included in Section. Each program estimated the association and dissociation rate constants ( $k_{a1}, k_{d1}, k_{a2}, k_{d2}$ ) simultaneously for each time series in the dataset it handled. Other unknown parameters ( $R, k_m$ ) were fitted for each time series. As a result, these parameters will be throughout this chapter be referred to as  $R_1, R_2, R_3$ , etc.;  $k_{m1}, k_{m2}, k_{m3}$ , etc., where the subscripted numeral will represent the whether the parameter is for the first, second or third etc. time series in the dataset. The time series will always be ordered from high concentration to low concentration.

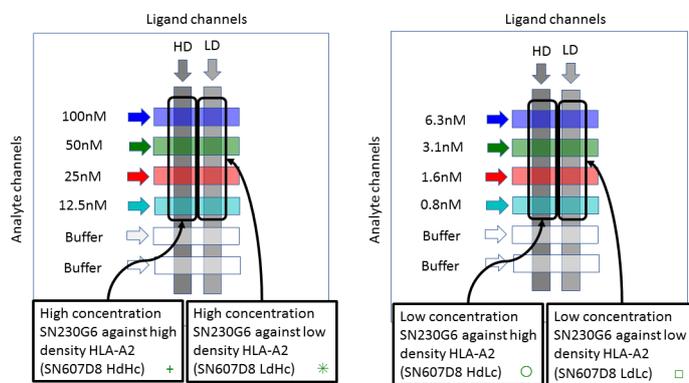
These programs (as discussed in 3.4.2) were used to fit parameters by integrating the model equations and varying model parameters. These were used alongside Differential Evolution, which kept tracks of populations of parameter vectors (as discussed in 3.4.1).

Each Facsimile program takes in a file generated by Differential Evolution running in Matlab. These files would contain a parameter vector taken from Differential Evolutions population of parameter vectors. The population of parameter vectors was 150 vectors each taken from a grid in parameter space defined by the maxima and minima and step-size presented in Table 6.1. These values were chosen to allow each parameter to vary over a greater range than what it would be expected to given the literature (Rich *et al.*, 2009; Yang *et al.*, 2016).

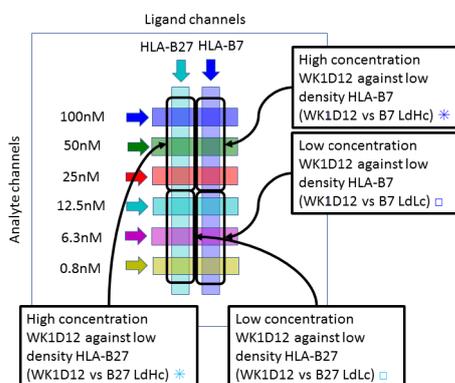
Differential evolution was typically run for 24 hours for each model on each dataset,



(a) Data sets taken for the SN607D8 against HLA-A2 sensorgrams.



(b) Data sets taken for the SN230G6 against HLA-A2 sensorgrams.



(c) Data sets taken for the WK1D12 against multiple antigen sensorgrams.

**Figure 6.3:** Diagrams of the interaction spots where the sensorgram curves in each data set comes from, with labels showing abbreviated names of the dataset, and colour coded symbols that will be used to represent the datasets through this chapter.

allowing approximately 100 populations of parameter vectors to be considered per model per dataset, and typically 12,500 individual parameter vectors to be passed to facsimile per model per dataset.

Each parameter vector is varied by Facsimile. Facsimile creates a log-file each time a program is run in it. A sample Log-file is shown in Subsection A.7. This illustrates the parameters being read in, varied logarithmically, their covariance being tabulated.

In some cases, Facsimiles, statistical analysis removes parameters from the fitting process for being not determining the value of a parameter within two orders of magnitude (described in *citetuguide*) .

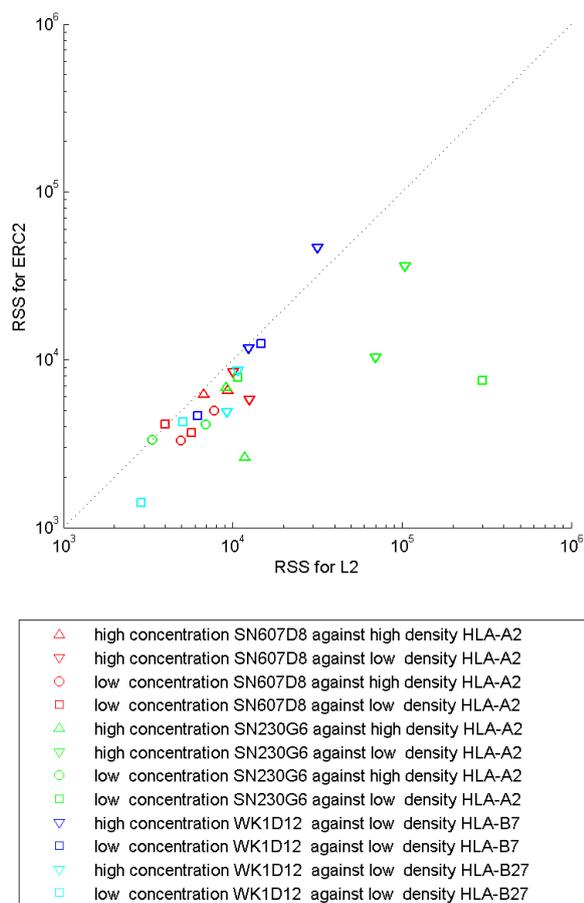
The values for these parameters are then fixed to the last value used in the statistical analysis before continuing with the parameter fitting and treated as unknown in subsequent statistical tests.

The parameters estimated that are not removed in this way are presented by Facsimile with a confidence interval for their fitting as well as an estimated standard deviation for the natural logarithm (SDLN) of the parameters value.

Parameters either removed from the fitting process or with an SDLN greater than 0.2 are considered not well determined (NWD). As a result in Tables 6.2 - 6.3, which show NWD parameter estimates these are marked with an asterisk.

The RSS for fits for both models to these data sets are presented along with a line marking equal RSS for both models in Fig. 6.4. AIC will be used to compare all the models together in the discussion at the end of the chapter.

The ERC model has a better RSS in all but three cases. As the output of the Langmuir model is a limiting case of the ERC model, there are parameter values for which the ERC model would have fit at least as well to the data as the Langmuir model (see subsection 5.1.3). With greater stimulation time we would see these beneath the line of equal RSS



**Figure 6.4:** Scatter graph of RSS for the Langmuir and ERC model

in Fig. 6.4.

In the majority of cases the RSS is reduced, but not more than halved. In 4 cases the RSS is reduced by a factor of ten or more - in each of these cases the antibody-antigen interaction being studied is SN203D8 vs HLA-A2, and with the antigen at high density. The combination of the high affinity of this interaction and the high density of the antigen means that concentration of free antibody able to bind on the chip's surface is much lower than in the sample. Notably these cases include both high concentration and low concentration experiments, showing that in this case the ERC approximation is useful for

low concentration experiments as well, and that density of ligand may be important in determining which model to use.

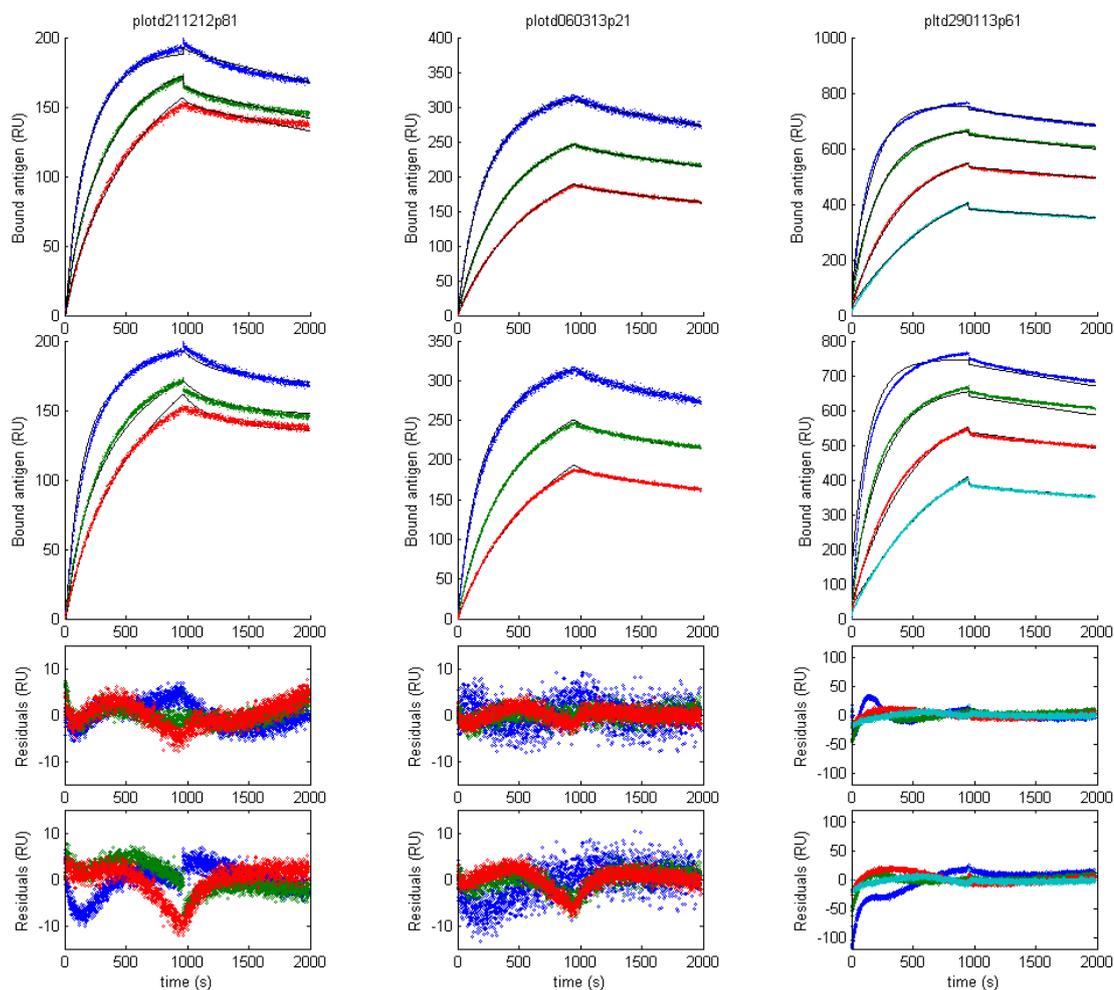
These improvements in RSS are due to the ERC2s better fitting in the association phase. This is illustrated in Figure 6.5(a) and 6.5(b). In some cases these better fits in the association phases lead to better estimates of the output at the end of the association phase, and as a result better dissociation phases for some channels.

Figure 6.5(a) illustrates the effect of the use of different step values on the fitting process. The fits of the Langmuir model (second panel) to data from the middle concentration (green) minimised RSS when the values that the step function used were incorrect.

In some datasets, the RSS was improved by a factor of 10 or more by the ERC2 model. An example of one of these datasets, along with both fits, is shown in Fig. 6.5(c). Here the ERC model had better fits for both models in either phase. One reason for this is because the dissociation constant does not just determine the model's behaviour in the dissociation phase. It also determines the equilibrium value of the association phase, and the amount of bound analyte at the end of the association phase. As a result, a model that struggles to fit the association phase well will also struggle to fit the dissociation phase.

Due to the size of the residuals for the L2 model shown in Fig. 6.5(c) the residual plot for this figure has a different scale. Throughout the chapter, as models improve, the discrepancy between the residuals for the datasets where the models perform worst and best will decrease, allowing residuals to be shown on more consistent scales. However, in the early sections of this chapter, as the greatest residuals for different modelled datasets vary by orders of magnitude a consistent scale cannot be used.

The improvements to RSS are limited by some features of the data discussed in subsections 4.3.1 - 4.3.3. For neither model does there exist a set of parameters for which the outputs



(a) SN607D8 low antigen density, low antibody concentration.

(b) WK1D12 low antigen density, high antibody concentration.

(c) SN203D8 high antigen density, low antibody concentration.

**Figure 6.5:** Fits of the ERC2 model in the top panel, L2 model in second panel and their residuals in the third and fourth panels respectively, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.3.

of two lanes of separate concentration may cross except at the start of the association phase. This happens in three of the eight SN607D8 vs A2 samples and five out of the eight SN230G6 vs A2 samples and all of the WK1D12 vs B7 data. Additionally, of the datasets in which there are no crossovers, there are humps in three of the SN607D8 vs A2 datasets, which similarly are a feature that the model cannot fit to. These features not only stop the reduction in RSS but undermine the reliability of the affinity estimates made by these models.

In Tables 6.2 - 6.3 the parameters estimated by these fitting processes are shown. These parameters are visualised in Fig. 6.6 and 6.7.

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	1.2E+6	2.1E+6	3.3E+6	1.5E+6	1.4E+6	1.2E+8	4.7E+6	5.5E+6
$k_{a2}(M^{-1}s^{-1})$	2.2E+6	1.2E+6	1.5E+6	3.1E+6	3.8E+6	9.6E+7	1.6E+6	4.7E+6
$k_{d1}(s^{-1})$	1.2E-4	8.1E-3	6.1E-3	1.1E-4	2.0E-3	1.6E-4*	4.8E-5	3.8E-5
$k_{d2}(s^{-1})$	8.6E-3	1.0E-4	1.3E-4	5.9E-3	5.1E-7*	3.1E-2	3.3E-3	4.8E-3
$R_1(\text{pgnm}^{-2})$	6.6E+2	6.0E+2	3.2E+2	2.9E+2	4.4E+2	9.4E+2	2.0E+2	2.1E+2
$R_2(\text{pgnm}^{-2})$	6.6E+2	5.6E+2	3.2E+2	2.6E+2	4.4E+2	8.7E+2	1.9E+2	2.2E+2
$R_3(\text{pgnm}^{-2})$	7.0E+2	6.3E+2	3.3E+2	3.0E+2	4.7E+2	8.7E+2	2.2E+2	2.8E+2

(a) Sn607 against HLA-A2.

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	6.2E+6	5.5E+6	6.1E+6	3.9E+6	2.5E+8	1.6E+7	2.1E+7	6.9E+5*
$k_{a2}(M^{-1}s^{-1})$	4.3E+6	2.8E+6	1.2E+7	7.7E+6	1.4E+7	1.6E+10	5.2E+9	2.2E+7
$k_{d1}(s^{-1})$	1.2E-4	1.0E-6*	9.7E-5	4.9E-5*	7.5E-1	8.6E-5	1.0E-4	1.9E-3*
$k_{d2}(s^{-1})$	1.7E-2	1.7E-3	1.3E-1	4.7E-3	8.6E-5	5.4E+1	2.1E+1*	1.7E-4
$R_1(\text{pgnm}^{-2})$	9.4E+2	9.2E+2	9.3E+2	3.2E+2	7.4E+2	6.9E+2	2.0E+2	2.3E+2
$R_2(\text{pgnm}^{-2})$	8.5E+2	9.0E+2	8.5E+2	3.3E+2	6.7E+2	6.2E+2	2.1E+2	2.1E+2
$R_3(\text{pgnm}^{-2})$	8.7E+2	7.7E+2	8.6E+2	3.0E+2	6.5E+2	5.9E+2	2.2E+2	2.0E+2
$R_4(\text{pgnm}^{-2})$	7.5E+2	8.2E+2	8.2E+2	3.1E+2	7.3E+2	5.7E+2	2.1E+2	2.2E+2

(b) Sn230 against HLA-A2.

	B7				B27			
	LdHc	LdHc	LdLc	LdLc	LdHc	LdHc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	1.3E+6	1.5E+6	1.3E+6	1.7E+6	1.1E+6	2.2E+6	2.3E+6	1.9E+6
$k_{a2}(M^{-1}s^{-1})$	1.3E+6	8.0E+5	2.7E+6	2.9E+6	2.5E+6	7.6E+5	2.9E+6	2.5E+6
$k_{d1}(s^{-1})$	6.1E-3	9.0E-3	5.2E-3	1.7E-2	2.2E-4	8.2E-3	9.3E-5*	5.7E-3
$k_{d2}(s^{-1})$	1.3E-4	1.7E-4	9.0E-5	1.6E-4	3.7E-3	4.3E-4	1.2E-3	3.7E-4
$R_1(\text{pgnm}^{-2})$	4.6E+2	2.8E+2	3.4E+2	1.9E+2	2.0E+2	1.3E+2	1.5E+2	8.9E+1
$R_2(\text{pgnm}^{-2})$	4.2E+2	2.6E+2	3.3E+2	1.8E+2	2.1E+2	1.3E+2	1.4E+2	9.2E+1
$R_3(\text{pgnm}^{-2})$	4.4E+2	3.0E+2	3.6E+2	2.0E+2	2.2E+2	1.5E+2	1.5E+2	8.9E+1

(c) Wk1d12 reacted against multiple antigen.

**Table 6.2:** Estimated parameters for each dataset for the L2 model, with parameters with no SDLN or an SDLN greater than 0.2 marked with an asterisk.

For both models, very few parameters are marked as NWD. The parameter most commonly

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	2.4E+6	2.7E+6	3.0E+6	2.9E+6	6.1E+6	2.1E+6	2.4E+6	9.7E+6
$k_{a2}(M^{-1}s^{-1})$	3.1E+6	5.3E+6	2.3E+6	1.9E+6	2.0E+6	5.8E+6	5.2E+6	1.2E+7
$k_{d1}(s^{-1})$	1.7E-4	1.9E-4	7.9E-3	7.0E-3	7.1E-5	4.4E-3	1.1E-2	1.7E-2
$k_{d2}(s^{-1})$	3.4E-2	4.2E-2	2.2E-4	1.7E-4	1.9E-2	6.2E-5	1.0E-4	1.4E-4
$R_1(\text{pgnm}^{-2})$	6.4E+2	5.8E+2	3.1E+2	2.8E+2	4.1E+2	4.3E+2	1.9E+2	1.9E+2
$R_2(\text{pgnm}^{-2})$	6.1E+2	5.2E+2	3.0E+2	2.5E+2	3.7E+2	4.2E+2	2.0E+2	1.9E+2
$R_3(\text{pgnm}^{-2})$	6.0E+2	5.4E+2	3.1E+2	2.8E+2	3.7E+2	5.2E+2	2.3E+2	2.1E+2
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.7E-3	1.3E-3	1.4E+11*	2.1E-3	4.2E-4	1.0E+0*	1.1E-2*	1.2E-4
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	7.1E-1*	1.2E-3	3.6E+6*	1.4E-3	3.9E+20*	3.6E-2	6.0E-3*	2.3E-4
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.7E+0*	5.8E+11*	1.1E-2	6.1E-2	7.7E+4*	2.2E-1*	3.9E-4	5.8E-4

(a) Sn607 against HLA-A2.

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	6.9E+6	1.2E+7	6.3E+7	4.9E+6	2.9E+8	1.6E+7	2.4E+7	2.0E+7
$k_{a2}(M^{-1}s^{-1})$	8.1E+6	5.9E+6	1.9E+8	9.3E+6	1.5E+7	1.6E+10	2.4E+7	5.4E+6
$k_{d1}(s^{-1})$	2.7E-2	5.9E-2	5.3E-2	1.1E-4	8.0E-1	8.6E-5	5.2E-2	1.5E-4
$k_{d2}(s^{-1})$	1.2E-4	1.4E-4	4.9E-4	5.5E-3	7.7E-5	5.4E+1	1.5E-4	4.4E-3
$R_1(\text{pgnm}^{-2})$	9.3E+2	9.3E+2	9.3E+2	3.2E+2	7.5E+2	6.9E+2	2.0E+2	2.3E+2
$R_2(\text{pgnm}^{-2})$	8.5E+2	9.1E+2	8.5E+2	3.2E+2	6.7E+2	6.2E+2	2.3E+2	2.1E+2
$R_3(\text{pgnm}^{-2})$	8.6E+2	7.8E+2	8.5E+2	2.9E+2	6.2E+2	5.9E+2	2.4E+2	2.0E+2
$R_4(\text{pgnm}^{-2})$	7.9E+2	9.3E+2	9.3E+2	3.2E+2	7.5E+2	5.7E+2	2.0E+2	2.0E+2
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.0E-3	3.6E-3	2.6E-3	1.8E+0*	7.7E-4	1.0E-3	3.6E-4	6.2E-4
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	8.4E-3	1.1E-2	2.4E-3	2.9E-2	1.1E+5*	4.6E-3	6.1E+1*	1.6E-3
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.8E-2	5.3E+2*	2.0E-3	4.6E-2	3.8E+10*	2.5E-2	5.0E+8*	1.5E-3
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	9.2E+3*	3.6E-3	2.6E-3	1.8E+0	7.7E-4*	1.7E-2	3.6E-4*	4.1E+8*

(b) Sn230 against HLA-A2.

	B7				B27			
	LdHc	LdHc	LdLc	LdLc	LdHc	LdHc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	2.6E+6	1.3E+6	1.4E+6	1.3E+6	8.4E+6	2.5E+6	7.2E+6	6.6E+6
$k_{a2}(M^{-1}s^{-1})$	2.2E+6	1.6E+6	3.9E+6	3.3E+6	1.4E+10	3.6E+7	1.2E+7	6.4E+6
$k_{d1}(s^{-1})$	2.9E-2	2.5E-4	8.5E-3	1.2E-2	1.4E-3	8.2E-4	1.6E-2	1.8E-2
$k_{d2}(s^{-1})$	2.2E-4	1.5E-2	1.2E-4	1.5E-4	4.6E+1	2.8E-1	7.0E-4	6.6E-4
$R_1(\text{pgnm}^{-2})$	4.5E-1	2.7E-1	3.3E-1	1.9E-1	2.0E-1	1.2E-1	1.4E-1	8.2E-2
$R_2(\text{pgnm}^{-2})$	4.0E-1	2.4E-1	2.9E-1	1.8E-1	2.0E-1	1.1E-1	1.3E-1	8.0E-2
$R_3(\text{pgnm}^{-2})$	4.0E-1	2.6E-1	2.9E-1	1.8E-1	2.0E-1	1.2E-1	1.3E-1	7.2E-2
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	8.8E-4	7.3E-4	3.0E-4	3.5E-4	2.6E-4	1.3E+0*	5.7E-5	3.4E-5
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.6E-3	7.6E-3*	7.7E-4	1.5E-3	3.0E-4	4.6E-4	7.2E-5	4.2E-5
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	5.3E-1*	6.0E-2*	3.9E+2*	4.0E+2*	3.4E-4	3.6E-4	8.4E-5	4.7E-5

(c) Wk1d12 reacted against multiple antigen.

**Table 6.3:** Estimated parameters for each dataset for the ERC2 model, with parameters with no SDLN or an SDLN greater than 0.2 marked with an asterisk.

marked as NWD is  $k_m$ , and for values greater than 7.6E-3.

In Subsection 5.1.3 it was noted that the L2 Model is also a limit for the ERC2 model as  $k_m$  tends towards infinity. Figure 5.7 illustrated both that as the value of  $k_m$  increased the trajectories of the state variables of the L1 and ERC1 models got closer together, but that as  $k_m$  increased changes in  $k_m$  had smaller effects on the trajectories.

A physical interpretation of large values of  $k_m$  being marked as NWD in these fits is that in the receptor layer at some interaction spots the analyte was generally better mixed than at others.

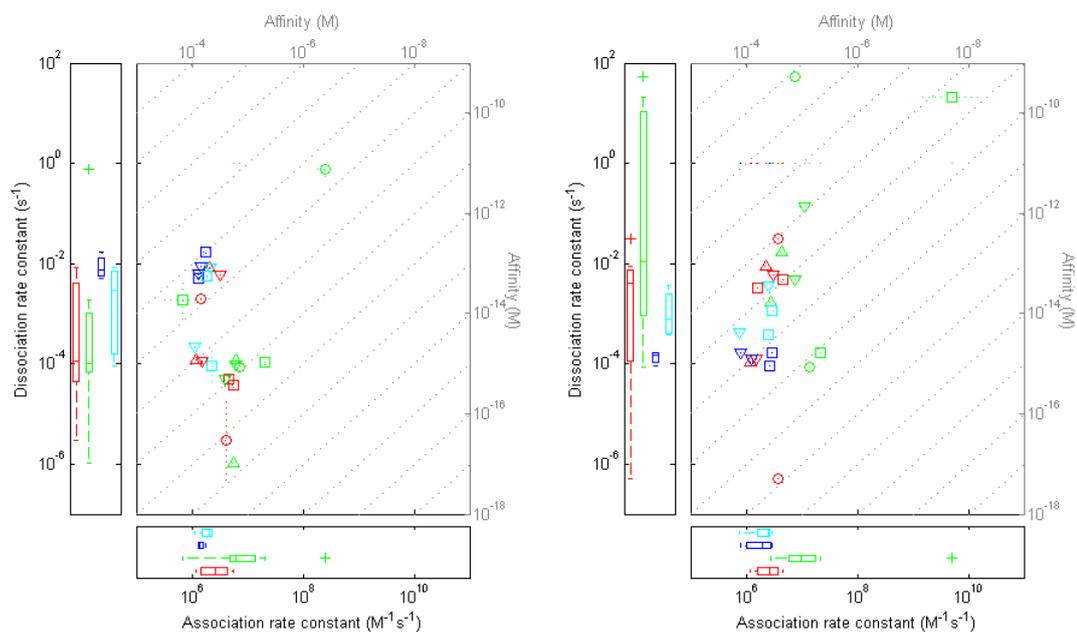
This makes sense because (Edwards., 2001.) demonstrates that the rates at which analyte comes into the receptor layer is dependent on the combined effects of transport and the receptor layer.

In Fig. 6.6(a) and Fig. 6.6(c) we see scatter graphs of the association and dissociation rate constants, for both models and for both primary and secondary dissociation rate, along with confidence intervals for the estimates, and box-plots showing the variation of estimates for a single parameter. As the affinity is the quotient of the dissociation rate constant divided by association rate constant, affinity can also be read on these graphs.

It should be noted that due to the structure of these models, the primary and secondary rate constants may not necessarily represent the rate constants for dimeric and trimeric binding.

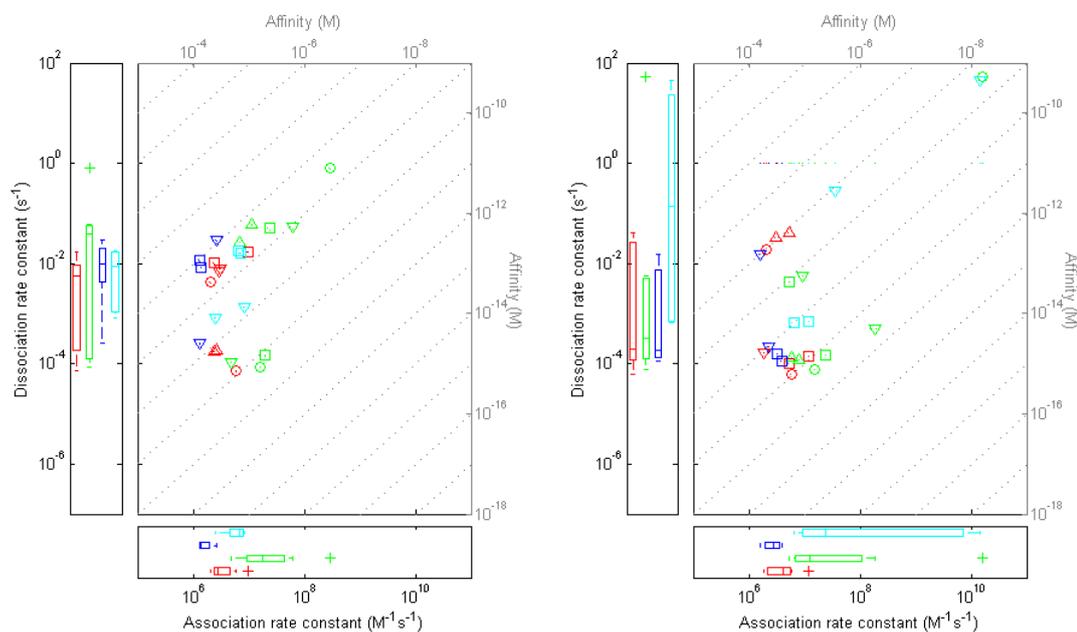
The primary constants estimated by the ERC model are clustered over a smaller area and in many cases have much tighter confidence intervals. However, for neither model is there much consistency in the estimates of any constant, either for any interaction or for any interaction at a given density and concentration. The only exception to this is that the Langmuir model estimates the rate constants WK1D12 vs B7, for each concentration and replicate to be within one order of magnitude of each other. However this appears not to be due to the model fitting the data well, as the RSSs for these fits are amongst the highest for the model (see Fig. 6.4).

We also see no real separation of the estimates of parameters between any interaction. That is, it is unclear based on the estimates of either model, which interactions have



(a) L2 model estimates for affinity, and rate constants for the primary interaction.

(b) L2 model estimates for affinity, and rate constants for the secondary interaction.



(c) ERC2 model estimates for affinity, and rate constants for the primary interaction.

(d) ERC2 model estimates for affinity, and rate constants for the secondary interaction.

**Figure 6.6:** Scatter graphs of estimates of association and dissociation rate constants, with boxplots and secondary axis representing affinity. See Figure 6.4 for legend

higher or lower rate constants.

The graphs also show the affinity of the interactions. Because of the way in which affinity is calculated, datasets that have outlying rate constant estimates, have similar estimated, affinities to those of other datasets for the same interaction. Although the only interaction for which all estimates of affinity were within one order of magnitude was WK1D12 vs B7.

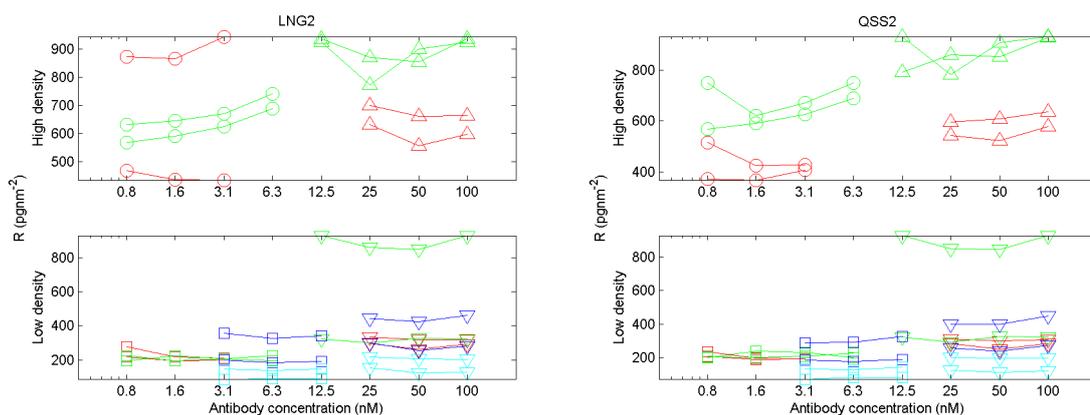
In Fig. 6.7 estimates of  $R$  and  $k_m$  are graphed for either model. Figure 6.7 (a) and (b) illustrate that generally estimates of  $R$  are higher for the high antigen density experiments than the low ones, but that for a given interaction the experiment with higher antibody concentration had higher estimates of  $R$ , and that both models generally gave similar estimates of  $R$  for data from the same interaction spot. Figure 6.7 (c) illustrates the wide range of  $k_m$  constants estimated.

Although the ERC model was able to fit the data with a reduced RSS it was not able to improve the consistency of the estimates of rate constants. This undermines its value for use in affinity estimation for monoclonal antibody - antigen interactions.

A key reason for this is that although it allows for two different kinds of bound state between the analyte and ligand, it does not allow movement from one state to the other without dissociation. As a result, it is not clear how these two bound states relate to the two bound states observed between IgG antibody and antigen, or how their affinities relate to that of the antibody-antigen pairs. Thus, both of these models were rejected for further work on monoclonal antibody.

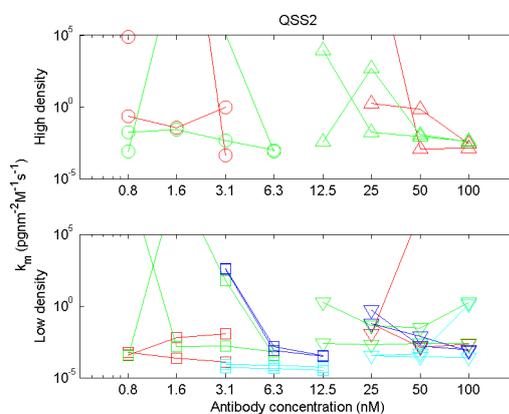
#### 6.4.2 Testing the BEREC model against the BL model

The bivalent Langmuir model was selected to test other models against. Like the two concentration ERC model, it is relatively simple and is one of the models used by the



(a) L2 model estimates for  $R$ .

(b) ERC2 estimates for  $R$ .



(c) ERC2 model estimates for  $k_m$ .

**Figure 6.7:** Estimates of  $k_m$  and  $R$  for the L2 and ERC2 models. See Figure 6.4 for legend

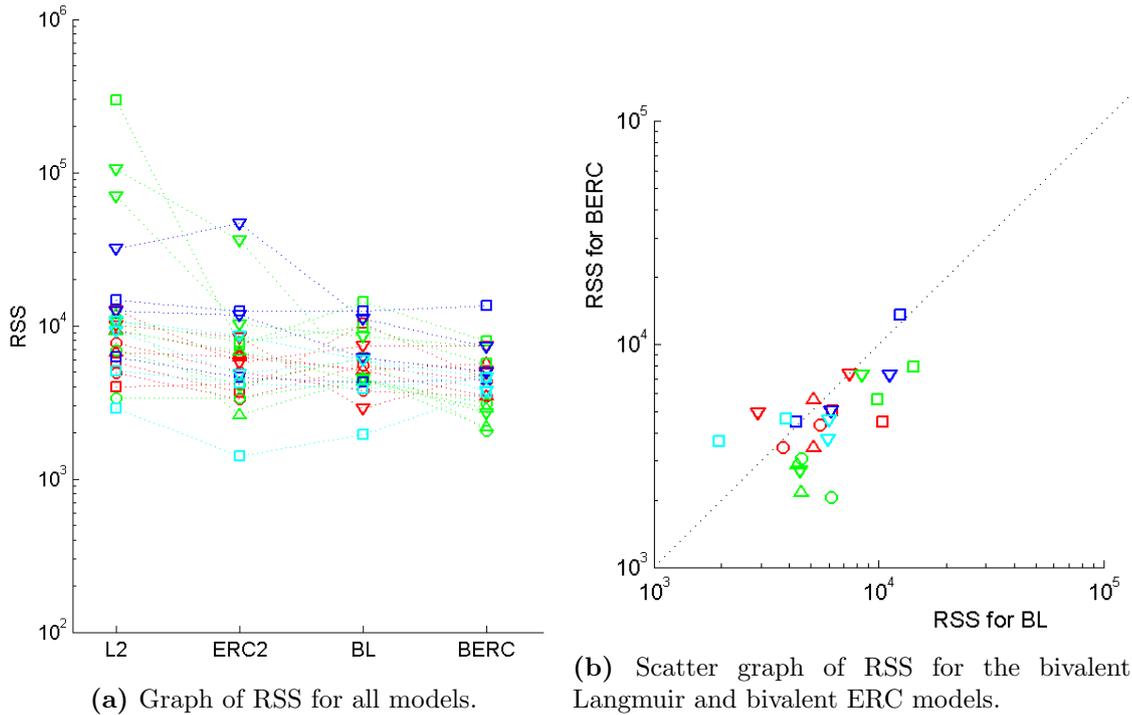
ProteOn Manager software.

The bivalent Langmuir model is structurally globally identifiable in the case where the concentration of the antibody sample is known. This is because it assumes that analyte in contact with the ligand is in the same concentration as analyte in the sample, it may also inaccurately estimate  $k_{a1}$ .

The bivalent ERC model corrects this problem, but unlike the two configuration ERC model it is structurally globally identifiable.

	Minimum	Maximum	Step size	Innitial estimate
$k_a$ ( $M^{-1}s^{-1}$ )	1E+4	1E+11	1E+0.1	1E+8
$k_a$ ( $nm^2pg^{-1}M^{-1}s^{-1}$ )	1E-4	1E+4	1E+0.1	1E+0
$k_d$ ( $s^{-1}$ )	1E-5	1E+2	1E+0.1	1E-2
$R$ ( $pgnm^{-2}$ )	1E-3	1E+3	1E+0.1	1E+0
$k_m$ ( $pgnm^{-2}M^{-1}s^{-1}$ )	1E-6	1E+6	1E+0.2	1E+0

**Table 6.4:** Details of parameter space considered by Differential Evolution.



**Figure 6.8:** Graphs comparing RSS. See Figure 6.4 for legend.

The experimental data was divided in the same way as it was with previous models (see Fig. 6.3), Facsimile programs were written to fit the BERC and BL models to each data-set, and Differential Evolution was used to manage a population of parameters, that were fed into Facsimile to be locally varied. The properties of the parameter space considered by Differential Evolution are shown in Table 6.4.

Figure 6.8(a) illustrates that both the bivalent model and the bivalent ERC have improved RSSs compared to the models in the earlier section.

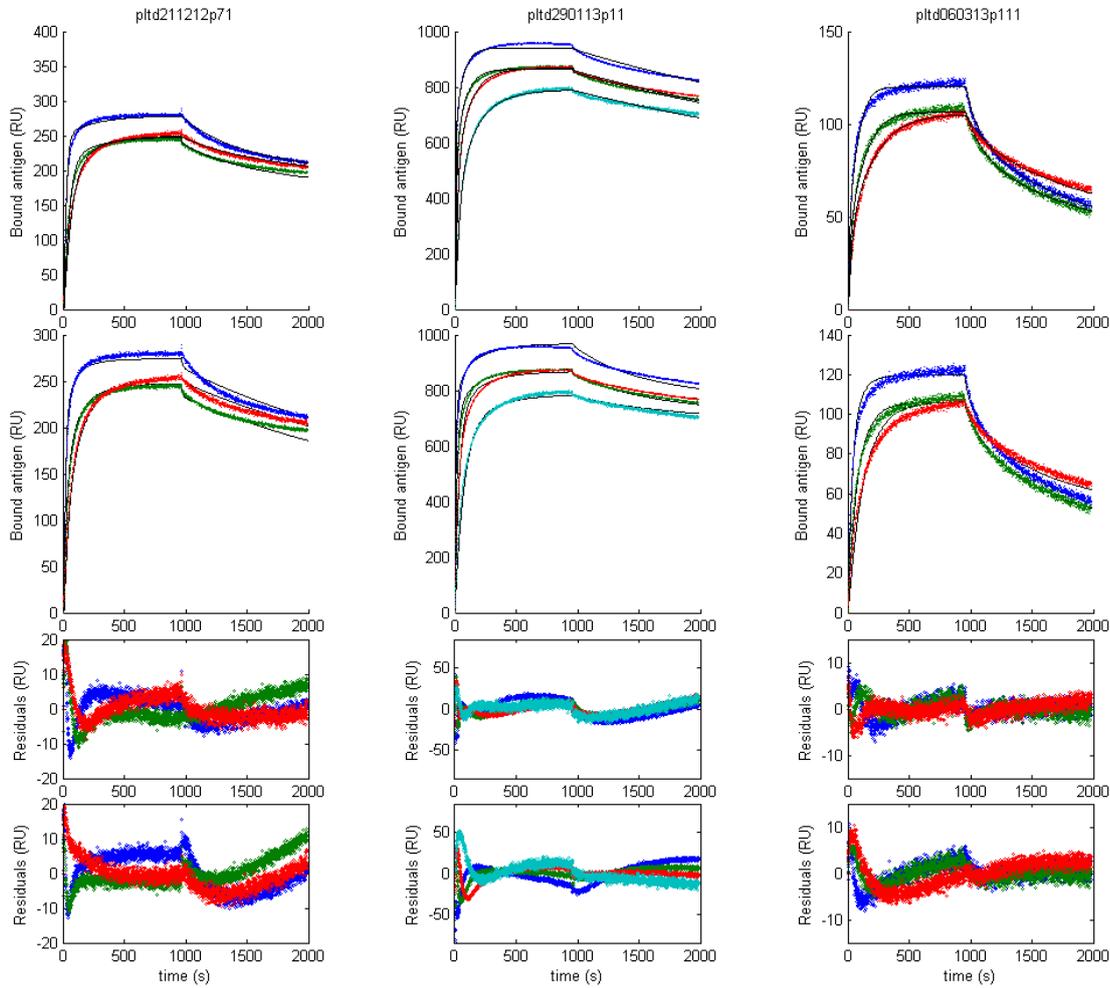
Figure 6.8(b) shows the RSS for the BL and BERC models fitted to each data-set. Similar to the two concentration models, the output of the BL model is a limiting case of the BERC model. So there are parameter values for which the bivalent ERC model would have fit at least as well to the data as the bivalent Langmuir model. With greater simulation time we would see these points moved beneath the line of equality.

The BERC model has a better RSS in most cases and halves the RSS for fits to five datasets. In four of these datasets the interaction is SN230 vs HLA-A2, in the fifth it is the SN607D6 vs A2 interaction.

The high affinity of this interaction and the high density of the antigen means that concentration of free antibody able to bind on the chip's surface is much lower than in the sample. Notably these cases include both high concentration and low concentration experiments, showing that in this case the ERC approximation is useful for low concentration experiments as well.

Similar to the previous two models compared, the improvements to RSS are limited by some features of the data. For neither model does there exist a set of parameters for which the outputs of two lanes of separate concentration may cross except at the start of the association phase. This happens in three of the eight SN607D8 vs A2 samples and five out of the eight SN230G6 vs A2 samples and all of the WK1D12 vs B7 data. Additionally, of the data-sets in which there are no crossovers, there are humps in three of the SN607D8 vs A2 data-sets, which similarly are a feature that the model cannot fit to.

Figure 6.9 shows examples of fits of these models to the data for each antibody. These particular data-sets were chosen because they illustrate the typical strengths and weaknesses of either models when fitting the data. By comparing the residual plots it can be observed that the ERC bivalent model has improved the fitting of the association phase. Whilst for both models error is higher at the start of the association phase than at any other point, the error of the ERC bivalent model is lower at this point and reduces faster after it.



(a) SN607D8 low antigen density, high antibody concentration.

(b) SN203D8 high antigen density, high antibody concentration.

(c) WK1D12 low antigen density, high antibody concentration.

**Figure 6.9:** Comparisons of fit of model and residuals of model for each antibody, with fit for BERC model in the top panel, fit for BL model in second from top panel and their residuals in the third and fourth from top panels respectively.

Whilst the ERC bivalent model fits better in the association phase, it fits little better than the bivalent model in the dissociation phases of these examples. Typically it struggles to fit a curve that remains within the data throughout the phase and instead passes in and out of the data during the phase.

In Tables 6.5 - 6.6 the parameters estimated by these fitting processes are shown. These parameters are visualised in Fig. 6.10 and 6.11.

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	8.9E+5	9.3E+5	4.7E+5	1.1E+6	1.5E+6	4.4E+6	9.6E+6	3.6E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	3.2E-2	4.4E-2	4.6E+4*	1.7E-1	1.1E-2	1.9E-1	5.6E-2	1.5E-1
$k_{d1}(s^{-1})$	5.8E-2	7.1E-2	9.4E+3*	7.3E-2	2.0E-2	5.1E-1	1.9E-1	1.1E-1
$k_{d2}(s^{-1})$	9.9E-5	8.8E-5	5.5E-4	1.5E-4	3.8E-5	6.4E-5	4.2E-5	9.1E-5
$R_1(pgnm^{-2})$	1.4E+3	1.2E+3	6.8E+2	5.8E+2	1.2E+3	1.0E+3	5.4E+2	4.8E+2
$R_2(pgnm^{-2})$	1.4E+3	1.2E+3	6.7E+2	5.2E+2	1.2E+3	1.0E+3	5.2E+2	4.8E+2
$R_3(pgnm^{-2})$	1.4E+3	1.3E+3	6.8E+2	5.8E+2	1.2E+3	1.1E+3	5.6E+2	5.3E+2

(a) Sn607 against HLA-A2.

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	2.5E+6	2.5E+6	2.5E+6	3.6E+6	8.6E+6	1.9E+7	1.0E+7	1.4E+7
$k_{a2}(nm^2pg^{-1}s^{-1})$	2.5E+3	7.1E+0	6.7E+0	7.6E+1*	7.8E+0	3.1E-1	2.0E-1	1.2E+0*
$k_{d1}(s^{-1})$	9.6E-4	1.1E-3	1.0E-3	7.0E-3	1.3E-4	2.0E+0	9.5E-2	2.2E-4
$k_{d2}(s^{-1})$	4.4E+0	1.1E-2	1.2E-2	6.6E-2*	1.1E+0	5.8E-5	1.1E-4	1.1E-1*
$R_1(pgnm^{-2})$	1.6E+3	1.6E+3	1.6E+3	5.8E+2	8.0E+2	1.6E+3	4.4E+2	2.3E+2
$R_2(pgnm^{-2})$	1.5E+3	1.6E+3	1.5E+3	5.9E+2	7.5E+2	1.5E+3	4.8E+2	2.1E+2
$R_3(pgnm^{-2})$	1.6E+3	1.4E+3	1.6E+3	5.4E+2	7.3E+2	1.5E+3	5.4E+2	2.0E+2
$R_4(pgnm^{-2})$	1.5E+3	1.5E+3	1.5E+3	5.2E+2	6.7E+2	1.5E+3	4.8E+2	2.0E+2

(b) Sn230 against HLA-A2.

	B7				B27			
	LdHc	LdHc	LdLc	LdLc	LdHc	LdHc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	6.1E+5	6.1E+5	1.4E+9*	1.3E+6	6.8E+5	3.9E+5	1.1E+6	1.1E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	1.3E-1	1.1E-1	3.7E-1	9.8E-2	1.7E-1	2.0E-1	1.8E-1	6.9E-2
$k_{d1}(s^{-1})$	8.9E-2	7.7E-2	5.0E+2*	8.3E-2	5.0E-2	3.7E-2	6.2E-2	2.4E-2
$k_{d2}(s^{-1})$	1.3E-4	1.5E-4	7.0E-5	1.1E-4	6.5E-4	1.1E-3	3.9E-4	3.3E-4
$R_1(pgnm^{-2})$	9.8E+2	5.9E+2	8.2E+2	4.9E+2	4.4E+2	2.8E+2	3.8E+2	2.2E+2
$R_2(pgnm^{-2})$	8.9E+2	5.4E+2	7.8E+2	4.8E+2	4.5E+2	2.8E+2	3.6E+2	2.3E+2
$R_3(pgnm^{-2})$	9.0E+2	5.8E+2	8.0E+2	4.9E+2	4.6E+2	3.0E+2	3.6E+2	2.2E+2

(c) Wk1d12 reacted against multiple antigen.

**Table 6.5:** Bivalent Langmuir.

There is an increase in the number of parameters estimated with the novel model marked with an asterisk for being NWD. Both the L2 and BL model have 8 of 176 parameters marked, and the ERC2 model had 27 of 256 but the BERC model has 43 of 256. The majority of these are NWD parameter estimates are for  $k_m$  at various interaction spots, with high values.

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	3.0E+6	2.0E+6	3.5E+6	1.4E+6	2.6E+6	3.3E+6	3.1E+6	9.0E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	6.9E-1	5.1E-2	1.3E+2*	3.5E-1	1.7E-2	6.8E-3	4.6E-3	3.4E+0*
$k_{d1}(s^{-1})$	2.1E-4	2.0E-4	3.3E-4	1.3E-3	2.3E-2	1.3E-4	1.2E-4	1.9E-4
$k_{d2}(s^{-1})$	3.1E-2	5.6E-3	5.8E+0	7.9E-4	3.4E-5	8.9E-4	5.1E-4	1.2E-1*
$R_1(pgnm^{-2})$	6.5E+2	5.9E+2	3.1E+2	4.4E+2	1.0E+3	4.7E+2	2.0E+2	2.0E+2
$R_2(pgnm^{-2})$	6.4E+2	5.3E+2	3.0E+2	4.0E+2	1.0E+3	4.6E+2	2.0E+2	2.1E+2
$R_3(pgnm^{-2})$	6.4E+2	5.6E+2	3.0E+2	4.4E+2	1.0E+3	5.5E+2	2.0E+2	2.2E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	1.3E-2	6.3E-3	3.3E-3	7.9E-3	5.7E-4	5.1E+0*	1.3E-2*	2.0E-4
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	8.1E-3	3.5E-3	2.5E-2*	2.3E-3	1.0E-3	1.2E+1*	1.2E-2*	2.5E-4
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	5.5E-3	8.0E+0*	1.0E+0*	5.7E+1*	9.8E-1*	7.2E+0*	1.3E-3	3.9E-4

(a) Sn607 against HLA-A2.

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	1.8E+8	1.4E+7	2.0E+7	4.3E+6	2.0E+7	4.3E+17*	2.2E+8*	4.6E+7
$k_{a2}(nm^2pg^{-1}s^{-1})$	1.4E-1	2.5E-1	2.6E-1	8.0E+1	2.8E-2	3.9E-2	7.9E-2	7.6E-2
$k_{d1}(s^{-1})$	1.4E+1	1.5E-4	1.5E-4	6.4E-3	1.5E-1	3.9E+9	8.7E-1*	1.7E-1
$k_{d2}(s^{-1})$	6.4E-5	1.4E-2	1.2E-2	7.3E-2	4.4E-5	4.7E-5	7.9E-5	8.7E-5
$R_1(pgnm^{-2})$	1.9E+3	9.3E+2	9.4E+2	5.7E+2	1.7E+3	1.5E+3	4.4E+2	5.0E+2
$R_2(pgnm^{-2})$	1.8E+3	9.1E+2	8.6E+2	5.9E+2	1.6E+3	1.4E+3	4.8E+2	4.7E+2
$R_3(pgnm^{-2})$	1.8E+3	7.8E+2	8.7E+2	5.3E+2	1.5E+3	1.4E+3	5.3E+2	4.5E+2
$R_4(pgnm^{-2})$	1.8E+3	8.4E+2	8.1E+2	5.2E+2	1.4E+3	1.4E+3	4.6E+2	4.4E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	5.9E-3	5.9E-3	5.7E-3	4.4E-3	9.3E-4	7.8E-4	2.6E-4	3.3E-4
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	1.3E-2	8.8E-3	7.0E-3	1.4E-2	1.8E-3	9.7E-4	3.8E-4	5.3E-4
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	2.7E-2	1.0E-2	8.5E-3	2.1E-2	1.7E-3	9.6E-4	4.9E-4	6.5E-4
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	1.1E+16*	1.6E-2	1.0E-2	2.6E+3*	7.1E-3	9.7E-4	7.5E-4	1.0E-3

(b) Sn230 against HLA-A2.

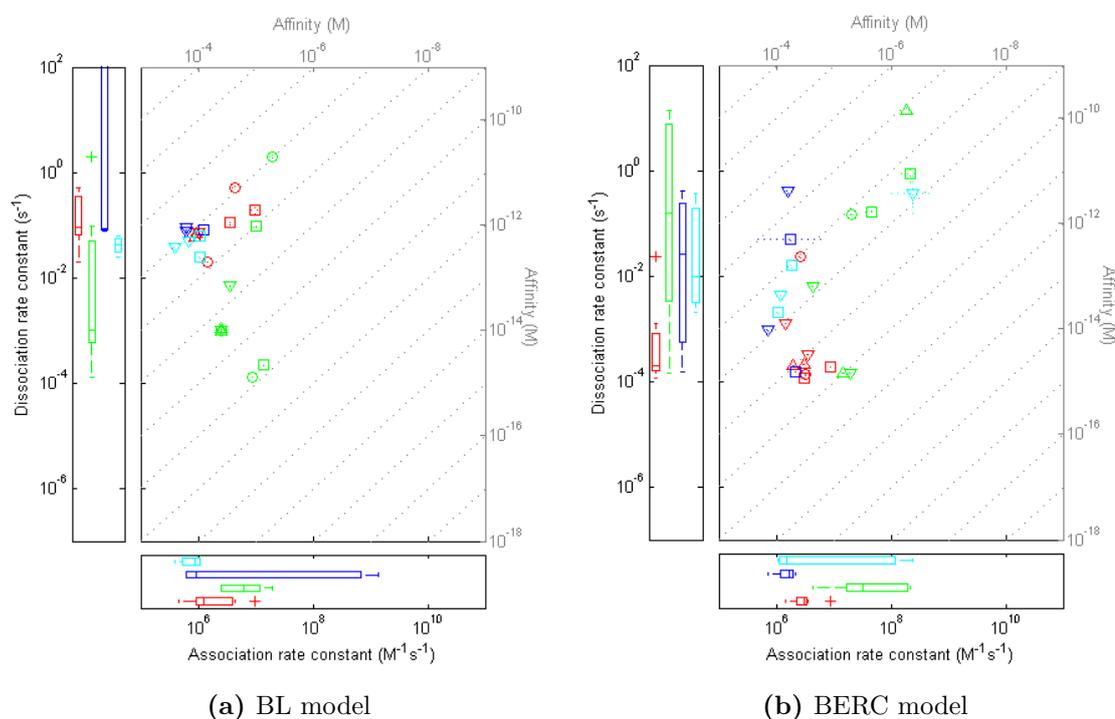
	B7				B27			
	LdHc	LdHc	LdLc	LdLc	LdHc	LdHc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	1.6E+6	7.1E+5	2.2E+6	1.8E+6*	1.2E+6	2.4E+8*	1.8E+6	1.1E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	3.3E-1	1.9E+0*	2.2E-2	1.7E-1*	4.6E+0	4.6E+0	1.1E-1	1.3E-2
$k_{d1}(s^{-1})$	4.2E-1	9.5E-4	1.5E-4	5.1E-2	4.4E-3	3.7E-1*	1.6E-2	2.0E-3
$k_{d2}(s^{-1})$	1.2E-4	8.9E-3	7.8E-3	1.1E-4	2.5E-2	4.6E-2	3.5E-4	1.4E-7*
$R_1(pgnm^{-2})$	9.6E+2	4.1E+2	3.3E+2	3.3E+2	3.3E+2	1.7E+2	3.2E+2	1.5E+2
$R_2(pgnm^{-2})$	8.8E+2	3.7E+2	2.9E+2	3.3E+2*	3.4E+2	2.2E+2	3.0E+2	1.6E+2
$R_3(pgnm^{-2})$	8.9E+2	3.9E+2	2.9E+2	3.5E+2*	3.5E+2	2.4E+2	2.9E+2	1.5E+2
$R_4(pgnm^{-2})$	9.6E+2	4.1E+2	3.3E+2	3.3E+2*	3.3E+2	1.7E+2	3.2E+2	1.5E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	1.2E+6*	5.6E-3	6.3E-4	1.2E+6*	1.6E-3	6.2E-5	2.9E-4	8.0E+3*
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	1.2E+6*	1.6E+5*	1.3E-3	1.0E+6*	7.8E-2	5.6E-6*	8.3E-4*	1.2E+4*
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	1.2E+6*	1.9E+8*	1.4E+3*	1.0E+6*	2.1E-2	5.6E-6*	2.6E-3	1.0E+4*

(c) Wk1d12 reacted against multiple antigen.

**Table 6.6:** Bivalent ERC.

In Subsection 5.1.3 it was noted that the L2 Model is also a limit for the ERC2 model as  $k_m$  tends towards infinity. Figure 5.7 illustrated both that as the value of  $k_m$  increased the trajectories of the state variables of the L1 and ERC1 models got closer together, but that as  $k_m$  increased changes in  $k_m$  had smaller effects on the trajectories. Again, this makes sense because the BL model is the limit of the BERC model as  $k_m$  tends towards infinity, and because we would expect there to be variation in the mixing of the analyte

in the receptor later between interaction spots.



**Figure 6.10:** Scatter graphs of estimates of  $k_{a1}$  and  $k_{d1}$ , with boxplots and secondary axis representing affinity. See Figure 6.4 for legend.

Figure 6.10(a) and 6.10(b) show estimates of the rate constants for both models. Unlike the estimates in the previous section there seems to be more of a pattern to the estimates here. That is, estimates are clustered to some degree by the underlying antibody - antigen interaction that was being measured in that experiment. In many cases in 6.10(a), points representing one interaction have as their closest neighbour points representing the same interaction, either measured with a different concentration of antibody or density of antigen, or a replicate of the same conditions. However one parameter estimate cannot be placed on the axis, because its parameter estimates took extreme values ( $k_{a1}$  of  $4.3E+17s^{-1}$  and  $k_{d1}$  of  $3.E+9M^{-1}s^{-1}$ ).

In 6.10(b) we see the estimates of primary affinity grouped according to interaction loosely

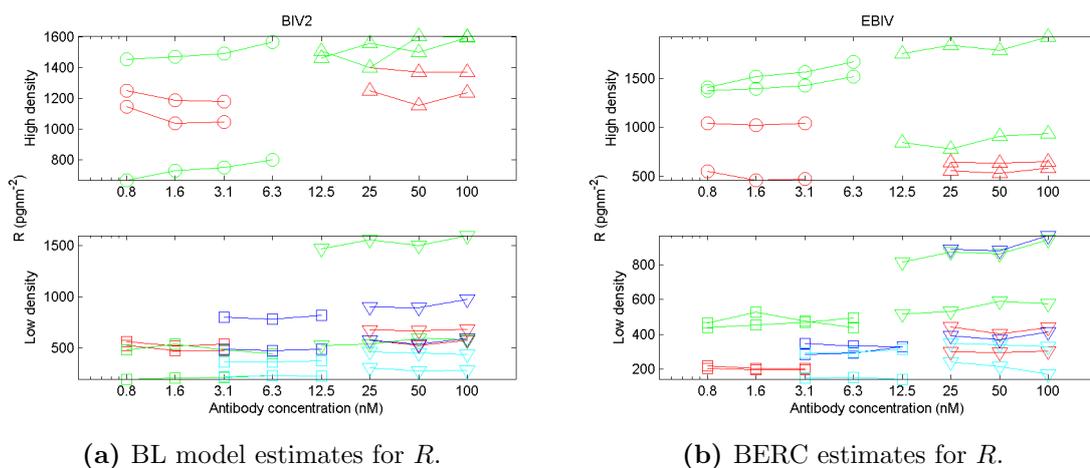
grouped into columns where there is little variation between association constant, and greater variation in dissociation constant. This suggests that the estimates of association rate constant may be more reliable, whereas the estimates of dissociation rate constant may not, in this model, be as reliable.

Estimates of secondary binding constants seem clustered along lines of equal affinity. This is because the models which were the most successful at minimising RSS retained similar ratios of secondary association rate constant and secondary dissociation rate constant, even though the estimates of these constants varied considerably. This clustering has the effect that these models would have similar proportions of bound antibody bound in each state, even though the rates at which antibody would move into and out of these states could vary considerably.

In Subsection 5.1.4 it was noted that in cases where there were no conformational changes to the antibody when it binds bivalently, that alter the rate at which it would disassociate from either antigen, then the primary and secondary dissociation rate constants should be equal. This clustering demonstrates that the models most successful in minimising RSS behave in very similar ways in terms of the ratios of their state variables - and as a result undermines the value of varying primary and secondary dissociation rate constants independently. Whilst these experiments cannot demonstrate that no conformational change is happening that alters these secondary disassociation rates, expanding parameter space to allow for such effects appears to be having little effect on minimising RSS.

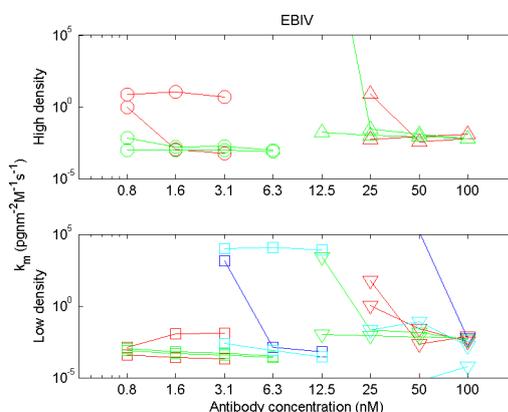
### 6.4.3 The bivalent ERC model with spatial effects

In half of all the datasets we see a crossing over of the response from interaction spots with separate concentrations (this may be viewed in Figures 4.2, 4.4 and 4.6). As a result, it is necessary to develop a model that can deal with this feature of the data. One explanation for these results is that not all antigen on each chip may be positioned so that an antibody



(a) BL model estimates for  $R$ .

(b) BERC estimates for  $R$ .



(c) BERC model estimates for  $k_m$ .

**Figure 6.11:** Estimates of  $k_m$  and  $R$  for the L2 and ERC2 models. See Figure 6.4 for legend.

once monovalently bound to one, may bind bivalently to another. On an interaction spot where there was a large number of spatially isolated antigen, we may see a more rapid increase in response than we would see on an interaction spot with a smaller number of antigen with a higher density. This is because when an antibody binds monovalently, there is an increase in the response signal, but when an antibody moves from being bound monovalently to bivalently, there is no increase in response signal, and the antigen that is bound becomes unavailable for antibody in flow to bind to. As a result, an interaction spot with more isolated antigen will also see a higher saturation point.

In the dissociation phase an interaction spot with more isolated antigen will have more antibody monovalently bound than otherwise. As a result, we will see a more rapid dissociation in the first part of the phase before the rate at which monovalent antibody dissociates becomes dependent on the combined effect of both dissociation rates.

These effects could explain the behaviour of the data, and allow for better parameter estimation in the half of the datasets with these features. As a result the bivalent model with spatial effects (presented in Subsection 5.1.6) was developed.

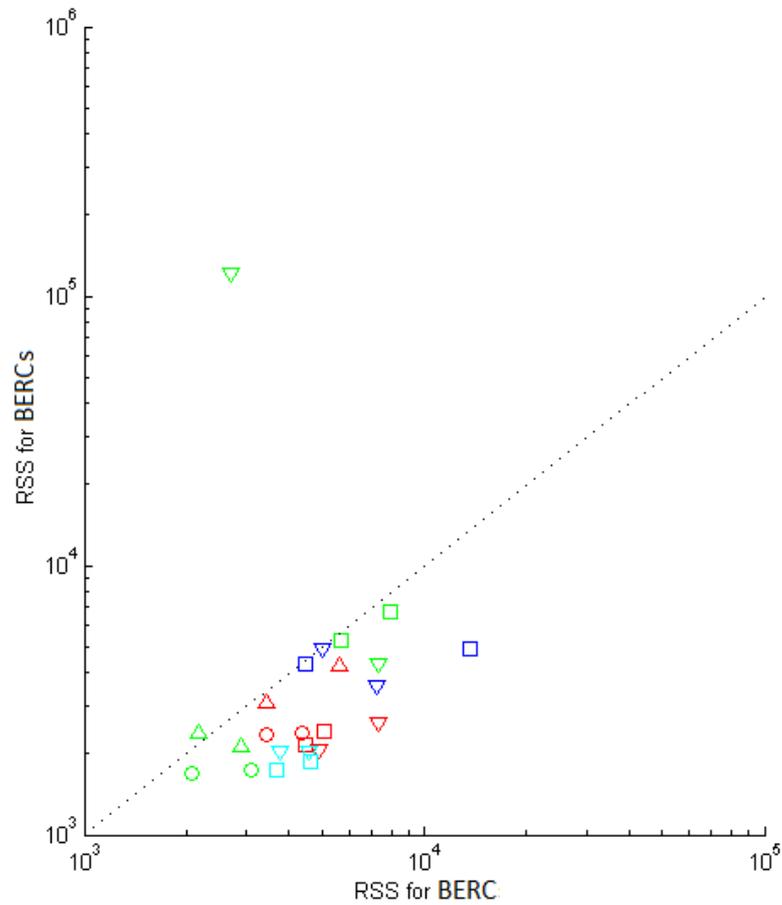
The experimental data was divided in the same way as it was with previous models (see Fig. 6.3), Facsimile programs were written to fit the BERC and BL models to each dataset, and Differential Evolution was used to manage a population of parameters, that were fed into Facsimile to be locally varied. The properties of the parameter space considered by Differential Evolution are the same as those for the BL and BERC model, and are shown in Table 6.4.

This model introduces two new parameters,  $R_b$  and  $R_m$ . These were fitted via a series of transformed parameters inside each Facsimile program, so that the files output by Differential Evolution could be consistent across the models. A new parameter  $S_b$  represented the proportion of antigen binding on which bivalent binding was possible was introduced so  $R_b$  and  $R_m$  could be varied together, with:

$$R_b = RS_b \tag{6.4}$$

$$R_m = R(1 - S_b). \tag{6.5}$$

As a result  $S_b$  had to vary between 0 and 1. Facsimile does not allow for parameters to be varied on an interval smaller than the largest and smallest real numbers it can store. As a result another parameter was introduced  $T_b$  which could be varied on the positive real numbers, and transformed to give a value on the interval  $(0, 1)$ . the parameter  $S_b$  was



**Figure 6.12:** Scatter graph of RSS for the BERC and BERCs model. See Figure 6.4 for legend.

calculated from  $T_b$  with the formula:

$$S_b = \frac{1}{1 + T_b}. \quad (6.6)$$

Each Facsimile program began by assigning a value of 1 to  $T_b$  and varied it with together with the other parameters for estimation.

Figure 6.12 displays the RSS of the BERCs model plotted against those of the bivalent Langmuir model. In all but two cases the BERCs model gives a lower RSS than those

given by the BERC model. For all outputs that may be generated with the BERC model may there is a set of parameters for which the BERCs model will produce the same output for the same input. As a result, these two points should be seen as a result of insufficient time spent fitting the models for parameter space to be fully explored, rather than the inferiority of the BERCs model.

In other cases RSSs were reduced by up to a factor of three. As expected the biggest reductions in RSS came in the datasets where crossovers were observed.

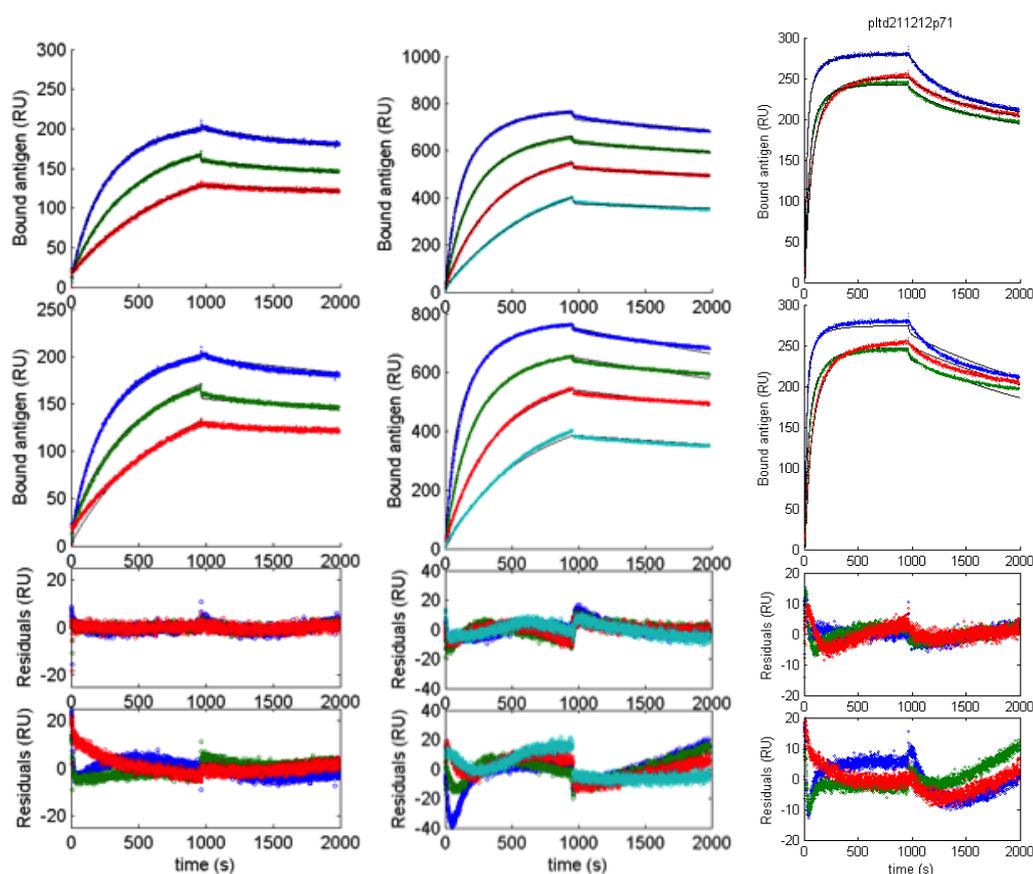
Figure 6.13 illustrates three datasets where the BERCs model outperforms the BERC model. In 6.13(a) (top) we see the BERCs model fitting the data well, the estimated response (black line) is within the data (coloured dots) at all concentrations (Blue 6.25nM Green, 3.12nM; Red, 1.51nM).

In 6.13(a) (middle), in the association phase (0-967s), the BL model underestimates the output in the 1.151nM lane (Red) and overestimates it in the 3.125nM lane (green), in the dissociation phase (967-2000s) its estimates of the 1.151nM lane (Red) stay within the data, but its estimates of the 3.125nM lane (green) and the 1.151nM lane (Red) are lower than observed for the most of the phase.

In 6.13(a) (bottom) we see the residuals for these two fits. The residuals for the BERCs model remain close to zero throughout both phases, although there appears to be a slight step in the residuals for the 6.25nM (blue) channel. However the residuals for the BL model vary in a non-random fashion.

In 6.13(b) what we see is similar to 6.13(a). In this case the reaction is SN607D8 vs A2 and there are four concentrations of analyte, but the data and the model behave similarly in either phase to what was presented in the previous figure.

Figure 6.13(c) shows the improvements made by the BERCs model. Notably these are much greater than those illustrated in the previous two figures. This is because the data



(a) SN607D8 high antigen density, low antibody concentration. (b) SN230G6 high antigen density, high antibody concentration. (c) SN607D8 high antigen density, low antibody concentration.

**Figure 6.13:** Comparisons of fit of model and residuals of model for each antibody, with fit for BERCs model in the top panel, fit for BERC model in second from top panel and their residuals in the third and fourth from top panels respectively, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.3.

the models were fitted to has the feature the BERCs model was designed to deal with, i.e. the crossing over of the response time series.

The problems illustrated in this figure are common to all the SPR experiments in the dataset. Both models allow analyte to bind in two configurations throughout the association phase; the Langmuir model assumes the concentration is uniform in the channel and

as a result wrongly estimates the response and the on-rate constants, badly fitting the association phase.

In each experiment multiple concentrations of antibody are used but the concentrations of antibody able to bind to ligand are not necessarily proportional to the concentrations of antibody used. As a result the model was often able to fit the middle channel well, but not either the top or the bottom channel.

In the small number of models where the bivalent ERC model fits well in the association phase, it has the wrong proportions of antibody bound in each configuration at the end of the phase, so in the dissociation phase (967-2000s), the model is still unable to fit the data well.

Because the bivalent ERC model assumes all ligands are the same, and can bind into trimers, it has a number of problems: in experiments where few ligand are isolated, antibody dissociates much more rapidly from the few isolated ligand, as it cannot go from dimer into trimer; similarly in experiments where most ligand are isolated, at the end of the dissociation phase most of the response is generated by ligand that are bound in trimers.

In Table 8.3 the parameters estimated by these fitting processes are shown. These parameters are visualised in Fig. 6.14. The proportion of parameters given an asterisk for having a NWD estimate remains similar to the BERC model, at 78 out of 338.

In Fig. 6.14 we see scatter graphs of the association and dissociation rate constants, for the BERCs model, along with confidence intervals for the estimates, and boxplots showing the variation of estimates for a single parameter. As the affinity is the quotient of the dissociation rate constant divided by association rate constant, affinity can also be read on these graphs.

This graph shows that experiments are clustered by interaction. The WK1D12 vs HLA-

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	2.3E+6	1.9E+6	1.3E+6	1.4E+6	2.6E+6	5.6E+6	2.8E+6	2.5E+7
$k_{a2}(nm^2pg^{-1}s^{-1})$	3.2E-1	3.2E+1	1.9E+0	4.3E+0*	5.5E+1*	9.7E+0*	2.3E+0	3.6E+2
$k_{d1}(s^{-1})$	2.2E-4	1.7E-4	1.5E-3	1.8E-3	9.2E-5	1.0E-4	1.5E-4	1.2E-3
$k_{d2}(s^{-1})$	1.9E-2	1.3E+0	2.8E-3	6.7E-3*	1.7E-4*	1.3E+0*	4.5E-3*	5.4E-1
$R_1(pgnm^{-2})$	6.7E+2	6.0E+2	4.7E+2	4.5E+2	4.2E+2	4.4E+2	2.4E+2	4.1E+2
$R_2(pgnm^{-2})$	6.5E+2	5.4E+2	4.7E+2	4.2E+2	4.7E+2	4.2E+2	2.3E+2	3.8E+2
$R_3(pgnm^{-2})$	6.5E+2	5.7E+2	4.7E+2	4.3E+2	4.5E+2	3.7E+2	2.4E+2	4.1E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	3.6E+0*	3.0E+1*	1.3E-2	6.5E-2*	1.6E+0*	5.9E-4	3.7E-2*	1.2E-2
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	7.4E-2*	2.6E+0*	1.0E-1*	7.2E-3	1.6E+0*	1.0E-3	8.7E-3*	2.8E-3
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	1.8E-2	9.8E+2*	1.1E-1*	3.3E+0*	1.8E+0*	7.9E+1*	1.1E-3	1.5E-2
$T_{1b}$	3.8E-2*	1.8E-1	5.2E-2	3.6E-2*	1.4E-1*	4.0E-7*	3.8E-7	4.4E-2
$T_{2b}$	0.0E+0	7.2E-2	6.8E-2	1.4E-3	1.7E+0	1.2E-4*	1.2E-4	1.1E-2
$T_{3b}$	1.4E-2*	1.2E-1	5.6E-2	3.4E-2	1.9E+0	9.4E-1*	1.6E+1	3.6E-2

(a) Sn607 against HLA-A2.

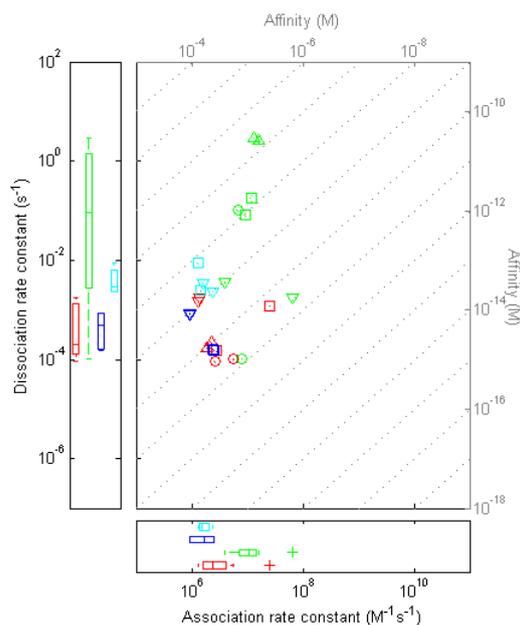
	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	1.3E+7	1.6E+7	6.5E+7	4.0E+6	6.8E+6	7.9E+6	9.3E+6	1.2E+7
$k_{a2}(nm^2pg^{-1}s^{-1})$	3.9E-1	3.2E-1	8.0E-6*	7.7E+1	6.6E-2	1.2E+1*	2.3E-1	4.4E-1
$k_{d1}(s^{-1})$	3.0E+0	2.6E+0	1.7E-3	3.6E-3	1.0E-1	1.1E-4	8.1E-2	1.8E-1
$k_{d2}(s^{-1})$	5.6E-5	5.7E-5	8.1E-5*	4.7E-2	3.5E-5	1.4E+0	7.0E-5	8.7E-5
$R_1(pgnm^{-2})$	2.2E+3	2.2E+3	9.1E+2	5.4E+2	1.9E+3	7.1E+2	5.1E+2	5.6E+2
$R_2(pgnm^{-2})$	2.3E+3	2.2E+3	8.3E+2	5.5E+2	1.8E+3	6.4E+2	5.9E+2	6.3E+2
$R_3(pgnm^{-2})$	2.6E+3	2.2E+3	8.0E+2	5.1E+2	2.6E+3	6.0E+2	6.5E+2	8.1E+2
$R_4(pgnm^{-2})$	6.8E+3	3.5E+3	8.3E+2	4.9E+2	1.1E+5	6.1E+2	4.5E+4	2.6E+3*
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	1.3E-2	7.3E-3	5.2E-3	2.2E-2	3.1E-3	1.1E+0*	8.0E-2*	3.2E-3
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	6.8E-2	1.4E-2	5.4E-3	5.3E+0*	9.4E-1*	6.6E-1*	1.4E-1*	4.7E-2*
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	9.5E+8*	4.5E-2	4.6E-3	2.6E-2	8.2E-1*	7.5E-1*	3.4E-1*	2.5E-3
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	1.0E+4*	7.3E-1	5.1E-3	1.1E+0*	3.7E+1	8.6E-2*	3.9E+2*	3.5E-2*
$T_{1b}$	1.6E-1	1.3E-1	1.0E+0*	2.4E-2	1.3E-1	3.6E-1	1.9E-1	1.8E-1
$T_{2b}$	2.3E-1	2.0E-1	1.0E+0*	3.6E-2	1.5E-1	3.8E-1	2.3E-1	3.0E-1
$T_{3b}$	3.1E-1	2.9E-1	1.0E+0*	4.3E-2	4.3E-1	4.5E-1	2.3E-1	4.8E-1
$T_{4b}$	7.5E-1	4.9E-1	1.0E+0*	7.4E-2	7.3E+1	4.0E-1	9.8E+1*	5.2E+0*

(b) Sn230 against HLA-A2.

	B7				B27			
	LdHc	LdHc	LdLc	LdLc	LdHc	LdHc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	9.3E+5	9.3E+5	2.4E+6	2.5E+6	2.5E+6	1.6E+6	1.5E+6	1.3E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	2.5E+0	1.6E+1*	1.1E+0*	2.4E-1	2.5E+3	1.6E+0	2.2E+1	1.5E-1
$k_{d1}(s^{-1})$	8.2E-4	8.7E-4	1.6E-4	1.5E-4	2.4E-3	3.4E-3	2.4E-3	9.1E-3
$k_{d2}(s^{-1})$	8.5E-3	5.6E-2*	1.0E-1*	2.9E-2	1.8E+1*	8.7E-3	1.2E-1	5.0E-4
$R_1(pgnm^{-2})$	7.3E+2	7.1E+2	3.5E+2	3.4E+2	2.7E+2	3.0E+2	2.5E+2	3.0E+2
$R_2(pgnm^{-2})$	6.4E+2	6.4E+2	3.1E+2	2.8E+2	2.8E+2	2.9E+2	2.4E+2	3.1E+2
$R_3(pgnm^{-2})$	6.5E+2	6.3E+2	2.8E+2	2.6E+2	2.7E+2	3.2E+2	2.4E+2	3.5E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	3.4E-3	3.2E-3	5.2E-4	4.5E-4	1.0E-3	8.3E-4	6.9E-4	9.9E-3
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	6.0E-3	5.4E-3	7.6E-4	6.0E-4	1.3E-3	1.3E-3	4.6E-3*	3.6E-2
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	4.9E-2*	1.4E-2	1.5E+0*	1.7E-2*	1.2E-3	3.0E-3	6.4E-3	4.6E-1
$T_{1b}$	9.5E-4*	2.7E-3	1.5E-2*	6.9E-3*	1.5E-3*	6.8E-3	3.8E-2	5.3E-3*
$T_{2b}$	6.9E-2	5.5E-2	4.3E-2*	3.1E-1	2.6E-4*	6.7E-2	6.6E-2	1.0E-1
$T_{3b}$	5.7E-2	6.0E-2	3.5E-1	5.2E-1	1.7E-2	1.3E-2*	1.1E-1	1.9E-1

(c) Wk1d12 reacted against multiple antigen.

Table 6.7: Bivalent ERC spatial.

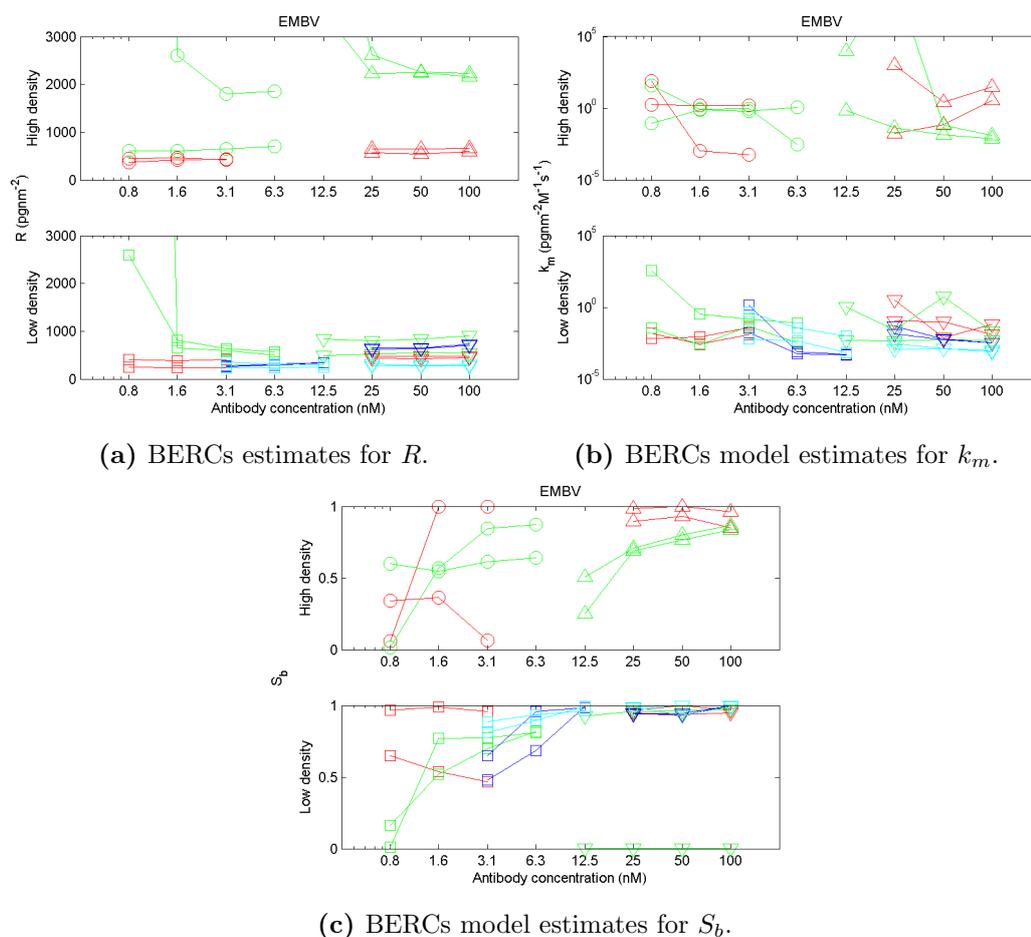


**Figure 6.14:** Scatter graphs of estimates of association and dissociation rate constants, with box-plots and secondary axes representing affinity. See Figure 6.4 for legend.

B27 estimates are clustered together, each within one order of magnitude for affinity estimate  $k_a$  estimate and  $k_d$  estimate. The WK1D12 vs HLA-B7 parameter estimates are grouped by concentration. Although the SN607D8 and SN230G6 vs HLA-A2 experiments are more scattered, they still form loose clusters. Interestingly these clusters are more vertically scattered than horizontally scattered. This raises two questions, whether estimation of  $k_d$  could be improved, and whether this variability in estimates comes from artefacts in the experimental data, the models or the fitting process.

#### 6.4.4 Variations of the BERCs model for the SN230G6 vs A2 dataset

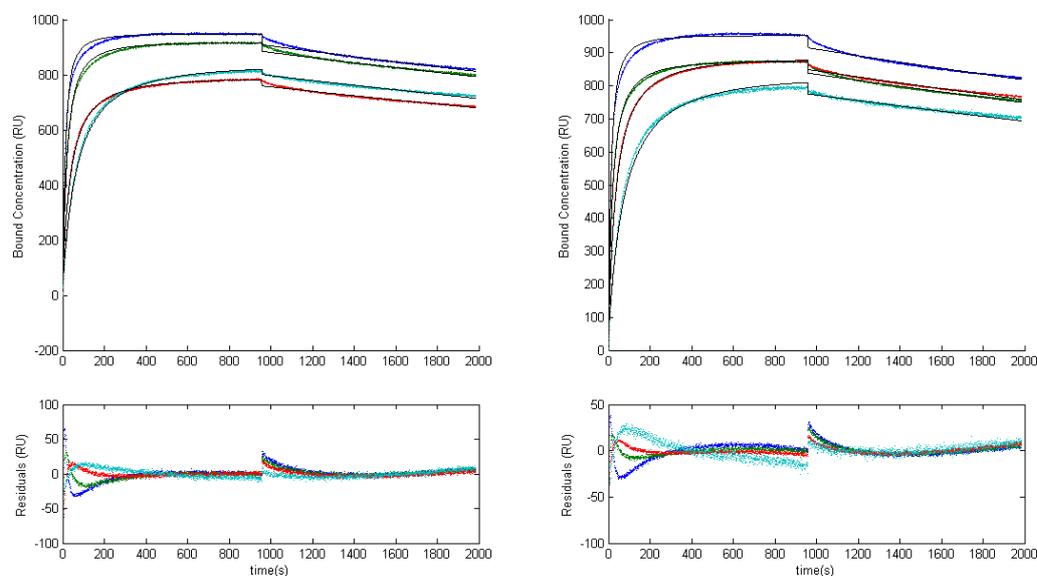
It was observed that in this dataset, the BERCs model fits often included systematic errors different to those observed in other datasets, as well as unusual parameter estimates (values of  $R$  orders of magnitude higher than for other datasets, illustrated in Fig. 6.15(a)).



**Figure 6.15:** Estimates of  $k_m$  and  $R$  for the BERCs model. See Figure 6.4 for legend

These systematic errors illustrated in Fig. 6.16. Whilst the model is generally well fit, it predicts extremely rapid dissociation at the start of the dissociation phase, followed by much slower dissociation throughout the phase. The parameter estimates that create this effect are shown in Table 8.3. Whilst most estimates of  $k_{d1}$  are between  $10^{-4}$  to  $10^{-2}$  s<sup>-1</sup> the estimates for files with this kind of error range from  $10^{-2}$  to  $10^1$  s<sup>-1</sup>, and the  $k_{d2}$  estimates which in other files range from  $10^{-4}$  to  $10^1$  s<sup>-1</sup> range from  $10^{-5}$  to  $10^{-4}$  s<sup>-1</sup>.

These dissociation rate constants mean that the affinity estimated with these datasets ( $10^{-8}$  to  $10^{-6}$ ) is lower than the affinity estimated for the other antibody-antigen inter-



**Figure 6.16:** Fits of the Bivalent ERC model with spatial effects with systematic errors, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.3.

actions ( $10^{-11}$  to  $10^{-8}\text{s}^{-1}$ ). This is a contradiction to other experimental results from outside this PhD project.

Whilst SPR experiments were being conducted on these antibodies a semi-quantitative measure of antibody affinity and abundance was also being used - Luminex. SN230G6 against HLA-A2 interaction had the highest Median Fluorescence Intensity (MFI) for any interaction investigated, suggesting that the interaction also had the highest affinity (Daga, 2015, Table 4-1).

To explore this, the BERCs extended model was developed (see Subsection 5.1.6 specifically Eq. 5.15 and Eq.5.18), and fit to the data. This model allowed the dissociation constants to take different values in the association and dissociation phases of the experiment.

The success of first extended version of the BERCs model motivated a second version (see Subsection 5.1.6 specifically Eq. 5.19) which also allowed the state variables to change

values discontinuously at the start of the dissociation phase ( $t_1$ ).

These variations of the model were motivated by the fact that, in the models where separate structural identifiability results for either phase had been obtained, the dissociation constants had been identifiable in either phase. These systematic errors that result in a characteristically bad fit of the dissociation phase coupled with a good fit of the association phase could be explained by the parameter values required to fit both phases well for these datasets differing between phases. This could be due to a conformational change that the antibody-antigen complex goes through that means its dissociation rate is initially high but becomes lower after a period of time or artefacts in the data

These fits were intended to help establish whether it was features of the data in the association or dissociation phases or both that lead to these unusual fits and unusual parameter estimates.

They were not intended to accurately describe the underlying chemical systems, but to allow us to understand the difficulties the BERCs model had had with these datasets.

Facsimile programs were written to fit both new variations of the BERCs model to each data-set, and Differential Evolution was used to manage a population of parameters, which were fed into Facsimile to be locally varied. The properties of the parameter space considered by Differential Evolution are the same as those for the generic BERCs model, and are shown in Table 6.4. The parameters  $R_b$  and  $R_m$  were fitted as exclusively in Facsimile, as discussed in the previous section.

To allow  $k_{d1}$  and  $k_{d2}$  to vary between association phases, Facsimile set them as  $k_{d11}$  and  $k_{d21}$  in the association phase, and  $k_{d12}$  and  $k_{d22}$  in the dissociation phase. The four new parameters initially took the values Differential Evolution sent to Facsimile for  $k_{d1}$  and  $k_{d2}$ , and were varied internally.

Parameters obtained by fits of the BERCs extended model are shown in Table 6.8(a).

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	1.9E+7	5.2E+6	4.7E+6	4.7E+6	6.2E+6	4.5E+6	6.9E+6	7.2E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	8.2E+0	3.2E-1	8.0E+1*	5.1E+2	3.4E-13*	8.4E-12*	5.7E+0	1.0E+0
$k_{d11}(s^{-1})$	5.7E+1	6.5E-1	7.1E-4	2.3E-3	7.3E-4	1.6E-3	1.5E-3	1.5E-3
$k_{d21}(s^{-1})$	1.0E-12*	1.0E-07*	1.6E+0	4.7E-1	3.6E-7*	1.1E-5*	1.2E-2	6.2E-3
$k_{d12}(s^{-1})$	2.1E-1	1.8E-02	1.6E-4	2.3E-3	7.8E-5	1.1E-4	1.7E-3	1.4E-3
$k_{d22}(s^{-1})$	4.1E-4	1.7E-04	2.6E-9*	5.3E-4*	3.6E-3*	7.6E-8*	6.1E-11*	1.2E-6*
$R_1(pgnm^{-2})$	4.5E+3	2.0E+3	1.0E+3	5.1E+2	8.0E+2	8.8E+2	3.5E+2	3.8E+2
$R_2(pgnm^{-2})$	3.1E+3	2.1E+3	9.3E+2	5.1E+2	7.7E+2	9.4E+2	3.8E+2	3.6E+2
$R_3(pgnm^{-2})$	4.8E+3	2.3E+3	1.0E+3	4.6E+2	7.9E+2	1.2E+3	4.0E+2	3.4E+2
$R_4(pgnm^{-2})$	1.7E+3	5.5E+3	1.0E+3	4.7E+2	9.4E+3	3.5E+2	3.6E+2	3.5E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	1.5E-2	1.2E-2	7.9E+2*	2.1E+1*	8.0E+7*	7.8E-1*	8.3E-3	1.9E+1*
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	3.5E-2	2.3E+1*	8.3E+0*	7.5E+0*	1.5E+0*	5.7E-2	3.6E-3	3.8E+1*
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	4.6E-2	1.5E-1	1.8E+3*	2.3E+1*	1.6E+1*	4.5E-3	1.9E-2	3.3E-1*
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	2.4E+8*	4.3E+0*	9.8E+1*	1.8E-2	1.7E+13*	1.4E-3	1.7E-1*	4.7E+0*
$T_{1b}$	5.8E-1	1.0E-1	3.4E-1	1.0E-2	1.0E+0	1.0E+0*	6.1E-2*	3.0E-7*
$T_{2b}$	4.4E-1	1.2E-1	4.4E-1	4.3E-2	1.0E+0	1.0E+0*	6.2E-2*	5.6E-2
$T_{3b}$	6.3E-1	3.2E-1	3.2E-1	7.6E-2	2.4E-2	9.9E-1*	7.4E-2*	1.4E-1
$T_{4b}$	3.4E-4*	7.0E-1	2.2E-1	7.9E-2	1.0E+0	1.0E+0*	2.0E-1*	1.8E-1

(a) BERCs extended model.

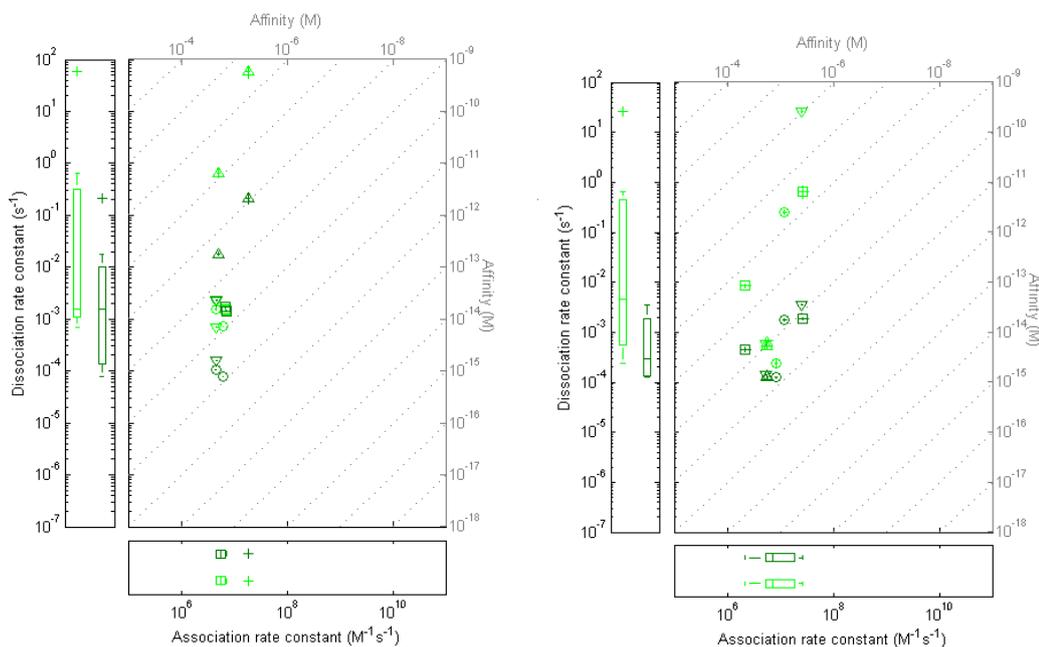
	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	5.6E+6	5.6E+6	5.3E+6	2.6E+7	1.2E+7	8.5E+6	2.7E+7	2.2E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	9.3E+0	5.7E+0*	5.8E+1	1.6E+1	3.6E-2	7.4E-2	2.1E-1	5.7E-5
$k_{d1}(s^{-1})$	5.4E-4	6.1E-4	5.8E-4	2.7E+1	2.5E-1	2.4E-4	6.4E-1	8.5E-3
$k_{d2}(s^{-1})$	3.3E-1	2.0E-1	1.6E+0	4.7E-5	2.5E-4	1.0E-2	4.8E-4	1.8E-4
$k_{d11}(s^{-1})$	5.4E-4	6.1E-4	5.8E-4	2.6E+1	2.4E-1	2.4E-4	0.6E-1	8.5E-3
$k_{d12}(s^{-1})$	1.3E-4	1.4E-4	1.4E-4	3.5E-4	1.8E-3	1.2E-4	1.9E-3	4.4E-4
$k_{d21}(s^{-1})$	1.9E-1*	1.6E+1*	4.7E-5	2.5E-4	1.0E-2	4.8E-4	1.7E-4	1.5E-3*
$k_{d22}(s^{-1})$	4.1E-4*	1.0E-7*	1.6E+0*	4.7E-1*	3.6E-7	1.1E-5*	1.2E-2*	6.2E-3*
$R_1(pgnm^{-2})$	1.0E+3	9.9E+2	1.0E+3	9.4E+2	1.9E+3	7.7E+2	5.8E+2	9.1E+2
$R_2(pgnm^{-2})$	9.4E+2	9.8E+2	9.4E+2	7.5E+2	2.0E+3	6.9E+2	6.5E+2	1.4E+3
$R_3(pgnm^{-2})$	9.9E+2	8.9E+2	1.0E+3	6.7E+2	2.0E+3	7.2E+2	7.2E+2	2.3E+3
$R_4(pgnm^{-2})$	9.6E+2	1.0E+3	9.9E+2	6.9E+2	2.1E+3	6.0E+2	9.3E+2	4.1E+3
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	2.2E+4*	4.2E-1*	3.5E+1*	4.3E-3	8.6E-3	1.2E-2	1.5E-2*	3.5E-2*
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	2.7E+2*	7.6E+0*	4.0E+2*	1.0E-2	5.1E-1*	3.9E-2*	6.9E-3	1.2E-3
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	7.9E-1	4.7E-1	5.1E+2*	1.5E-2	5.2E-1*	5.9E-3	2.7E-3	6.4E-4
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	1.1E+1*	5.3E-1	1.0E+2*	9.1E-3	1.1E-2	7.9E-2*	9.3E-4	5.0E-4
$T_{1b}$	1.5E-1	2.4E-1	2.3E-1	3.2E-1	1.3E-4*	3.2E-2	7.4E-2	1.2E-2*
$T_{2b}$	2.0E-1	2.3E-1	2.8E-1	1.3E-1	1.9E-2	1.8E-1	7.0E-2	3.0E-2*
$T_{3b}$	2.1E-1	1.4E-1	2.5E-1	1.2E-1	2.6E-2	1.3E-2*	5.1E-2	6.9E-3*
$T_{4b}$	1.8E-1	1.6E-1	2.1E-1	1.6E-1	4.0E-2	2.8E-1	1.9E-1	7.2E-2*
$b_{11}(pgnm^{-2})$	6.6E+1	7.5E+1*	3.5E+2	2.2E-1*	7.6E+2	5.1E+2	1.3E+2	8.0E+1
$b_{21}(pgnm^{-2})$	2.0E+1	2.4E+1*	2.9E+2*	2.0E+2	3.1E+2	1.6E+2*	1.3E+2*	6.6E+1
$b_{31}(pgnm^{-2})$	1.0E+2*	5.3E-1*	2.9E+2*	2.8E+2	1.8E+2	1.3E+1*	2.9E+1*	1.0E+2
$b_{41}(pgnm^{-2})$	1.8E+2*	6.4E+2	2.8E+2	1.5E+2	2.9E+2	1.8E+2	1.3E+1*	8.62E+1
$b_{12}(pgnm^{-2})$	7.6E+2	6.1E+2	3.4E+2	1.3E+0*	4.0E+0*	1.6E+2*	3.4E+1*	1.E+1
$b_{22}(pgnm^{-2})$	6.5E+2	6.5E+2	2.9E+2	5.2E+0*	3.2E+2	3.1E+2	3.9E+1*	1.0E+2
$b_{32}(pgnm^{-2})$	6.4E+2	6.3E+2	3.0E+2	2.3E-1*	1.6E+1*	1.8E+2*	1.5E+2	6.1E+1
$b_{42}(pgnm^{-2})$	6.0E+2	1.2E+1*	2.8E+2*	9.2E+0*	8.0E+0	1.1E+1*	6.7E+1	7.0E+0*

(b) BERCs extended 2 model.

Table 6.8: Parameter estimates.

Of 144 parameters 43 were NWD, part of the motivation for the development of the BERCs extended 2 model was obtaining a higher proportion of well determined parameter estimates.

Figure 6.17(a) shows the parameter estimates made by this models. As the models fitted the association rate constant across both phases, and the dissociation constant separately in either phase, each light green symbol is vertically above a dark green symbol and the



(a) BERCSs extended model. (b) BERCSs extended 2 model.

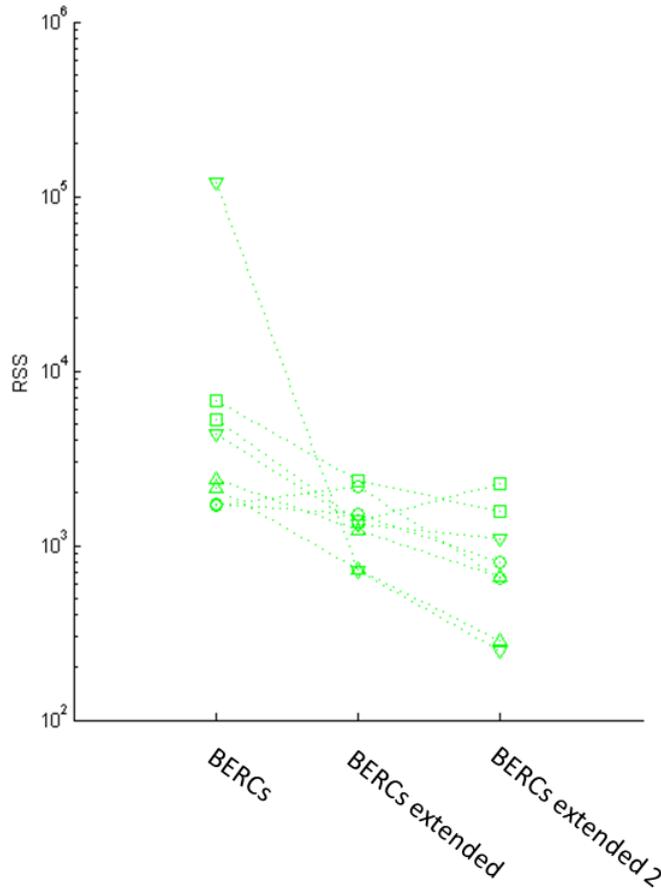
$\triangle (k_{a1}, k_{d11})$ for HDHC SN230G6 vs HLA-A2	$\triangle (k_{a1}, k_{d12})$ for HDHC SN230G6 vs HLA-A2
$\nabla (k_{a1}, k_{d11})$ for LDHC SN230G6 vs HLA-A2	$\nabla (k_{a1}, k_{d12})$ for LDHC SN230G6 vs HLA-A2
$\circ (k_{a1}, k_{d11})$ for HDHC SN230G6 vs HLA-A2	$\circ (k_{a1}, k_{d12})$ for HDHC SN230G6 vs HLA-A2
$\square (k_{a1}, k_{d11})$ for HDHC SN230G6 vs HLA-A2	$\square (k_{a1}, k_{d12})$ for HDHC SN230G6 vs HLA-A2

**Figure 6.17:** Scatter graphs of estimates of association and dissociation rate constants, with confidence intervals, boxplots and secondary axis representing affinity.

box plots for the association rate constants at the base of the graph are identical.

These show  $k_{d12}$  was estimated at a lower value for every data set than  $k_{d11}$ , and the boxplots show less variation for either  $k_{d12}$  or  $k_{d11}$  than the equivalent parameter in the BERCS model (see Fig. 6.14).

The Facsimile program created for the BERCS extended 2 model, like the BERCS extended model also allowed  $k_{d1}$  and  $k_{d2}$  to change between the two phases, and used the same mechanisms for this. It also fit the parameters  $b_1$  and  $b_2$  which were the initial conditions of  $x_1$  and  $x_2$  at the start of the dissociation phase. These were also varied internally, and



**Figure 6.18:** RSS values for the BERCs and BERCs extended models.

initial assigned 1 as a value.

Parameters obtained by fits of the BERCs extended 2 model are shown in Table 6.8(b), and their rate constant estimates are shown in Fig. 6.17(b).

Similar to the BERCs extended model, for the BERCs extended 2 model, show  $k_{d12}$  was estimated at a lower value for every data set than  $k_{d11}$ . Additionally the boxplots show less variation in estimates of either  $k_{d12}$  or  $k_{d11}$  than the equivalent parameter in the BERCs model (see Fig. 6.14), or in the equivalent parameter in the BERCs extended model (see Fig. 6.17(b)).

A comparison of the RSSs obtained for these two new models with the RSSs of the original BERCs model is shown in Fig. 6.18. RSSs are reduced when the dissociation constants are allowed to vary between phases and again when the state variables are allowed to discontinuously change values at the start of the dissociation phase.

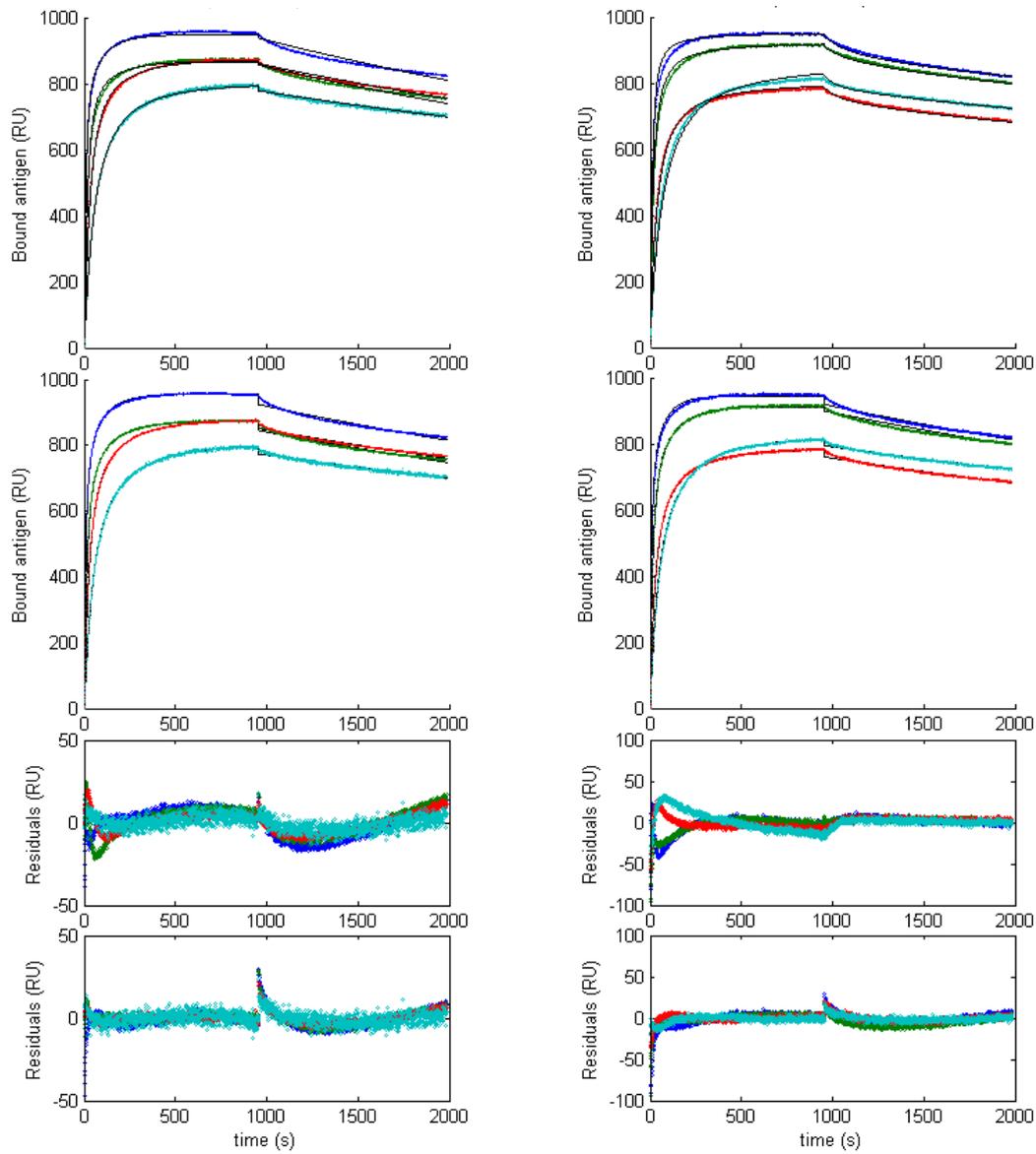
Fits to the data made with the new models to the data shown in Fig. 6.16 are shown in Fig. 6.19(a) and 6.19(b). In both cases the models reduce the residuals in the start of the dissociation phase, but do not eliminate them.

The extended BERCs models resulted in improvements in fits (See Fig. 6.18) and distribution of dissociation rate constant estimates (See Fig. 6.17(a) and Fig. 6.17(b)). However when more experiments were conducted on the same interaction (see Fig. 4.9), and the BERCs simplified model was fitted to them their fits had none of the systematic errors that existed in the sensogram fits that motivated these models (see Section 6.6). As a result these systematic errors were judged to be caused by artefacts, and no more modeling was done with the data taken from the original HLA-A2 experiments.

#### **6.4.5 Initial fitting the simplified BERCs model**

In Section 5.1.6 it was noted that the number of parameters could be reduced by equating the primary and secondary dissociation constants. This would mean that the model assumed that no conformational changes would happen when the antibody-antigen dimer trimerises, that would cause it to have a slower rate constant for splitting into a dimer and an antigen than the dimer would for becoming a free antibody and a single antigen.

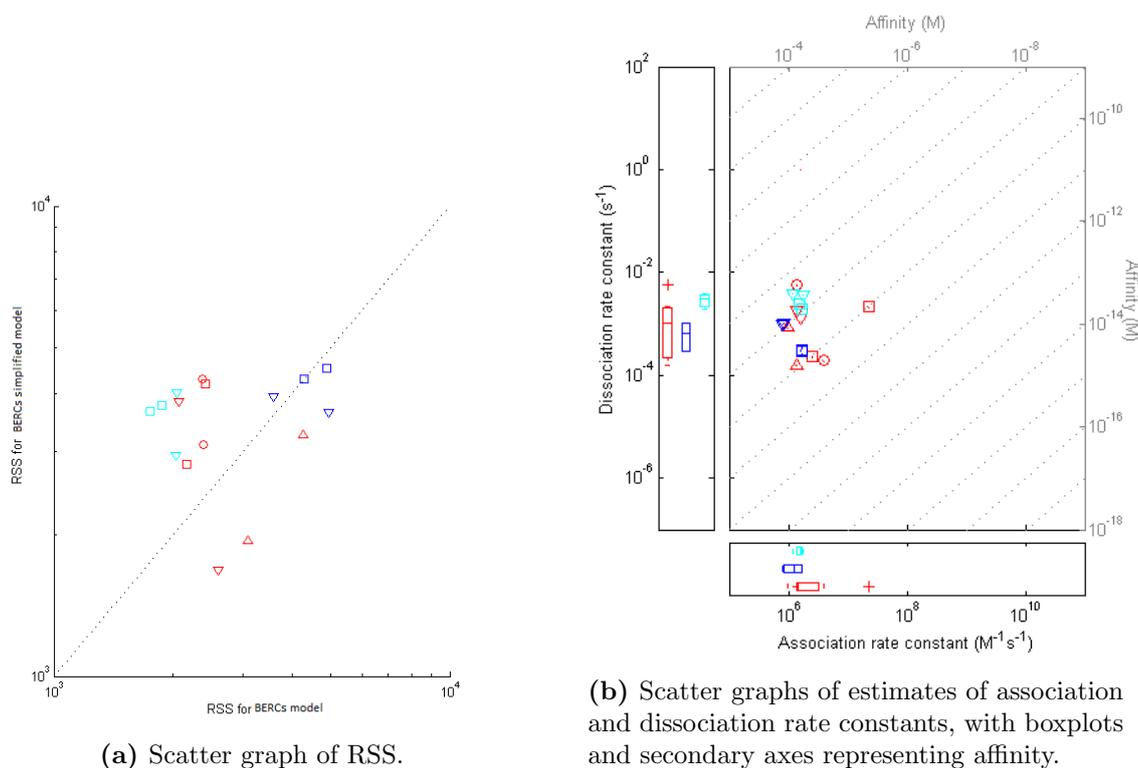
Typically good fits of the BERCs model had estimated values of the primary and secondary dissociation rate constants within one order of magnitude of each other (see Table 8.3), so were in a much smaller area of the parameter space than the area searched. As a result it was thought that this simplification may allow for faster processing of data, so it was used



(a) Comparison of fits between the two extended BERCs models.

(b) Comparison of fits between the two extended BERCs models.

**Figure 6.19:** Illustrations of the fits of the extended BERCs model, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.3.



**Figure 6.20:** Graphs showing the RSS and affinity estimates made by the simplified BERCs model. See Figure 6.4 for legend.

on the SN607D8 and WK1D12 experiments. The data from the SN230G6 experiment was not processed in this way, as it appeared to have complexities that limited the use of the BERCs model for processing.

The BERCs simplified model was fitted by a combination of Differential Evolution and Facsimile programs nearly identical to the BERCs model, except that the parameter  $k_d2$  was not read in by Facsimile (see Subsection 6.4.3). An initial fitting run was conducted for 6 hours per model per sensorgram dataset, 1/4 of the time used to fit the previous models.

In Fig. 6.20(a) we see that the range of RSS achieved by either model is similar, but that it was improved by the simplified model in six out of sixteen cases, but was made worse in

	HdHc	HdHc	LdHc	LdHc	HdLc	HdLc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	9.6E+5	1.4E+6	1.4E+6	1.6E+6	1.4E+6	3.9E+6	2.5E+6	2.2E+7
$k_{a2}(nm^2pg^{-1}s^{-1})$	1.6E-1	5.1E+1	6.5E-1	1.3E+3	1.0E+1	4.8E-1	1.8E+4	5.9E-1
$k_{d1}(s^{-1})$	8.5E-4	1.6E-4	1.9E-3	1.3E-3	5.7E-3	2.0E-4	2.4E-4	2.1E-3
$R_1(pgnm^{-2})$	1.0E+3	5.9E+2	5.0E+2	4.4E+2	8.0E+2	5.8E+2	2.7E+2	4.6E+2
$R_2(pgnm^{-2})$	1.0E+3	5.1E+2	5.0E+2	3.8E+2	8.6E+2	5.7E+2	2.8E+2	4.3E+2
$R_3(pgnm^{-2})$	1.0E+3	5.2E+2	5.1E+2	4.4E+2	1.0E+3	5.7E+2	2.6E+2	4.6E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	7.9E-3	1.1E-2	1.4E-2	4.4E-1	1.2E+3	3.5E-4	3.9E-4	7.5E-3
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	1.3E-2	1.3E+0	4.9E-2	4.4E-1	1.3E+1	6.3E-4	6.0E-4	5.2E-3
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	7.2E-1	2.4E-1	2.3E-1	4.4E-1	2.3E-2	1.7E+3	3.7E-2	1.4E-1
$S_{1m}$	9.6E-1	9.3E-1	4.1E-1	1.4E-1	8.5E-1	9.8E-1	7.7E-1	5.6E-1
$S_{2m}$	3.2E-1	3.8E-2	7.1E-1	2.3E-1	1.4E-1	2.1E-1	2.2E-1	4.3E-1
$S_{3m}$	2.2E-1	1.4E-1	3.5E-1	9.9E-1	1.7E-1	7.0E-1	1.8E-1	6.6E-1

(a) Sn607 against HLA-A2.

	B7				B27			
	LdHc	LdHc	LdLc	LdLc	LdHc	LdHc	LdLc	LdLc
$k_{a1}(M^{-1}s^{-1})$	8.2E+5	8.4E+5	1.7E+6	1.7E+6	1.2E+6	1.8E+6	1.7E+6	1.5E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	1.7E-1	1.6E-1	2.1E-1	3.3E-1	5.2E-1	3.1E-1	4.3E-1	5.9E-1
$k_{d1}(s^{-1})$	9.9E-4	1.0E-3	3.0E-4	3.1E-4	3.8E-3	3.7E-3	1.9E-3	2.4E-3
$R_1(pgnm^{-2})$	7.5E+2	7.5E+2	4.9E+2	4.9E+2	3.2E+2	3.0E+2	2.5E+2	2.6E+2
$R_2(pgnm^{-2})$	6.7E+2	6.7E+2	4.2E+2	4.2E+2	3.4E+2	3.2E+2	2.3E+2	2.4E+2
$R_3(pgnm^{-2})$	6.7E+2	6.7E+2	3.8E+2	3.8E+2	3.5E+2	3.3E+2	2.3E+2	2.4E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	2.1E-3	2.1E-3	3.1E-4	3.3E-4	1.8E-3	8.0E-4	5.2E-4	7.4E-4
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	5.8E-3	5.7E-3	6.3E-4	6.7E-4	2.7E-3	1.3E-3	1.2E-3	3.0E-3
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	8.1E+0	3.5E-1	1.5E-2	1.3E-2	1.0E+2	1.8E-3	2.7E-3	9.8E-3
$S_{1m}$	3.0E-1	2.0E-2	8.5E-1	8.8E-2	5.2E-1	3.6E-1	5.9E-1	2.8E-1
$S_{2m}$	4.8E-1	3.9E-1	6.4E-1	8.8E-2	6.4E-1	6.0E-1	2.4E-1	4.9E-2
$S_{3m}$	4.9E-1	4.4E-1	1.1E-1	6.9E-1	1.9E-2	2.4E-1	1.4E-1	1.3E-1

(b) Wk1d12 reacted against multiple antigen.

**Table 6.9:** Parameter estimates made with the Bivalent ERC spatial model.

ten cases. Notably whenever data the WK1D12 vs B27 was processed it was made worse by the simplified model.

However, as the simplified model was run for a dramatically shorter period of time, it is likely that better parameter estimates could be found. The fact that the RSSs remained within the same approximate range, and in some cases were improved was seen as sufficient reason to believe that it could produce similar quality fits faster.

Graphs showing the output of this model along with data are not shown because they are characteristically similar to those produced by the BERCs model shown in Fig. 6.13.

The estimates of parameter values are shown in Fig. 6.20(b). Notably these are very similar to those produced by the BERC model shown in Fig. 6.10(a), except that generally

there is a reduction in the range of the estimates of a parameter for a specific interaction. All parameters for these fits are shown in Table 6.9, unlike previous parameter tables, this does not use an asterisk to mark NWD parameter estimates, as these were initial estimates using the model, and the model was used on these data and more in the next section.

## 6.5 Processing expanded datasets with the simplified BERCs model

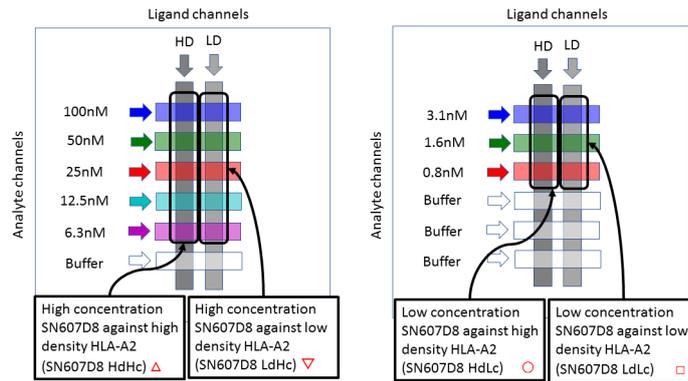
The previous subsection showed fits with the BERCs simplified model made in dramatically reduced time with consistent rate constant estimates, RSSs better than most previous models, and with a similar distribution to the best-fitting previous model (BERCs).

This motivated using the model on larger datasets. Figure 6.3 shows how data had been previously divided - two interaction spots per ligand density were not modelled in the high concentration SN607G8 vs HLA-A2 experiments, and the WK1D12 vs multiple antigen experiments had been divided into two sets of three interactions per dataset.

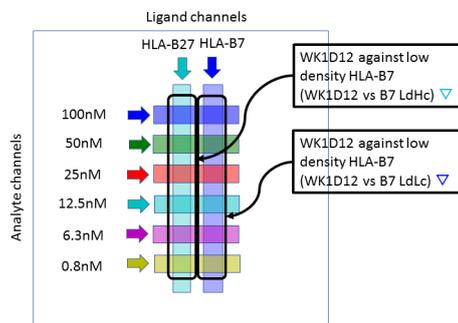
It was wondered if the BERCs model could be used to process larger datasets. As a result the experimental data was redivided so that all sensorgrams produced from a single chip at with a single ligand density were included in the same dataset, as shown in Fig. 6.21.

These datasets were processed identically to those in the previous section, except that fitting was performed for each sensorgram for 24 hours rather than 6.

Sensorgrams produced by these fits are shown in Fig. 6.22. In Fig. 6.22 (a) and (b) we see the model lines (black) remain within the coloured datapoints they are fitted to throughout the duration of both experiments on sensorgram curves. The residual plots



(a) Data sets taken from the SN607D8 against HLA-A2 sensorgrams.



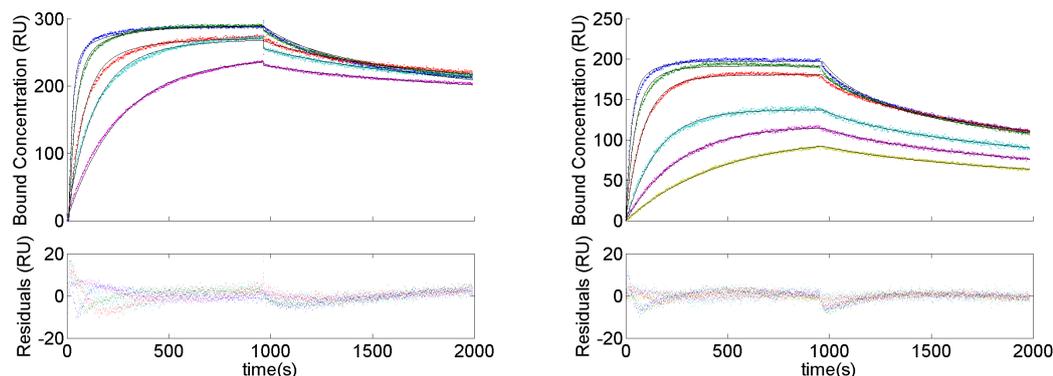
(b) Data sets taken from the WK1D12 against multiple antigen sensorgrams.

**Figure 6.21:** Diagrams of the interaction spots where the sensorgram curves in each data set comes from, with labels showing abbreviated names of the dataset, and colour coded symbols that will be used to represent the datasets through this subsection.

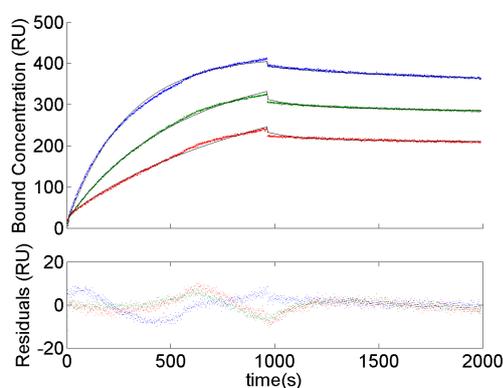
in these subfigures show that the weakest parts of these fits are the early parts of either phase, although residuals are not as large as fits done with models from the literature (compare Fig. 6.9 and Fig. 6.5) even though more data is processed.

Fig. 6.22 (c) shows a fit to a dataset that contains the artefact identified as humps (see Subsection 4.3.1). Other than the humps in the association phase, the model fits well, but generally has higher residuals than the two previous sensorgrams.

The general improvements in fit are also reflected in the scaled RSS values of the model,



(a) High concentration SN607D8 vs low density HLA-A2 antigen. (b) All concentrations of WK1D12 vs HLA-B27 antigen.

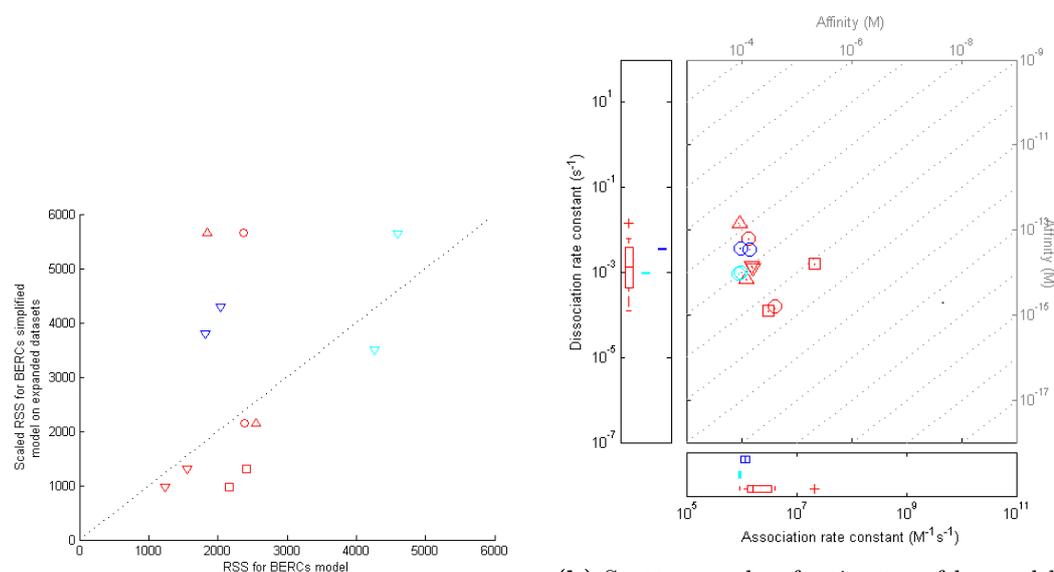


(c) Low concentration SN607D8 vs high density HLA-A2 antigen.

**Figure 6.22:** Fits of the BERCs simplified model with residuals, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.21.

which are shown in a scatter graph against those for the previous best performing model, Fig. 6.23(a). In this graph, RSS is scaled because additional data were included in these fits.

For the BERCs simplified model in the high concentration SN607 experiments is multiplied by  $3/5$ , because the simplified BERCs model had  $5/3$  times as many datapoints as the BERCs model. The RSS for the simplified BERCs model in the WK1D12 experiments is multiplied by  $1/2$  because the simplified BERCs model had 2 times as many data points as either of the BERCs models processing the same dataset (compare Fig. 6.3 and 6.21), and



(a) Scatter graph of scaled RSS for the BERCs simplified and BERCs models. (b) Scatter graphs of estimates of  $k_{a1}$  and  $k_{d1}$ , with boxplots and secondary axes representing affinity. See Figure 6.4 for legend.

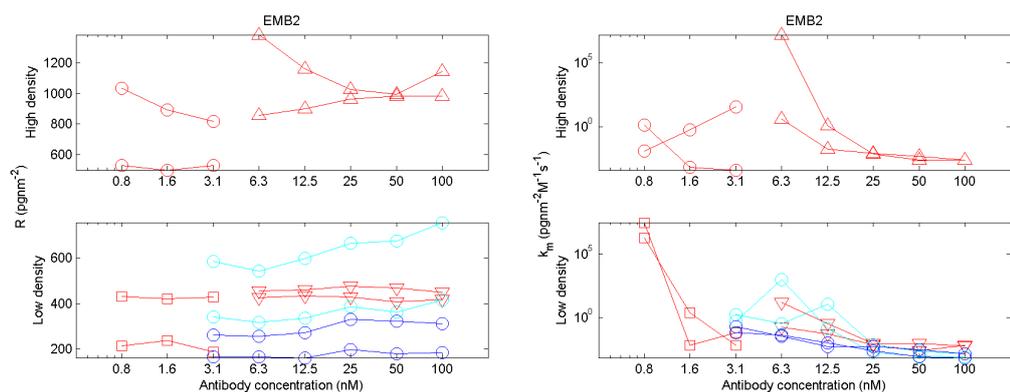
**Figure 6.23:** Graphs of RSS and rate constants for the BERCs simplified model.

it is plotted against the average RSS for the BERCs model for those merged datasets.

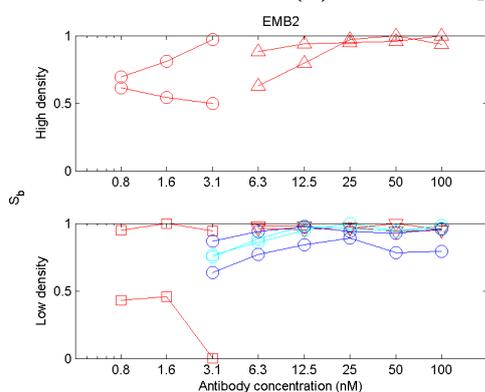
This figure shows a general improvement in fits. With two high-density ligand SN607D8 vs A2 experiments showing a reduction in quality of fit, and three WK1D12 experiments showing a reduction in quality of fit.

Parameter estimates from these fits are shown in Table 6.10 and graphed in Fig. 6.23(b) and Fig. 6.24. Comparatively few of these parameter estimates are marked as NWD, 10/84 for the WK1D12 data-set, 14/120 for the SN607D8 data-set. The parameter values marked as NWD are most commonly the  $k_{ds}$  for the lanes with the lowest concentration of antibody (and correspondingly the highest number as a subscript). Generally these also have the highest values for their  $k_m$  estimates (compare Fig. ).

This makes sense, as these lanes have much lower levels of reactant so are less likely to be limited by mass transport as observed by Richard B. M. Schasfoort (2008) and Marquart (2013).



(a) BERCs simplified estimates for  $R$ . (b) BERCs simplified estimates for  $k_m$ .



(c) BERCs simplified estimates for  $S_b$ .

**Figure 6.24:** Estimates of  $k_m$  and  $R$  for the L2 and ERC2 models. See Figure 6.4 for legend

The rate constant estimates are shown in Fig. 6.23(b) have a very similar distribution to those made with the BERCs model (see Fig. 6.14) and are very consistent. For all points representing WK1D12 experiments, their closest neighbour is a point from a replicate of the same experiment, and estimates of both  $k_{d1}$  and  $k_{a1}$  vary by much less than an order of magnitude. There is a greater spread of points representing the SN607D8 against HLA-A2 experiment (across two orders of magnitude for both  $k_{a1}$  and  $k_{d1}$ ), but again all points have as their closest neighbour a point from a version of the same experiment, although in many cases with a different analyte concentration or ligand density.

The accuracy and consistency of these estimates is in contrast to estimates from benchmark

$k_{a1}(M^{-1}s^{-1})$	1.3E-3	9.1E-4	1.5E-3	1.6E-3	8.4E-5	2.5E-4	1.9E-4	1.3E-3
$k_{a2}(nm^2pg^{-1}s^{-1})$	2.1E-1	8.6E+1	7.8E-1	7.9E-1	1.9E+1	5.7E-1*	2.4E+3*	1.1E+0
$k_{d1}(s^{-1})$	7.0E-4	1.4E-2	1.4E-3	1.3E-3	6.3E-3	1.6E-4	1.2E-4	1.6E-3
$R_1(\text{pgnm}^{-2})$	9.8E+2	1.1E+3	4.5E+2	4.2E+2	8.2E+2	5.3E+2	1.9E+2	4.3E+2
$R_2(\text{pgnm}^{-2})$	9.8E+2	9.9E+2	4.7E+2	4.1E+2	8.9E+2	5.0E+2	2.4E+2	4.2E+2
$R_3(\text{pgnm}^{-2})$	9.6E+2	1.0E+3	4.8E+2	4.3E+2	1.0E+3	5.3E+2	2.1E+2	4.3E+2
$R_4(\text{pgnm}^{-2})$	9.0E+2	1.2E+3	4.6E+2	4.3E+2				
$R_5(\text{pgnm}^{-2})$	8.6E+2	1.4E+3	4.6E+2	4.3E+2				
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.4E-3	2.3E-3	6.1E-3	6.8E-3	3.5E+1*	4.0E-4	6.4E-3	6.5E-2
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.7E-3	2.3E-3	9.6E-3	2.4E-3	5.6E-1*	6.5E-4	2.3E+0*	6.6E-3
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	8.1E-3	7.4E-3	9.1E-3	7.6E-3	1.3E-2*	1.3E+0*	1.9E+6*	2.9E+7*
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.7E-2	1.2E+0*	3.3E-1*	5.3E-2				
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.0E+0*	1.7E+7*	1.5E+1*	1.8E-1*				
$T_{1b}$	5.8E-4*	6.8E-2	5.2E-2	4.9E-2	3.2E-2	1.0E+0	2.1E+3*	6.0E-2
$T_{2b}$	4.3E-2	9.4E-7*	6.3E-2	8.3E-4*	2.3E-1	8.4E-1	1.2E+0	2.0E-6*
$T_{3b}$	5.5E-2	2.9E-2	3.0E-2	4.0E-2	4.4E-1	6.3E-1	1.3E+0	5.3E-2
$T_{4b}$	6.6E-2	2.6E-1	5.1E-2	1.9E-2				
$T_{5b}$	1.4E-1	5.9E-1	3.8E-2	1.9E-2				

(a) Sn607 against HLA-A2.

	HLA-B27		HLA-B7	
$k_{a1}(M^{-1}s^{-1})$	9.8E+5	9.1E+5	1.4E+6	9.8E+5
$k_{a2}(nm^2pg^{-1}s^{-1})$	4.1E-1	3.5E-1	8.4E-1	2.4E+0
$k_{d1}(s^{-1})$	1.0E-3	9.5E-4	3.4E-3	3.7E-3
$R_1(\text{pgnm}^{-2})$	7.5E+2	4.2E+2	3.1E+2	1.9E+2
$R_2(\text{pgnm}^{-2})$	6.7E+2	3.6E+2	3.2E+2	1.8E+2
$R_3(\text{pgnm}^{-2})$	6.6E+2	3.9E+2	3.3E+2	2.0E+2
$R_4(\text{pgnm}^{-2})$	6.0E+2	3.4E+2	2.7E+2	1.6E+2
$R_5(\text{pgnm}^{-2})$	5.4E+2	3.2E+2	2.6E+2	1.7E+2
$R_6(\text{pgnm}^{-2})$	5.8E+2	3.4E+2	2.6E+2	1.7E+2
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.4E-3	8.4E-4	1.4E-3	7.1E-4
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.3E-3	9.8E-4	3.2E-3	8.1E-4
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	7.2E-3	1.8E-3	5.1E-3	2.6E-3
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.1E+1*	1.4E-1*	5.2E-3	1.0E-2
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.3E-1*	9.9E+2*	3.5E-2	4.3E-2*
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.7E+0*	5.1E-1*	1.9E-1*	6.4E-2*
$T_{1b}$	1.6E-2	3.8E-2	4.5E-2	2.7E-1
$T_{2b}$	6.1E-2	5.7E-2	7.7E-2	2.7E-1
$T_{3b}$	4.0E-2	4.7E-5*	6.3E-2	1.2E-1
$T_{4b}$	2.3E-2	5.2E-2	2.6E-2	1.9E-1
$T_{5b}$	1.3E-1	1.6E-1	6.4E-2	3.0E-1
$T_{6b}$	3.3E-1	3.1E-1	1.5E-1	5.7E-1

(b) WK1D12 against multiple antigen.

**Table 6.10:** Parameter estimates made with the BERCs simplified model.

studies in the literature (see Fig. 1.15) where estimates of either constant have varied across multiple orders of magnitude.

Interestingly the immunising antigen for the WK1D12 antibody was B27 (see Table 4.2), although B7 presents the same epitope, the antibody's affinity for the antigen is lower. This decrease in affinity appears to come from an increased dissociation rate constant.

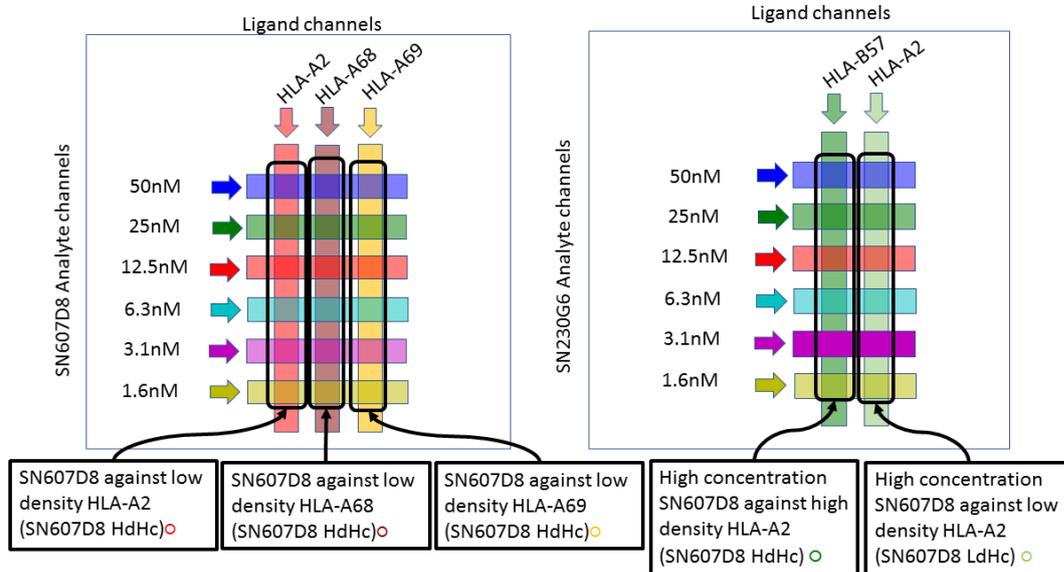
This greater variation in rate constant estimates for the SN607D8 against HLA-A2 experiments may be due to the artefacts in most of the sensorgrams discussed in Subsection 4.3.1. Conversely the low variation in rate constant estimates for the WK1D12 may be due to the fact that the sensorgrams had much fewer artefacts (see Subsection 4.3.3).

## 6.6 Processing new data with the BERCs model

Whilst the modelling in the previous section was being carried out, a new dataset became available (see Section 4.4). These experiments were carried out like the WK1D12 against multiple antigen experiments, which had much fewer artefacts, such as jumps in signal between phases, humps within phases. The only difference was that lower concentrations of antibody were used, the highest concentration used was 50nM, rather than 100nM. This is because the highest concentration lanes had generally been worst affected by the previously mentioned artefacts. Notably, this dataset still included crossovers, suggesting that these may be a feature of these experiments rather than an artefact.

Like the WK1D12 against multiple antigen experiments, this data featured each antibody reacted against multiple HLA which all present the same epitope. The goal of this linked experimental - theoretical project was twofold; to establish models that could estimate association rate and dissociation rate constants so accurately and consistently that the differences between these affinities could be measured, and to measure these affinities.

In part, the goals of this project had already been accomplished in the previous section. The simplified BERCs model gave such consistent estimates (see Fig. 6.23(a)) that the difference in the affinity of WK1D12 for HLA-B7 and HLA-B27 had already been measured extremely accurately, allowing for the observation that affinity was higher for the immunising antigen and that the difference was largely due to an increased dissociation



(a) Diagram of interaction spots on the XPR 36 chips for the SN607D8 antibody vs HLA-A2, HLA-A68 and HLA-A69 antigens experiment.

(b) Diagram of interaction spots on the XPR 36 chips for the SN230G6 antibody vs HLA-B57 and HLA-A2 antigens experiment.

**Figure 6.25:** Diagram of interaction spots on XPR 36 chips.

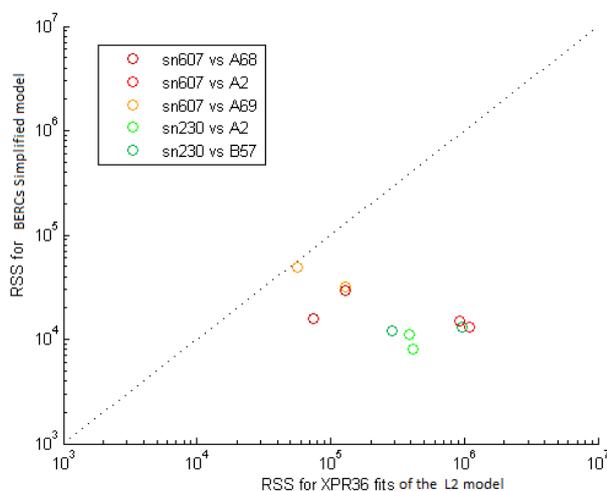
constant for the interaction with HLA-B7.

As a result, data was divided identically to the previous section (see Fig. 6.25), and processed with the simplified BERCs model, with an identical design for Differential Evolution and FACSIMILE programs.

This was because the faster speed it was able to achieve fits in, and ability to process more data simultaneously.

As a result, the new data were treated in some ways like a test set, in which the best model developed in the previous section could be validated.

The RSSs for the processing of the data are shown in Fig. 6.26 as well as the RSSs achieved by the XPR modelling of the same data. The model the XPR software fits to the



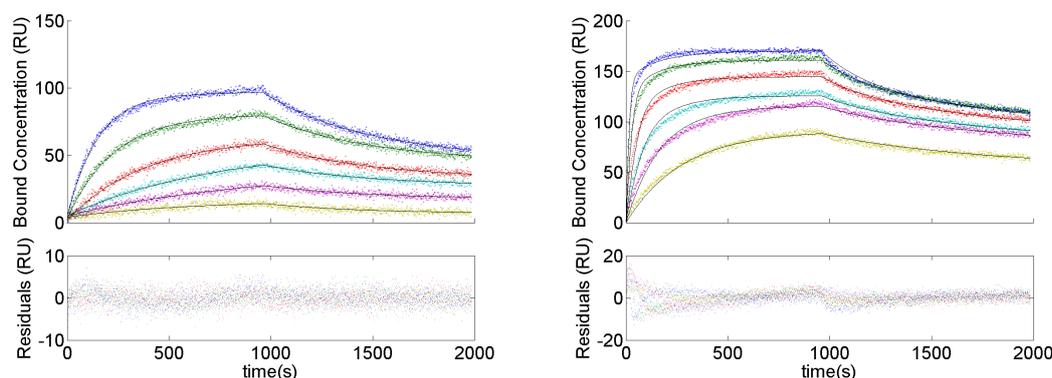
**Figure 6.26:** Scatter Graph of RSS for both models.

data is the same as the bivalent Langmuir model presented in Section 5.1.4 and fitted to data in Section 6.4.2. As a result, this figure shows the overall improvement of the fitting methods described in this chapter as well as the models developed in the previous chapter to those commercially available. In four cases the RSS is reduced by a factor of 10, in the remaining six cases the RSS is reduced by less than a factor of 100.

Fits of this model are shown in Fig. 6.27. In this figure we see the model lines (black) largely remain within the coloured datapoints they are fitted to throughout the duration of both experiments on sensorgram curves. The residual plots in these subfigures show that the weakest parts of these fits are the early parts of either phase.

These graphs are a strong contrast to the XPR fits to the same data (see Fig.4.8 and Fig. 4.9) in which the model lines (black) in some cases are outside of the data for individual sensorgram curves for the majority of an experiment.

The estimates for the parameters are shown in Table 6.6.1, and graphed in Fig. 6.28 and Fig. 6.29. For the SN230G6 experiments, very few parameters are NWD (7/84); for the SN607D8 experiments a larger number are (26/126). Notably, for the SN607D8 experi-



(a) High concentration SN607D8 vs low density HLA-A2 antigen.

(b) All concentrations of WK1D12 vs HLA-B27 antigen.

**Figure 6.27:** Fits of the BERCs simplified model with residuals, with colours corresponding to concentration of analyte in channels as shown in Fig. 6.25.

ments, 9 parameters are missing from the table. These are the spot specific parameters for the interaction spots whose outputs was omitted as noted in Subsection 4.4.

The accuracy of these estimates is also shown visually in Fig. 6.28, with confidence intervals around the estimates. In 6.28 (b) the domain in which the estimates are is enlarged, and the confidence intervals around the estimates are large enough to be seen. With previous models, the variability between estimates meant that although there were tight confidence intervals around most estimates, they were practically invisible due to the scale that had to be used in previous figures to show all estimates (See Fig. 6.6 and Fig. 6.10). This illustrates both the importance of using replicate experiments when estimating parameters rather than simply relying on statistics like SDLNs or confidence intervals on the parameters, and the success of this model, in creating estimates with low variability between replicates, small confidence intervals and few SDLNs large enough to mark parameters as NWD.

Because fewer data were available, and more data were processed together, there are fewer points on this graph than on previous graphs of the same type. However, we may now note that the affinity estimates are constrained to a much smaller area of the graph, and

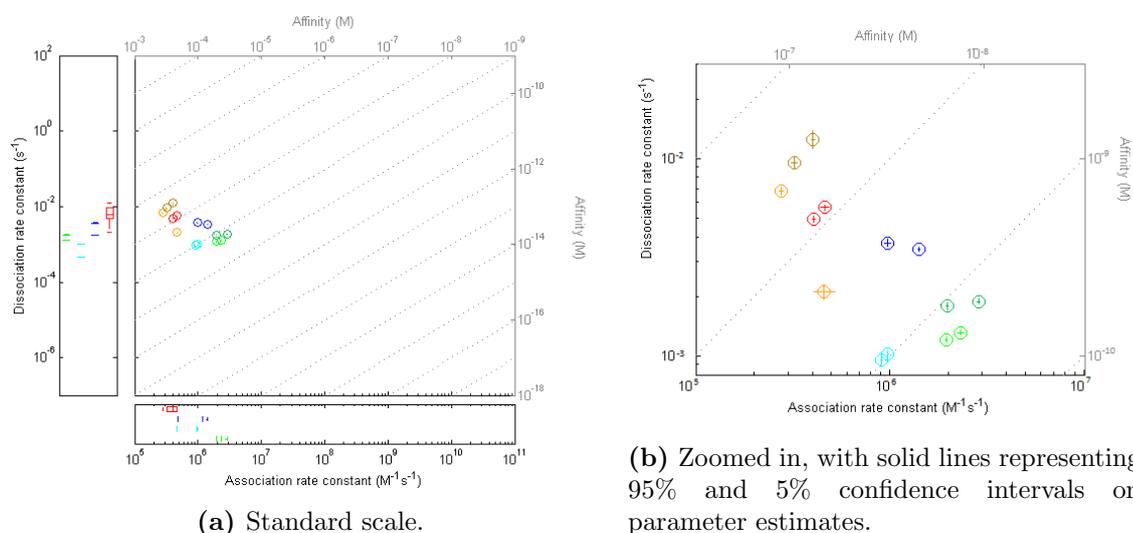
	HLA-A2		HLA-B57	
$k_{a1}(M^{-1}s^{-1})$	2.3E+6	2.0E+6	2.9E+6	2.0E+6
$k_{a2}(nm^2pg^{-1}s^{-1})$	3.6E-1	5.2E-1	8.6E-1	1.2E+0
$k_{d1}(s^{-1})$	1.3E-3	1.2E-3	1.9E-3	1.8E-3
$R_1(pgnm^{-2})$	3.3E-1	3.2E-1	2.4E-1	2.5E-1
$R_2(pgnm^{-2})$	3.4E+2	3.0E+2	2.5E+2	1.9E+2
$R_3(pgnm^{-2})$	3.2E+2	3.0E+2	2.4E+2	2.3E+2
$R_4(pgnm^{-2})$	3.0E+2	2.7E+2	2.2E+2	2.1E+2
$R_5(pgnm^{-2})$	2.9E+2	2.7E+2	2.2E+2	1.8E+2
$R_6(pgnm^{-2})$	2.6E+2	2.2E+2	1.8E+2	1.7E+2
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	1.8E-3	1.3E-3	1.4E-3	2.9E-3
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	1.8E-3	1.5E-3	1.7E-3	1.7E-2
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	2.4E-3	2.7E-3	1.8E-3	3.7E+1*
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	2.8E-3	3.5E-3	4.2E-3	2.6E+12*
$k_{m5}(pgnm^{-2}M^{-1}s^{-1})$	9.5E-3	2.0E-2	7.0E-3	1.6E+10*
$k_{m6}(pgnm^{-2}M^{-1}s^{-1})$	1.3E-1*	9.2E+0*	2.9E+3*	1.6E+17*
$T_{1b}$	1.75E-6*	5.13E-2	2.27E-2	4.92E-2
$T_{2b}$	1.21E-2	3.73E-2	2.24E-2	1.29E-1
$T_{3b}$	1.74E-10*	4.32E-2	3.59E-2	5.84E-2
$T_{4b}$	6.51E-2	7.36E-2	3.21E-2	6.85E-2
$T_{5b}$	5.89E-2	1.13E-1	4.07E-2	1.69E-1
$T_{6b}$	1.59E-1	1.94E-1	1.27E-1	5.15E-1

(a) SN230G6 against multiple antigen.

	A2		A68		A69	
$k_{a1}(M^{-1}s^{-1})$	4.6E+5	4.0E+5	3.2E+5	4.0E+5	2.8E+5	4.6E+5
$k_{a2}(nm^2pg^{-1}s^{-1})$	6.0E+0	3.8E+0	1.4E+1	2.2E+1	3.3E+0	1.8E+1*
$k_{d1}(s^{-1})$	5.6E-3	4.9E-3	9.5E-3	1.2E-2	6.8E-3	2.1E-3
$R_1(pgnm^{-2})$	5.3E+2	4.1E+2	4.5E+2	3.1E+2	3.5E+2	1.6E+2
$R_2(pgnm^{-2})$	4.0E+2	3.2E+2	5.2E+2	2.4E+2	4.4E+2	1.5E+2
$R_3(pgnm^{-2})$	5.5E+2	3.3E+2	6.1E+2	3.1E+2	4.7E+2	1.5E+2
$R_4(pgnm^{-2})$	5.1E+2	4.6E+2	7.0E+2	3.7E+2	5.8E+2	1.6E+2
$R_5(pgnm^{-2})$	8.7E+2*	4.3E+2	7.9E+2	4.4E+2	9.7E+2	1.7E+2
$R_6(pgnm^{-2})$		6.5E-1		4.9E-1		1.5E-1
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	5.8E-4	2.2E-3	7.5E-4	4.5E-4	2.7E+2*	2.9E-4*
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	1.6E-2*	3.9E-2*	1.4E+2*	7.9E-1*	8.6E+4*	1.5E-3*
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	1.0E-2*	1.1E-1*	1.6E+4*	3.1E+2*	6.8E+7*	2.0E-3
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	4.4E-3	1.0E+0*	1.0E+6*	9.5E+4*	5.5E+10*	2.7E-4
$k_{m5}(pgnm^{-2}M^{-1}s^{-1})$	1.9E-2*	3.1E+0*	6.4E+7*	1.3E+5*	4.4E+13*	2.9E-4
$k_{m6}(pgnm^{-2}M^{-1}s^{-1})$		2.0E+1*		1.9E+9*		1.9E-2*
$T_{1b}$	5.5E-2	9.3E-2	2.5E-1	2.5E-1	4.4E-2	7.0E-1
$T_{2b}$	2.7E-1	1.4E-1	5.7E-1	5.7E-2	6.1E-1	6.2E-1
$T_{3b}$	4.3E-1	3.5E-1	7.4E-1	3.7E-1	6.4E-1	9.6E-1
$T_{4b}$	9.3E-1	5.5E-1	9.0E-1	6.6E-1	1.2E+0	1.3E+0
$T_{5b}$	2.2E+0	8.2E-1	1.1E+0	1.0E+0	4.1E+0	1.8E+0
$T_{6b}$		2.0E+1		1.9E+9		1E+0

(b) SN607D8 reacted against multiple antigen.

Table 6.11: Bivalent ERC spatial.



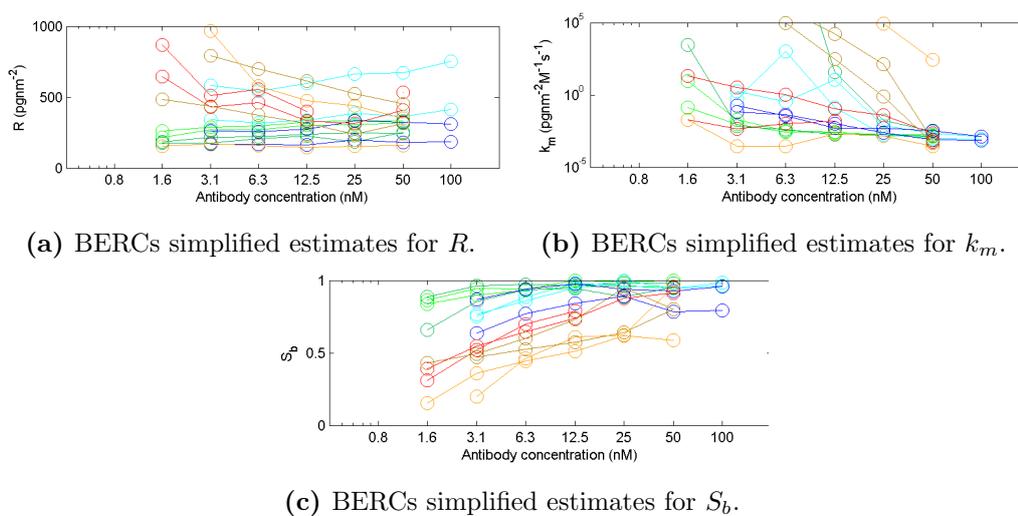
**Figure 6.28:** Estimates of primary and secondary association and dissociation rate constants made with the BERCS model. See Figure 6.26 for legend.

Human monoclonal	HLA protein	$k_a$	p value	$k_d$	p value	$K_D$	p value	$K_{D1}/K_{D2}$
SN230G6	A2	1.31E-03	<b>2.58E-05</b>	2.32E+06	<b>1.33E-04</b>	5.63E-10	0.244	0.75
		1.20E-03		1.96E+06		6.13E-10		
	B57	1.88E-03		2.87E+06		6.56E-10		
		1.78E-03		1.98E+06		9.02E-10		
WK1D12	B27	1.02E-03	<b>4.63E-05</b>	9.80E+05	<b>3.20E-04</b>	1.04E-09	<b>0.0407</b>	0.33
		9.47E-04		9.10E+05		1.04E-09		
	B7	3.45E-03		1.42E+06		2.43E-09		
		3.72E-03		9.77E+05		3.81E-09		
SN607D8	A2	5.63E-03	<b>2.58E-05</b>	4.63E+05	<b>1.33E-04</b>	1.22E-08	0.244	0.4
		4.90E-03		4.05E+05		1.21E-08		
	A68	9.45E-03		3.23E+05		2.93E-08		
		1.24E-02		4.03E+05		3.07E-08		

**Table 6.12:** Difference in rate constants and affinities for interaction with different alleles compared to immunising allele, with p values for significant differences between alleles.

are clustered so that generally, most estimates are closer to estimates from their replicates, than estimates from other interactions.

This clustering allows us to begin stratifying and comparing the rate constants and affinity of the interactions. Table 6.12 shows the estimated rate constants and affinity, as well as p-values for the significance of the difference of these estimates between alleles. These were calculated by using T-Tests on log-transformed parameter values. The A69 allele was



**Figure 6.29:** Estimates of  $k_m$  and  $R$  for the BERCs simplified model. See Fig. 6.26 for legend.

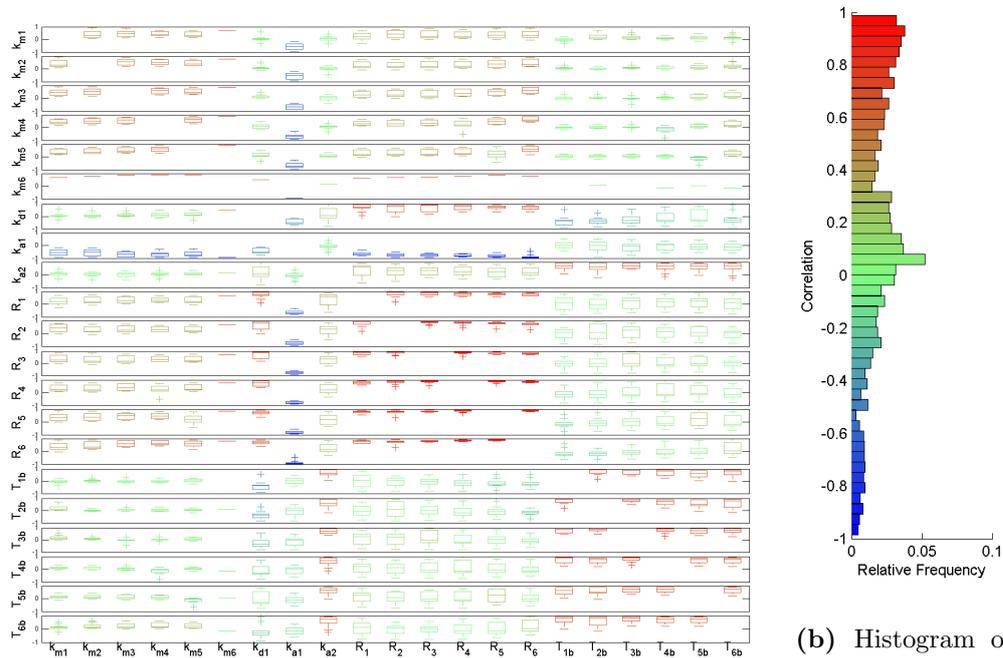
removed from these tests because unlike the other experiments the replicates had dissimilar values for their rate constants. As a result, this number represents the significance of the difference between its reaction with A2 and A68; rather than the significance of the differences between the group of three alleles.

The significance of this work is great, as it establishes the BERCs simplified model as a method for estimating antibody affinity accurate enough to distinguish between multiple configurations of the same epitope. This led to the paper (Daga *et al.*, 2017), in which these results were published (see Appendix).

### 6.6.1 Correlation of parameters

Whilst FACSIMILE fits the models it also analyses the correlation between the parameters used in fitting. At the end of the fitting process, it outputs a matrix showing the correlation coefficient for each pair of parameters.

Box-plots were made to visualise the distribution of the correlation coefficients for each



(a) Box-plot of parameter correlations.

(b) Histogram of parameter correlations.

**Figure 6.30:** Graphs of parameter covariances for the parameter estimates shown in Table 6.6.1.

pair of parameter estimates from Table . These box-plots are shown in Fig. 6.30 along with a histogram showing the frequency of correlation coefficients across all pairs.

Figure 6.30 (b) illustrates that in the fitting of the 10 sensorgrams a large number of the parameter pairs had high correlations. Comparing this with Fig. 6.30 (a) we see that there are consistently high correlations within the  $R_i$  parameters, and within the  $T_{iB}$  parameters. Interestingly the parameter  $k_{a1}$  has a consistently negative correlation with most of the other parameters.

This makes sense, parameters that represent the same quantity but for a different lane are correlated across lanes.

These box-plots suggest that there may be ways to reduce the complexity of the model.

For example, the parameters  $R_i$  could be combined into a single parameter  $R$  representing the maximum density of analyte that can attach to the ligand on the sensor chip in  $\text{pg}^{-1} \text{nm}^2$  at each of the interaction spots. The parameters  $T_{iB}$  could be combined to a single parameter  $T_B$ , allowing FACSIMILE to vary the proportion of antigen isolated or non-isolated at each interaction point simultaneously (See Eq. (6.6) ).

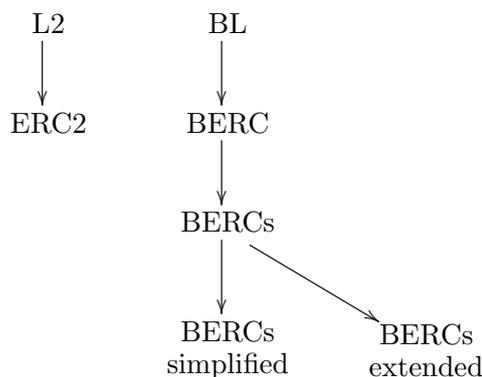
However such reductions were rejected. The parameters  $R_i$  and  $T_{iB}$  are only marked as NWD 1/57 and 2/57 times respectively. In part, this is because the BERCS model accurately represents the features of the data, but additionally because each sensorgram contains a large number of measurements, 6 measurements of the change of angle of SPR every 0.9 seconds for close to 2000 seconds. This means for every parameter there are over 600 individual SPR measurements.

## 6.7 Discussion

In this chapter, we have seen improvements in modelling lead to improvements in fitting measured by decreased RSS (see Fig. 6.8, Fig. 6.12, Fig. 6.20(a), Fig. 6.23(a) ) and decreased variation in parameter estimation between replicates (see Fig. 6.6, Fig. 6.10, Fig. 6.14, Fig. 6.20(b), Fig.6.23(b)) .

The process of model development is illustrated in 6.31. Initially, the 2 configuration versions of the Langmuir and ERC models were used. The ERC2 model got consistently better RSSs but its parameter estimates were no more consistent. This appears to be because they did not accurately model the interactions of antibody bound into dimers and into trimers. As a result, the BL model was fit to data along with the BERCS model. Although RSSs continued to reduce, and parameter estimates improved (particularly for the BERCS model), they did not reach the level of consistency desired.

The BERCS model was developed to deal with features of the data that were not well fit by



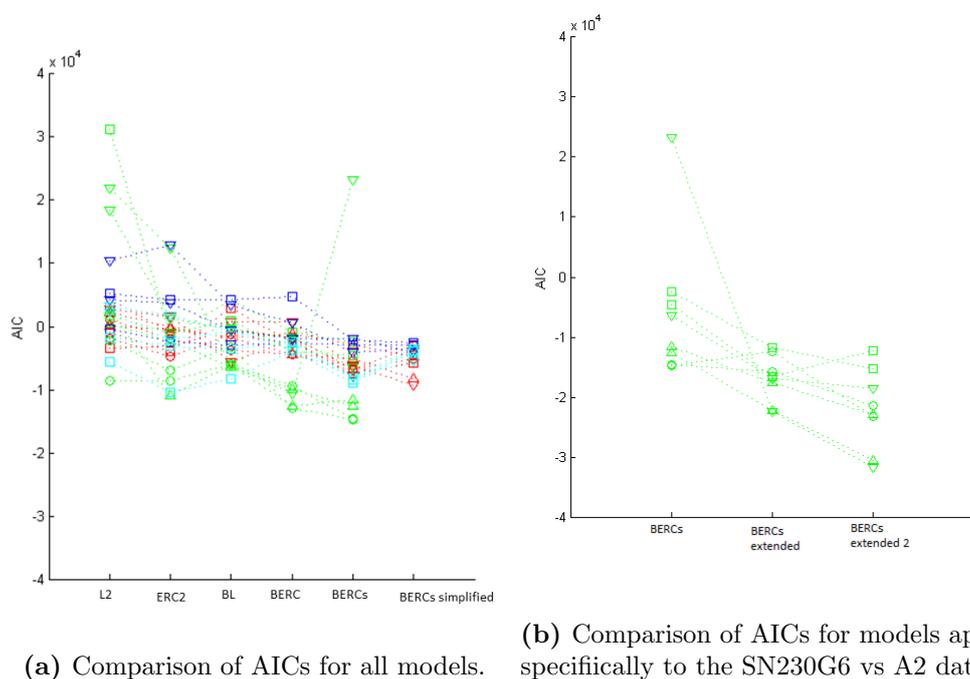
**Figure 6.31:** Diagram of model development.

the BERC model, such as the output at interaction spots of differing concentration crossing over or converging. This model was simplified to allow for faster parameter estimation, and extended to investigate systematic errors in a specific dataset. Finally, the simplified model was used to fit an expanded version of the first dataset and a second dataset.

However, these improved models have generally also been more complex. It is necessary to weigh the complexity of the models against the improvements they make in explaining the data. Figure 6.32 displays the Akaike information criterion (AIC) for each of the fits presented in this chapter. This illustrates, that although the number of parameters has increased from model to model, this has been justified by improved AIC.

The BERCs simplified model achieved a reduction in the variation of parameter estimates, for both the primary dataset used in this chapter (Section 6.4) and a secondary dataset (Section 6.6). The significance of the former is that that this validates the model for use with data taken at a wide range of analyte concentrations and ligand densities, despite the presence of artefacts.

The significance of the later is that improved experimental techniques, reducing the concentration of analyte and density of ligand which lead to curves with fewer artefacts,



(a) Comparison of AICs for all models. (b) Comparison of AICs for models applied specifically to the SN230G6 vs A2 dataset.

**Figure 6.32:** Comparison of AICs.

also allowed a level of precision vastly greater in the key SPR benchmark studies (Rich *et al.*, 2009; Yang *et al.*, 2016) (see Fig. 1.15).

This new level of precision allows us to detect the differences in association and dissociation rate constants for the reaction of a monoclonal antibody against the same epitope when it is on a separate antigen (see Fig. 6.28 and Table 6.12). These improvements in modelling made the paper (Daga *et al.*, 2017) possible (shown in the appendix, Section B.1).

It is hard to compare the fits made in this chapter to those in the literature this is because most papers do not include residual plots, or numerical measures of error such as RSS, AIC or root-mean-square error (RMSE). This is because most of the software in which sensorgrams are generally processed doesn't include these features (Luo, 2013; Richard B. M. Schasfoort, 2008; Marquart, 2013).

As a result, the strategy that is advised for the best fitting of curves is in experiment

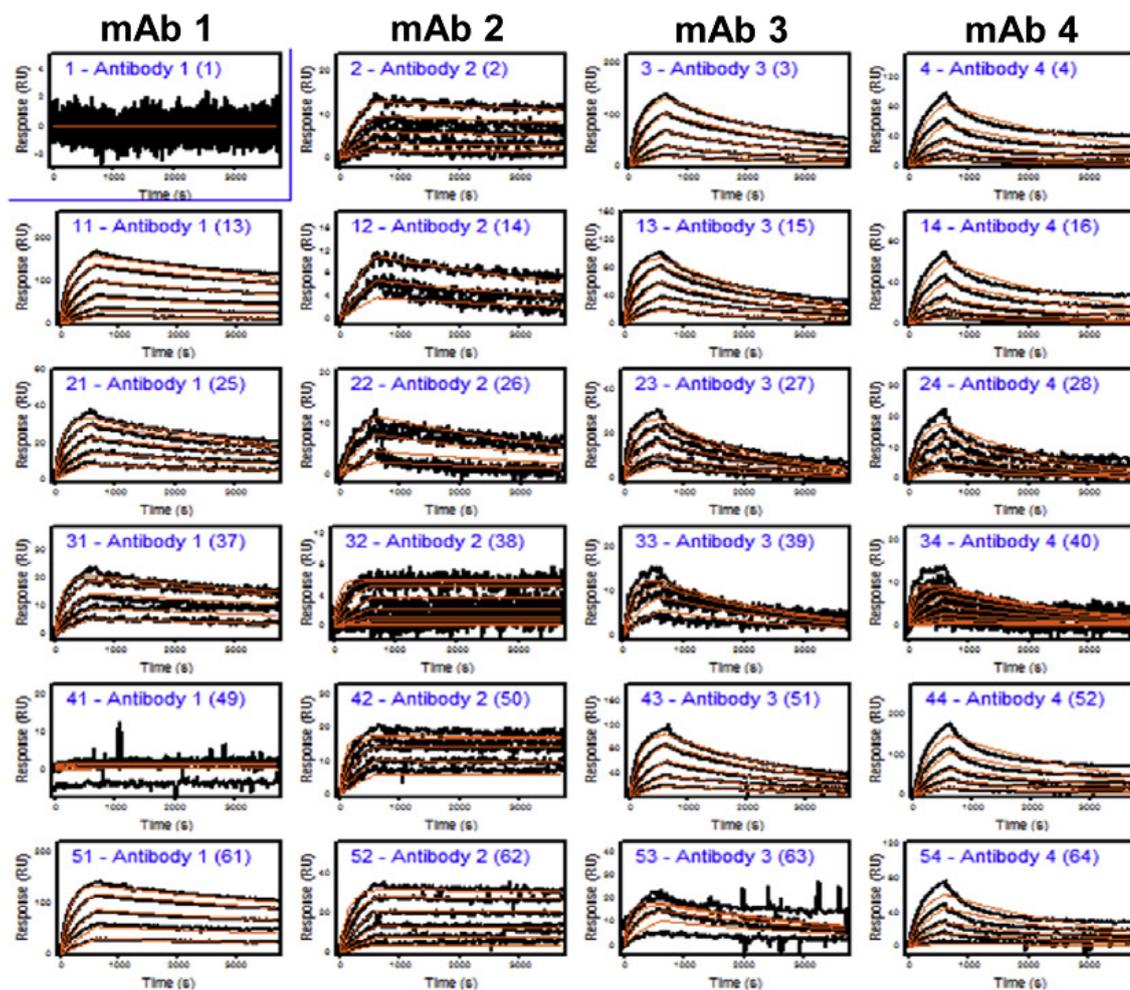
design, for the experimenter to minimise the amount both of analyte and ligand in the experiments (Richard B. M. Schasfoort, 2008; Marquart, 2013). This reduces the effect of mass transport and multivalent binding but potentially makes the experiments less useful.

For an analyte that binds *in vivo* multivalently, it is less useful to have experimental data that primarily shows the effect of monovalent binding.

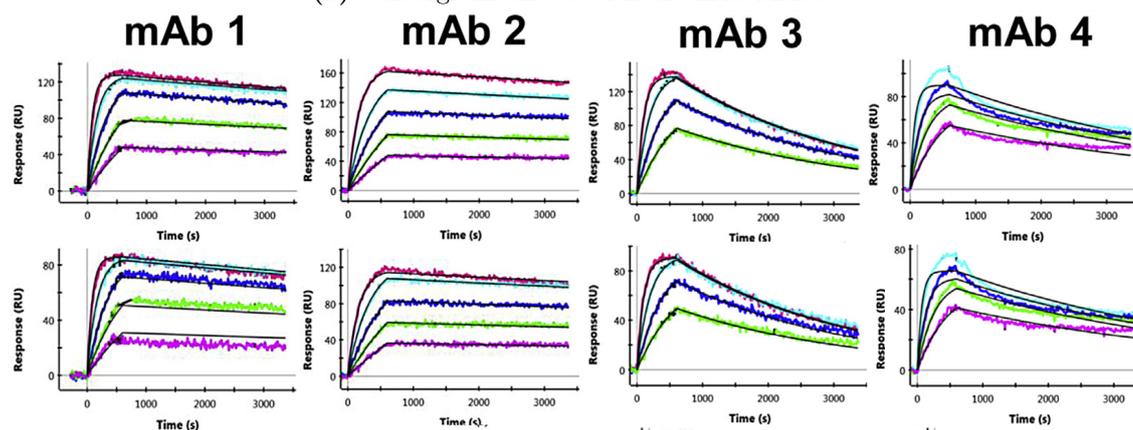
Additionally, this may be a reason for the higher level of variance between rate constants in the literature. Even when the concentration of analyte and density of ligand are low some bivalent binding should be expected to occur. This might only have a small effect in the association phase, but in the dissociation phase, the bivalently bound antibody remains attached to the chip longer than a monovalently bound antibody. As a result, it should have a notable effect late in the association phase.

Rather than just leading to poor fits, this can lead to unusual estimates for the dissociation rate constant, as the model is fitted to try and allow for rapid dissociation of monovalently bound antibody in the early dissociation phase, and the slower dissociation of antibodies that have been bivalently bound throughout most of the phase.

Consulting the figures of Rich *et al.* (2009) and Yang *et al.* (2016), it is hard to make statements about the goodness of fit and distribution of the errors of their curves in the same way as comments have been made in this chapter. This is because they are printed very small, with one page in Yang *et al.* (2016) containing as well as text, 160 sensorgrams, each with up to individual 6 curves, without residual plots. This is indicative of the fact that there is not a focus on the goodness of fit in these benchmark studies. The first 24 sensorgrams from the 160 sensorgram figure are shown for comparison in Fig. 6.33(a), as well as in Fig. 8 sensorgrams showing the same interactions on a different SPR platform.



(a) Sensorgrams made with the IBIS MX96.



(b) Sensorgrams made with the XPR36.

**Figure 6.33:** Sensorgrams for interactions of 5 antibodies with antigen, taken with various SPR platforms. Adapted from Yang *et al.* (2016).

The best fits in these papers are generally achieved by using very low analyte concentration and ligand density. This results in curves with very low maximum responses (in many cases less than 50 RU), meaning the sensorgram data is more affected by measurement error. Additionally, because these are fitted with the L1 model, which assumes all analyte is monovalently bound, the model can only produce dissociation curves that exponential decays. As the data generally shows faster dissociation at the beginning of the dissociation phase (due to initially monovalently bound antibody rapidly dissociating) and slower dissociation at the end of the phase (due to initially bivalently bound antibody needing to become monovalently bound before dissociating), it is never perfectly fit with an exponential decay curve. As a result, the best fits typically show two systematic errors, particularly in their highest concentration curves, the peaks of the output at the transition being undercut (Fig. 6.33(a) column 1 rows 2-5), or the model leaving the data at the end of the dissociation phase (Fig. 6.33(a) column 1 rows 2-5).

In some cases, a fit worse than the worst fit shown in this chapter is achieved, where the model curve remains outside the data for at least one phase for a single curve per sensorgram (Fig. 6.33(a) column 2 rows 2, column 3 row 3). In some cases, the model remains wholly outside of the data for the majority of either phase for multiple curves within a sensorgram ( Fig. 6.33(a) column 4 all rows, Fig. 6.33(b) row 4 all columns).

These fits are a strong contrast to fits with the BERCs model, like those shown in Fig. 6.27, which have model curves entirely within the data, as shown by their plots of residuals. They highlight the need for the development and use of a model that incorporates features of antibody-antigen binding that the L1 model does not, such as the BERCs model.

Another way in which these sensorgrams highlight the value of the modelling work in this chapter is that they exhibit the features that motivated the development of the BERCs model - crossovers and convergences. It is hard to distinguish between the two features in Fig. 6.33(a) because all lanes have their data in black, however numerous examples

exist which are either one or the other, column 1 row 4, column 2 row 2-5, column 3 row 2-5, column 4, row 3,4. We see these same features more clearly in Fig. 6.33(b) in every sensorgram except column 2 row 1.

The BERCs model was specifically designed to deal with these features. The commonly available models on software like the L1 model are unable to produce fits with these features, so always have systematic errors when processing data with them.

A variety of avenues exist for further work. More SPR experiments could be conducted on the same interactions, to produce more points on Fig. 6.28, and further establish the differences in the rate constants of the various interactions. SPR experiments could be conducted on other interactions, and their rate constants could be established and compared. Alternatively, a benchmark study similar to Rich *et al.* (2009) or Yang *et al.* (2016) could be conducted, where the BERCs simplified model is used exclusively to model the data. This could help establish the model as the standard model for SPR experiments with antibody analytes and antigen ligands.

A more cost-effective form of this last option would be to request the data from previous benchmark studies and process it with the BERCs simplified model.

The methods here developed are not sufficient to do any kind of population analysis of affinity, for two reasons, firstly the affinity of an interaction is determined by both the antibody and antigen, and secondly that humans have constantly changing populations of antibodies inside them (see Section 1.2). However, the success of these methods lay the groundwork for theoretical work in Chapters 7 and fitting in Chapter 8 which will allow experiments conducted with patient samples in Section 4.5 - 4.9 to begin to stratify patient antibody-antigen affinity.

An interesting development parallel to this study is that It has been proposed that HIV and possibly other pathogens use the same mechanisms the BERCs model was developed

to incorporate to avoid antibody-mediated neutralization (Klein & Bjorkman, 2010; Cai, 2005); i.e. it is proposed that they have evolved so that their antigen are far enough that an IgG antibody binding to one cant bind to another- effectively reducing the avidity of of an antibody to its affinity.

-

# Chapter 7

## Modelling the reaction kinetics of polyclonal and mixtures of monoclonal antibody in SPR experiments

SPR experiments with polyclonal mixtures of antibody as analyte are a relatively novel idea. As a result, only one paper analysing these kinds of experiment has been published, Evans *et al.* (2010). This paper used the LwT1 and ERC1 models, defined by Eq. (5.3) and Eq. (5.6) respectively, and achieved fits characterised by large systematic errors (see Fig. 1.17, and discussion in Subsection 1.4.2).

In Chapter 6, it was established that models that did not allow for the analyte to bind bivalently with the antigen had higher AICs, RSSs, and produced less consistent estimates of affinity between estimates (See Fig. 6.32). As the models used in Evans *et al.* (2010) did not allow for these features of their reactants, this may explain their poor fits. However, as patient-derived polyclonal antibody is expected to bind to multiple epitopes of the same antigen (see Sections 1.2 and 1.3) more complex antibody-antigen interactions may be possible than with monoclonal experiments.

Three models in the literature, have been identified as possessing features somewhat appropriate for simulating SPR experiments with polyclonal mixtures of antibody as analyte, and individual antigen as the ligand, the Langmuir two ligand model and the langmuir two concentration model (Bio-Rad, 2010), and the two concentration ERC model (Evans *et al.*, 2013).

It may be possible to construct experiments where the interactions of the analyte and ligand would be similar to the assumptions of these models. As each model is discussed corresponding experimental designs will all also be discussed.

However the experiments these models will be fit to, those of Section 4.5 - 4.9 were not designed to limit bivalent binding, multiple antibodies binding to multiple epitopes on single antigen, so a more general model is also presented.

This general model is modelled on the underlying chemistry of antibody-antigen interactions and allows for multiple antibodies binding to multiple epitopes on the same antigen, and binding bivalently to identical epitopes on multiple antigens.

Versions of it include the assumptions of the BERCs model, as defined in Eq.(5.15), namely bivalence of antibody, spatial inhomogeneity of antigen distribution, and variation in the concentration of antibody close to the receptor layer. As the BERCs model was demonstrated in Chapter 6 to be dramatically better at minimising the RSS and AIC of fits to data and producing consistent affinity estimates for systems where only one antibody binds to one antigen, these features were expected to lead to the best possible fits and affinity estimates for the data.

## 7.1 Previously Published Models

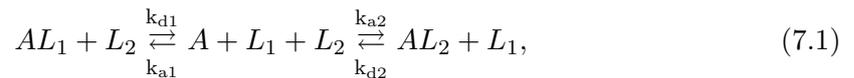
### 7.1.1 Langmuir model

Extensions to the single configuration Langmuir model, discussed in Section 3.3, have been made and are among the models available to researchers using the Biorad Proteon Manager (Bio-Rad, 2010, 2011) that allow for heterogenities in either the ligand or the analyte.

The *Langmuir two ligand model* (L2L) is an adaptation of the Langmuir model, that makes additional assumptions that could represent some of the features of polyclonal antibody binding, recomended in Bio-Rad (2010) for use with polyclonal antibody.

Each antigen molecule has multiple epitopes, each of which can bind to an individual antibody with a corresponding paratope. To reduce complexity analyte is modelled as homogeneous, but can bind into two separate dimers with the two separate epitopes.

Its state-space equation is derived from the reaction equation:



where  $A$ ,  $L_1$ ,  $L_2$ ,  $AL_1$ ,  $AL_2$ , represent an antibody, epitopes of either type respectively, an antiobdy-antigen dimers of either type respectively. Notably whilst earlier models have used  $L$  to represent antigen, for this model it represents the antigens epitopes. This is because some antigen on the sensor chip may be bound in such a way as that only one epitope is available for binding, or such that two are available. As a result it is more useful to think in terms of antibody-epitope reactions rather than antibody-antigen reactions.

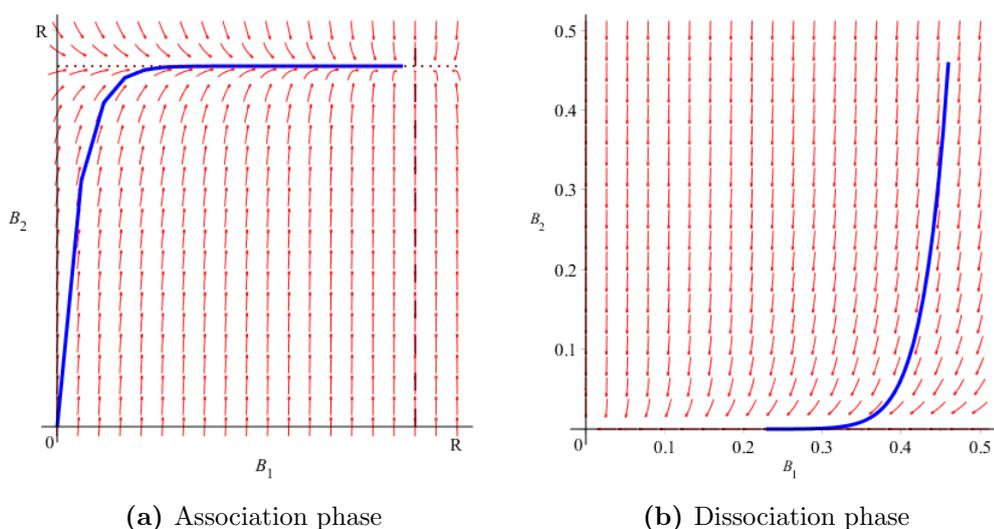
The model is defined as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} C_T & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = k_{a1}I(R_1 - x_1) - k_{d1}x_1 \\ \frac{dx_2}{dt} = k_{a2}I(R_2 - x_2) - k_{d2}x_2 \\ x(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, \mathbf{p}) = \alpha(x_1(t, \mathbf{p}) + x_2(t, \mathbf{p})), \end{cases} \quad (7.2)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_T, R_1, R_2)$ . The parameters  $k_{a1}$  and  $k_{a2}$ ;  $k_{d1}$  and  $k_{d2}$ ; are the rate constants for association and dissociation of the two binding species respectively measured in  $\text{M}^{-1}\text{s}^{-1}$  and  $\text{s}^{-1}$  respectively,  $C_T$  is the concentration of analyte flowing over the sensor chip in nM and  $R_1$  and  $R_2$  are the maximum density of antibody that can attach to antigen on the sensor chip in  $\text{pg nm}^{-2}$ . The state variable  $(x_1, x_2)^\top = \mathbf{x}(t; \mathbf{p}, I) \in \mathcal{X}$ , represents the concentrations of two configurations of analyte-ligand binding in  $\text{pg nm}^{-2}$  and the state-space is defined as  $\mathcal{X} = [0, R_1] \times [0, R_2]$ . The input function  $I(t)$  represents concentration of analyte at the machines inlet and  $T_a$  and  $T_d$  are intervals on the real line representing the association and dissociation phase, defined as  $T_a = [0, t_1], T_d = [t_1, t_2]$  respectively, and  $\alpha$  is a known constant  $1 \text{ RU pg}^{-1} \text{ nm}^2$ .

As we would expect the process by which antigen are bound to the chip to be as likely to make one epitope inaccessible as the other we generally expect  $R_1 = R_2$ , and this assumption can be used to reduce the number of unknown parameters of the model when fitting.

A phase portrait for the two configuration binding system is shown in Fig. 7.1. This phase portrait is a contrast to the ones shown in Chapter 5; this is because the area of the phase plane that corresponds to concentrations of bound antibody that could occur in the



**Figure 7.1:** Phase portraits of the two ligand Langmuir model in the association and dissociation phase, with null clines (dotted and dash lines).

Parameter	$k_{a1}$	$k_{d1}$	$k_{a2}$	$k_{d2}$	$R_1$	$R_2$	$C_T$	$t_1$	$t_2$
Unit	$M^{-1}s^{-1}$	$s^{-1}$	$M^{-1}s^{-1}$	$s^{-1}$	$pg\ nm^{-2}$	$pg\ nm^{-2}$	nM	s	s
Value	6E-4	7E-3	6E-3	7E-2	500	500	100	500	1000s

**Table 7.1:** Values taken by parameters in Fig. 7.1

experimental system is square rather than triangular (compare Fig. 5.8). This is because the amount of one antibody bound has no effect on how much of the other antibody can be bound.

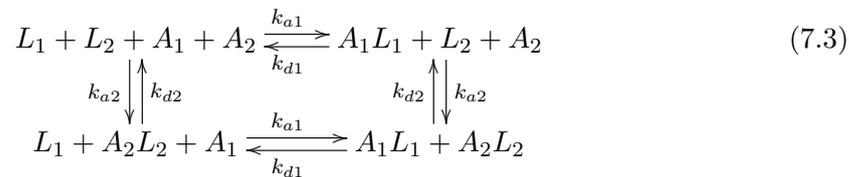
In Fig. 7.1 (a) the nullclines enclose a rectangular region of phase space in which either component of the state variable is increasing. At the start of the association phase analyte initially binds mainly in the configuration with the higher association rate, but as it approaches the nullcline fewer antigen have the right epitope available to bind. As a result the increase in total bound concentration becomes dominated by increase in concentration of the antibody with the lower rate constant.

In the dissociation phase, Fig. 7.1 (b), the nullclines move so that they run along the axes and the steady state is at the origin. The system equations become identical to those of

the L2 model (Eq. (5.1)) and the components of the state variable rapidly decrease.

In Chapter 6 it was established that bivalent binding, as well as transport effects, were components of models necessary for them to fit the data accurately and give consistent estimates of association and dissociation rate constants.

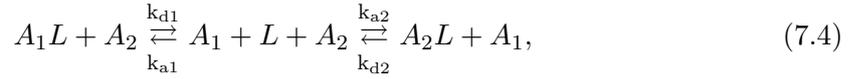
Ignoring these two features of the model, the model assumes all antibody in the experiment can bind to all of the epitopes, and that the binding interaction is solely determined by the epitope in the reaction (see Eq. (7.1)), whereas antibody should be expected to bind only to epitopes its paratope is comparable with. With a chemical equation such as:



where  $A_1$ ,  $A_2$ ,  $L_1$ ,  $L_2$ ,  $AL_1$ ,  $AL_2$ , represent antibody with paratopes compatible with either type of antigen respectively, epitopes of either type respectively, and antibody-antigen dimers of either type respectively.

The experiments the L2L model would be most appropriate for fitting would be measuring the binding affinity of paratopes separated from a monoclonal antibody, developed specifically against a single epitope on an antigen, which also binds non-specifically to a second epitope on the antigen, hence having the reaction equation Eq. (7.1). One potential explanation for why this model despite being recommended for use with polyclonal antibody binding to antigen, has only been published in studies without antibody or antigen reactants (Kuziemko *et al.*, 1996; Nakajima *et al.*, 2001; Ricard-Blum *et al.*, 2004) is that the disparity between Eq. (7.1)) and the reactions we would expect in polyclonal antibody are so great that the model doesn't allow for sufficiently good fitting or parameter estimation.

The *Langmuir two concentration model* (L2C) assumes that ligand is homogeneous but that there are two types of analyte, and that two separate kinds of dimer may form when ligand reacts with either. Its state-space equation is derived from the reaction equation:

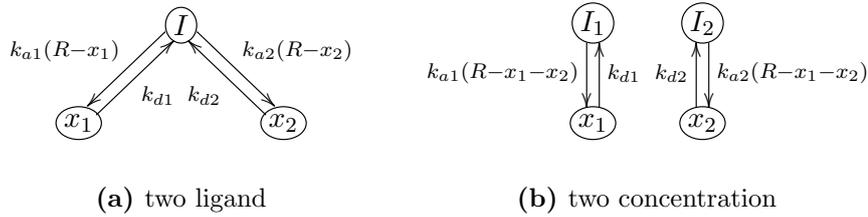


where  $A_1$ ,  $A_2$ ,  $L$ ,  $A_1L$  and  $A_2L$ , represent two antibody types respectively, antigen, and antibody-antigen dimers of either type respectively.

The model is defined as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} 1 & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = k_{a1}IC_{T1}(R_1 - x_1 - x_2) - k_{d1}x_1 \\ \frac{dx_2}{dt} = k_{a2}IC_{T2}(R_2 - x_1 - x_2) - k_{d2}x_2 \\ x(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, \mathbf{p}) = \alpha(x_1(t, \mathbf{p}) + x_2(t, \mathbf{p})), \end{cases} \quad (7.5)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_{T1}, C_{T2}, R)$ . All parameters except  $C_{T1}, C_{T2}$  and  $R$  have the same definitions and are measured in the same units as the L2L mode (Eq. (7.2)). The parameters  $C_{T1}$  and  $C_{T2}$  represent the concentration of either antibody type respectively (nM) and  $R$  is the maximum density of antibody that can attach to antigen on the sensor chip in  $\text{pg nm}^{-2}$ . The state variable  $(x_1, x_2)^\top = \mathbf{x}(t; \mathbf{p}, I) \in \mathcal{X}$ , represents the concentrations of two configurations of analyte-ligand binding in  $\text{pg nm}^{-2}$  and the state-space is defined as  $\mathcal{X} = [0, R] \times [0, R]$ . The input function  $I(t)$  represents whether antibody is flowing in from the machines inlet and  $T_a$  and  $T_d$  are intervals on the real line representing the association and dissociation phase,



**Figure 7.2:** Compartmental diagram of the two configuration Langmuir model with analyte concentration  $I$  and concentrations of bound analyte  $x_1(t)$  and  $x_2(t)$ .

defined as  $T_a = [0, t_1], T_d = [t_1, t_2]$  respectively, and  $\alpha$  is a known constant  $1 \text{ RU pg}^{-1} \text{ nm}^2$ .

This system is a reparameterisation of the two configuration Langmuir model (introduced in example 2.10) where  $IC_{T_1}k_{a,1}$  and  $IC_{T_2}k_{a,2}$ , replace  $Ik_{a,1}$  and  $Ik_{a,2}$ , and all other parameters remain the same. As a result Figure 2.5 is also a phase portrait of this system.

The antibody experiments the L2C model would be most appropriate for fitting would be measuring the binding affinity of paratopes separated from a polyclonal antibody sample in which only two antibody clones bind to the antigen, and both bind to the same epitope, hence having the reaction equation Eq. (7.4).

Compartmental model diagrams for two forms of the Langmuir model are shown in Fig. 7.2.

Besides the weaknesses mentioned in the previous section, both models do not allow antibody concentration to vary in the proximity of the antigen, or to bind bivalently, the two features that were identified as being key to good fits and accurate parameter estimation in the previous section.

However there is another model in the literature that overcomes the first of these problems.

### 7.1.2 Effective rate constant approximation model

The two concentration version of the ERC model (ERC2C) was presented in (Evans *et al.*, 2013). It represents two species of antibody competing over a single species of antigen as in Eq. (7.4), but with the concentrations of antibody at available to bind to ligand being in quasi-steady-state.

The model is defined as:

$$\Sigma = \begin{cases} I(t) = \begin{cases} 1 & : t \in T_a \\ 0 & : t \in T_d \end{cases} \\ \frac{dx_1}{dt} = \frac{k_{a,1}IC_{T1} \left( R - \sum_{i=1}^2 x_i \right) - k_{d,1}x_1}{1 + (k_{a,1}/k_m) \left( R - \sum_{i=1}^2 x_i \right)} \\ \frac{dx_2}{dt} = \frac{k_{a,2}IC_{T2} \left( R - \sum_{i=1}^2 x_i \right) - k_{d,2}x_2}{1 + (k_{a,2}/k_m) \left( R - \sum_{i=1}^2 x_i \right)} \\ x(0; \mathbf{p}, I) = (0, 0)^\top \\ y(t, \mathbf{p}) = \alpha (x_1(t, \mathbf{p}) + x_2(t, \mathbf{p})), \end{cases} \quad (7.6)$$

where the parameter vector is defined as  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, k_m, C_{T1}, C_{T2}, R)$ . All parameters and variables except  $k_m$  have the same definitions and are measured in the same units as the L2C model (Eq. (7.5)). The parameter  $k_m$  is the transport coefficient, describing diffusive movement of analyte between the flow and the volume in contact with the surface measured in  $\text{pgnm}^{-2} \text{M}^{-1}\text{s}^{-1}$ .

This system is a reparameterisation of Eq. (5.7) where  $I_1k_{a,1}$  and  $I_2k_{a,2}$ , replace  $Ik_{a,1}$  and  $Ik_{a,2}$ , and all other parameters remain the same. As a result Figure 5.8 is also a phase portrait of this system.

As this model is based on the same chemical equations as the L2C model, it is appropriate

for the same kinds of experiment design, that is measuring the binding affinity of paratopes separated from a polyclonal antibody sample in which only two antibody clones bind to the antigen, and both bind to the same epitope.

As this model uses the effective rate constant approximation rather than assuming that the concentration of antigen in contact with the ligand is the same as in the sample this model may give better fits and parameter estimates than the L2C model (Chaiken *et al.*, 1992).

Similar to this a two ligand ERC model could be developed, however one was not. This is because of the criticisms made in the previous section of the chemical equation assumed by that model, and the fact that none of these models include bivalence.

The monoclonal experiments of Chapter 4 typically showed a dissociation phase that was characterised by dissociations that looked visually similar to a sum of two exponential decays with vastly different rates. In the early part of the dissociation phase dissociation was rapid, in the later part of the phase, despite still high measurement values change is very slow. Whilst these features lead to huge variations in rate constant estimates for models that didn't include bivalence (See Fig. 6.6), bivalent models gave consistent estimates (see Fig. 6.23(b) and Fig. 6.28) because they included chemical equations that accurately represented the interactions of the reactants.

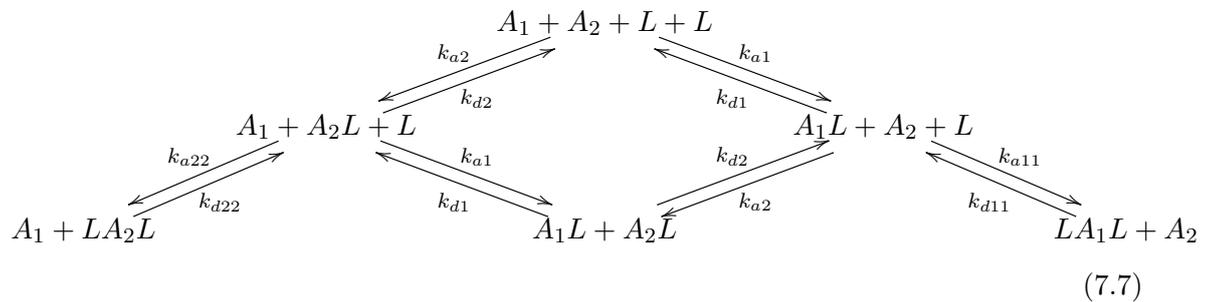
The polyclonal experiments of Chapter 4 typically showed the same features. As a result it is expected that parameter estimation with non-bivalent models would lead to similarly inaccurate parameter estimates. Particularly to match the fast and slow changes in the data caused by the bivalence of the antibody in the experiments, the L2C, L2L and ERC2C models would give a large value for one of  $k_{d1}$  and  $k_{d2}$  and a small value of the other, regardless of the mixture of antibody-antigen interactions involved in the experiment, making the parameter estimates useless as measurements of properties of the antibody-antigen interactions.

As a result it was necessary to develop a model sophisticated enough to deal with bivalent antibody interactions as well as multiple simultaneous antibody-epitope interactions.

## 7.2 General model of polyclonal antibody binding

Previously Eq. (7.3) and Eq. (7.4) were given as the chemical equation for two antibodies binding to a single antigen-epitope and two separate antigen-epitopes respectively, when constrained so as the interactions are exclusively monovalent. When bivalent interactions are allowed these interactions scheme becomes dramatically more complicated, even if we limit our consideration to complexes with two antibody or fewer.

The chemical equation for two antibodies competing over a single epitope is:



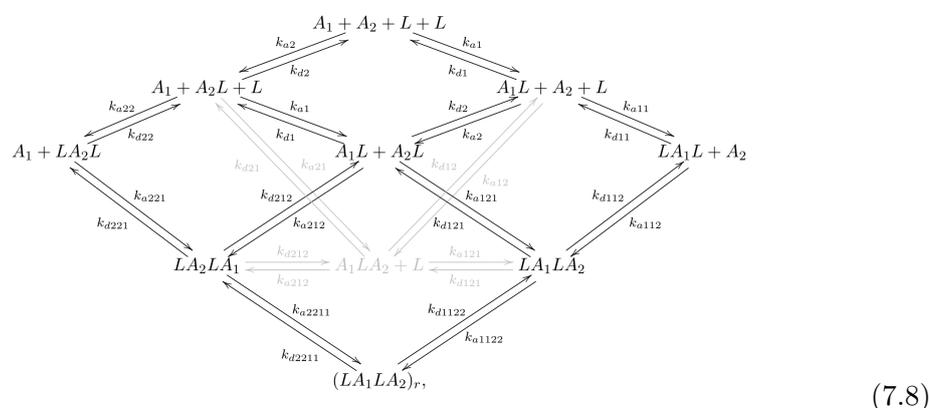
where  $A_1$  and  $A_2$  represent two species of antibody clones with the same epitope target, and  $L$  is an antigen with this epitope;  $A_1L$  and  $A_2L$  represent antibody-antigen dimers, containing antibody of either species respectively;  $LA_1L$  and  $LA_2L$  represent antigen-antibody-antigen trimers, containing antibody of either species respectively;  $k_{a1}$ ,  $k_{d1}$ ,  $k_{a2}$  and  $k_{d2}$ , are the primary association and dissociation rate constants, and  $k_{a11}$ ,  $k_{d11}$ ,  $k_{a22}$  and  $k_{d22}$ , are the secondary association and dissociation rate constants.

In the Chapters 5 and 6, the BERCS simplified model was developed and was demonstrated

to give better fits and more consistent parameter estimates than the other models. It is assumed that secondary dissociation rate was double that of the primary, because when an antibody is bivalently bound, it has twice as many bonds that it may dissociate from, i.e.  $k_{d11} = 2k_{d1}$  and  $k_{d22} = 2k_{d2}$ . This can be used to simplify Eq. (7.7). Additionally we may assume that the secondary association rates are each a product of the primary association rates and a constant  $\beta$  representing the effective change in concentration that occurs when an antibody is bound to an antigen and its free paratope is constrained to a smaller region (see Fig. 7.3 (a)), i.e.  $k_{a11} = \beta k_{a1}$  and  $k_{a22} = \beta k_{a2}$ , where  $\beta$  is measured in  $\text{Mnm}^{-2}\text{pg}^{-1}$ .

The chemical equation for two antibodies binding to separate epitopes on the same antigen is significantly more complex. We are ultimately interested in SPR experiments, where because of the way the antigen is attached to the surface of the reaction chip, complexes with large numbers of antigen and antibody are unlikely to form, so in the chemical equations we limit our consideration to tetrameric and smaller complexes. If we did not, these chemical equations would be vastly more complex and would explain how complexes such as those pictured in Fig. 1.8 are formed. Even constrained to this level of complexity, some of the interactions cross each other so it is necessary to print them in two colours.

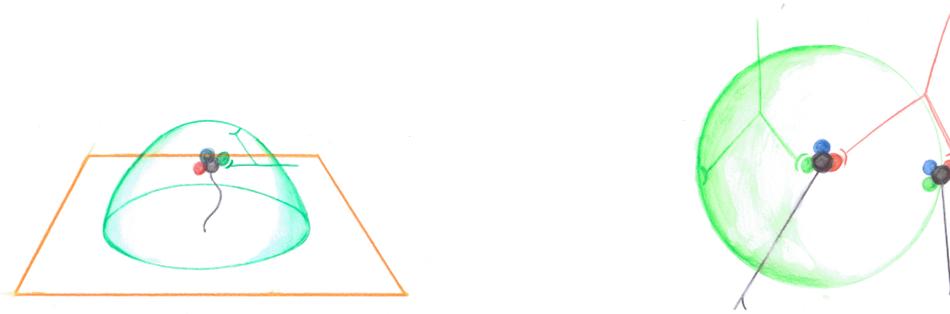
The chemical equation for two antibodies binding to separate epitopes on the same antigen is:



where  $A_1$  and  $A_2$  represent two species of antibody clones, and  $L$  is an antigen with this epitope;  $A_1L$  and  $A_2L$  represent antibody-antigen dimers, containing antibody of either species respectively;  $LA_1L$  and  $LA_2L$  represent antigen-antibody-antigen trimers, containing antibody of either species respectively;  $A_1LA_2$  represents the only possible antibody-antigen-antibody trimer;  $LA_1LA_2$  and  $LA_2LA_1$  represent antigen-antibody-antigen-antibody tetramers; and  $(LA_1LA_2)_r$  represents a tetramer in which both antibody are bivalently bound, forming a ring like structure.

The association and dissociation rate constants are named so that the final index represents whether the association or dissociation involves the paratope of an antibody of the first or second type, and previous indexes represent the indexes of the paratopes that remain bound in the case of dissociation, or become bound in the case of association. That is  $k_{a212}$  represents the association rate constant of a trimer bound together by paratopes of the first and second kind, binding to another antigen through a paratope of the second kind binding to its corresponding epitope.

The possibility of the tetramer in which every antibody is bivalently bound,  $(LA_1LA_2)_r$ , may explain the results of Section ??, in which experiments showed such slow rates of



(a) The free paratope of an antibody attached to a single antigen is constrained to a hemisphere centred where the linker joins the chip.

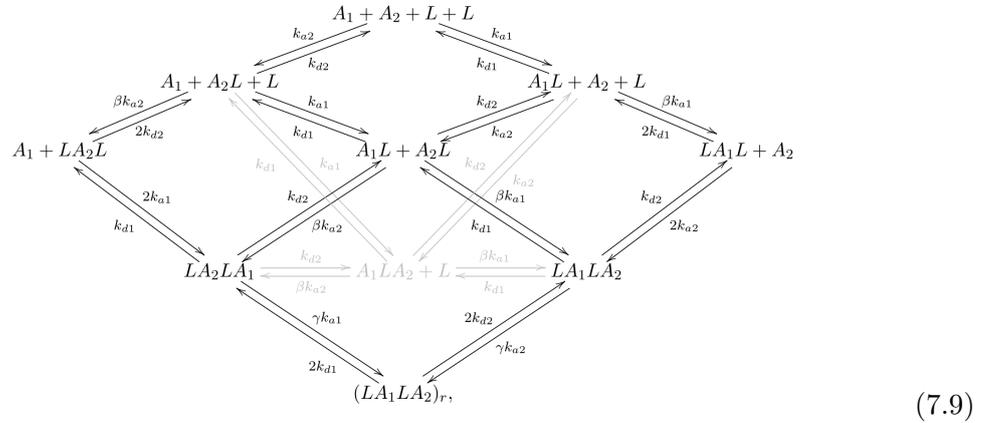
(b) The free paratope of an antibody attached to a single antigen is constrained to a sphere centring on the epitope of the antigen it is bound to.

**Figure 7.3:** Diagrams of the regions in which the free paratope of a bound antigen can move.

dissociation that the SPR estimates of  $k_D$  using the default software produced values as extreme as  $3.8E - 24$ .

These experimental results and this theory motivated the experiments of Section 4.8 which showed that when the antigen HLA-A2 was used as ligand in experiments with a identical concentrations of antibody, maximum binding was achieved when a mixture of two antibodies was used (SN607D8 and SN230G6). The analysis of these experiments is discussed in detail in Section 8.3, but the provisional result that the presence of the other antibody leads to greater total bound concentration at the end of the experiment, suggested again that in some way that the antibody were helping each other to remain bound.

Making the same assumptions as for the two antibodies competing over a single epitope, we can reduce the numbers of rate constants,  $k_{d11} = 2k_{d1}$  and  $k_{d22} = 2k_{d2}$ ,  $k_{a11} = \beta k_{a1}$  and  $k_{a22} = \beta k_{a2}$ . Similarly if neither antibody causes an allosteric effect on the antigen, or creates steric hindrance for the other antibody, then the chemical equation for two antibodies binding to separate epitopes on the same antigen becomes:



where  $\gamma$  is the constant that  $k_{aj}$  is multiplied by to give the rate constant for the free epitope of a monovalently bound antibody in a tetramer binding to the free epitope in the same tetramer (see Fig. 7.3 (b)), measured in  $\text{Mpg}^{-1}\text{nm}^2$ .

However, even if the antigen on the SPR chip are spread so that only two antigen can be a part of the same complex, there are reasons to expect the chemistry of patient polyclonal antibody binding to antigen in vitro to be more complex.

This is because we expect the number of species of antibody clones in patient serum to be vastly more than 2. In Section 1.2 the processes of hypermutation and B-cell proliferation were discussed. The immune system rapidly creates new antibodies and selects them on the basis of their affinity for foreign antigen. As a result large numbers of antibody clones may be targeted at the same antigen.

This motivates the creation of a model that can be used for experiments with varying scales of complexity. A model that could be used for mixed monoclonal antibody, as well as clinical samples. This is why the *generalised polyclonal model* (GPM) was proposed.

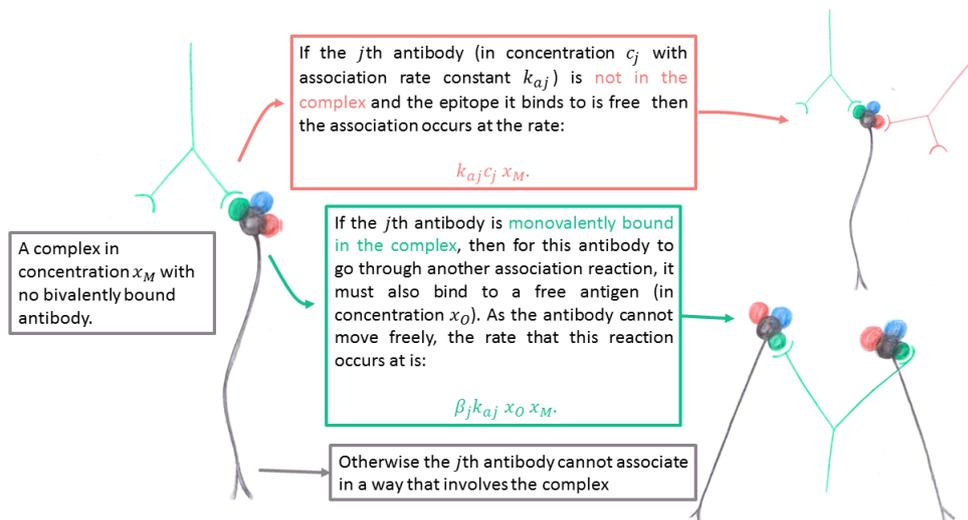
Using the assumptions behind Eq. (7.9) we can determine the rates at which many of the

interactions will happen in a system with more antibody, even though we cannot draw a chemical equation for the system without knowing the number of antibody. In these diagrams  $\beta$  is replaced with  $\beta_j$ , and  $\gamma$  is replaced with  $\gamma_j$ . This will allow us to later make substitutions like  $\beta_j = \beta \forall j$  and  $\gamma_j = \gamma \forall j$  if we later want to simplify the model, but to leave it at this level of complexity if required.

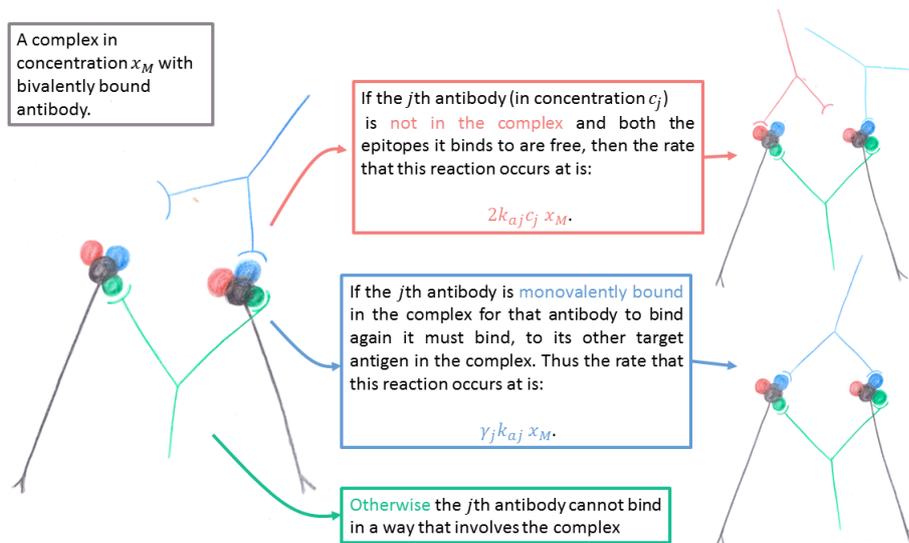
These interactions, together with their rates are visualised in Fig. 7.4 - Fig. 7.6. In these diagrams the concept of an index for antibody is used, a generic antibody is termed the  $j^{th}$  which has a concentration close to the antigen of  $c_j$ , and it has a association rate for its target antigen-epitope of  $k_{aj}$  and dissociation rate of  $k_{dj}$ . To show the different ways that reaction can occur colour coded information boxes associated with each potential reaction are shown. Their outline reflects the colour of the antibody that in the text corresponds to the  $j^{th}$  index.

As the kinds of reaction a complex can undergo, an index  $M$  used for complexes;  $x_M$  represents the concentration of a generic complex. In the diagrams we see the way the  $j^{th}$  antibody reacts is determined by properties of the complex - for example if it has 0, 1, or more than 1 bivalently bound antibody, and if the antibody we are considering is bivalently bound or not.

Whilst building a model based on the chemical equations, rather than a vast simplification, as in the cases of the L2L, L2C and ERC2C models, has the advantage that each parameter and variable would directly represent a property or an amount of a specific complex, and the possible behaviours of the experimental system would be matched by the possible behaviours of the model, it has the disadvantage of requiring new concepts and mathematical tools. This growth in complexity is illustrated in Fig. 7.4 - Fig. 7.6, whilst it might be relatively easy to keep track of an index corresponding to an antibody, an index for the complexes created will require addition features to deal with the complexity of possible complexes and their interactions both with free antibody, antigen and other



(a) Complexes with monovalently bound antibody only



(b) Complexes with one or more bivalently bound antibody only

**Figure 7.4:** Rates at which complexes grow by the addition of monomers.

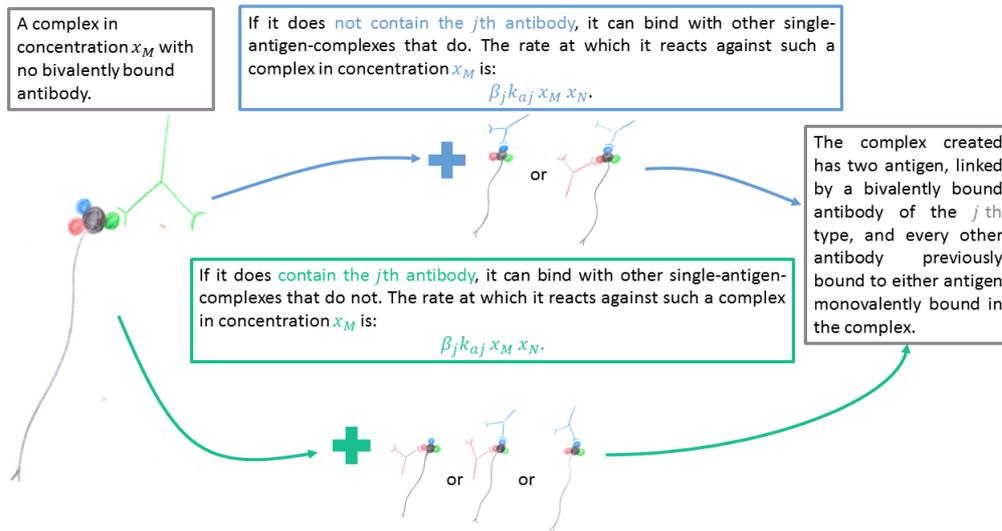
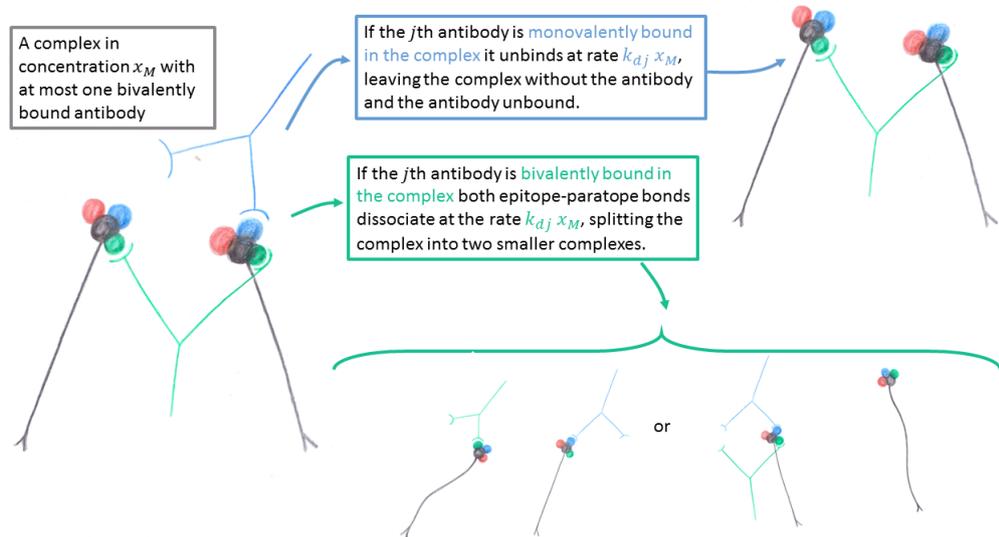


Figure 7.5: Rates at which complexes grow by binding with other complexes.

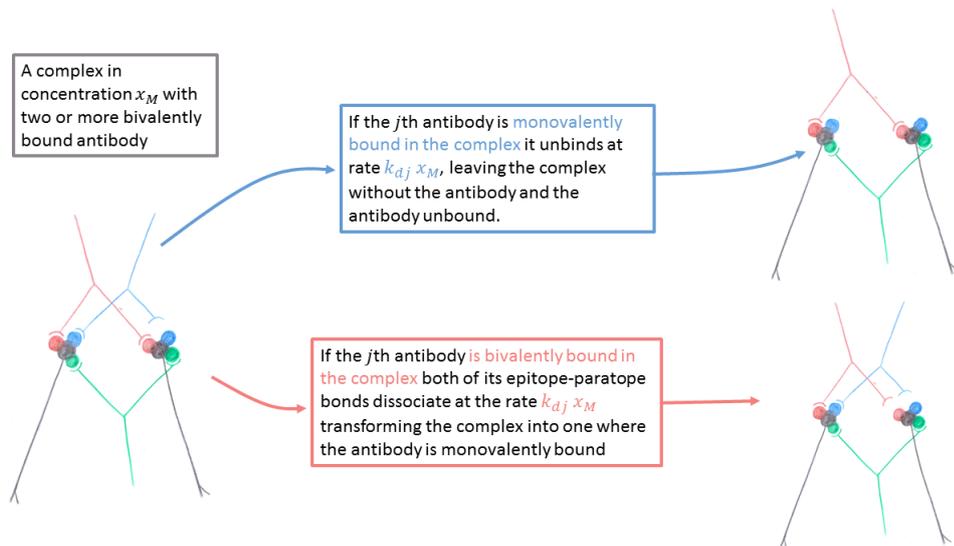
complexes. As a result the paratope-epitope matrix was developed.

### 7.2.1 Paratope-epitope matrices and related functions

The first element of this language that must be introduced is the paratope-epitope matrix. This matrix will act as a label for each of the complexes we would expect to observe in the reaction. The columns of this matrix represent the epitopes of the antigen, and the rows represent species of antibody, each entry of the matrix will have a numeral written in it showing the number of antibody paratopes bound to an epitope of that type. Here

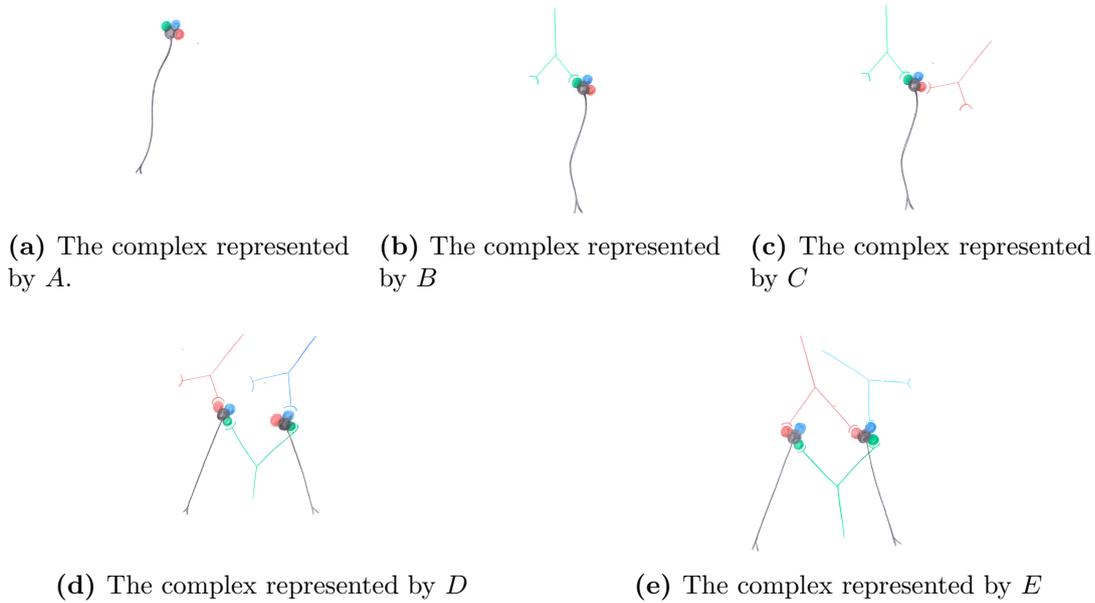


(a) Complexes with monovalently bound antibody only



(b) Complexes with one or more bivalently bound antibody only

**Figure 7.6:** Rates at which complexes shrink by the dissociation of monomers.



**Figure 7.7:** The complex represented by the paratope-epitope matrices of Eq. (7.10).

are some example paratope-epitope matrices:

$$\begin{aligned}
 A &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & B &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} & C &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\
 D &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} & E &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}
 \end{aligned} \tag{7.10}$$

These matrices represent a system where four species of antibody are binding to an antigen with three epitopes. The first antibody species binds to the first epitope, the second binds to the second, and the third and fourth antibody compete over the third epitope. The entries associated with the individual epitope-paratope pairs are coloured, so they can be cross-referenced with Fig. 7.7, in which the complexes represented by the matrices are

illustrated.

Whilst the GPE could be used for antibody-antigen interactions where there is non-specific binding, in this example, it is assumed that each antibody binds to one epitope only.

The GPE model assumes that the largest complexes will only consist of two antigens. This is for the reasons discussed in section 5.1 (see Fig. 5.1). This means that the largest number that might occur as an entry to an antibody epitope matrix would be 2.

This model also assumes that a complex will include at most one antibody of each species. This is because an antibody bound to a single antigen will dissociate much faster than an antibody bound to two separate antigen - which would have to dissociate twice.

The matrix  $A$  represents an antigen attached by linker to the chip, that is not bound to any antibody. The matrix  $B$  represents an antigen bound to a single antibody of the first type. The matrix  $C$  represents an antigen bound to an antibody of the first type and an antibody of the second type. The matrix  $D$  represents a complex including pair of antigen both bound to a single antibody of the first type, with antibodies of the second and third type also somehow bound. Notably, this notation does not distinguish between which antigen an antibody is bound to once it is in a complex with multiple antigens. The matrix  $E$  represents a complex including a pair of antigen both bound to a single antibody of the first type, and to a single antibody of the second type with an antibody of the fourth type bound to one of the pair.

The model we are about to discuss assumes that a single complex will not contain two antibodies competing for the same epitope. It is assumed that when a complex like  $D$  or  $E$  is formed the rate at which the antibody monovalently bound in the complex will become bivalently bound to the other antigen in the complex is so fast that it is unlikely a competing antibody will join the complex.

The complexes represented by these matrices are very different so they will interact in

very different ways. The complexes represented by  $D$  and  $E$  have twice as many epitopes than those represented by  $A$ ,  $B$  and  $C$ , because they include two antigens rather than one. As a result, the antibodies linking these antigens can dissociate, leading to the complexes splitting apart or the monovalently bound antibody can react with an epitope on the other antigen.

As there are such big differences between the behaviour of these complexes it is useful to have some functions that will tell us the number of monovalently or bivalently antibody in a complex, these are:

$$\text{Mono}(M) = \text{the number of 1s in matrix } M \quad (7.11)$$

$$\text{Biv}(M) = \text{the number of 2s in matrix } M. \quad (7.12)$$

It is also useful to have others that take a 0 or 1 value that can be used to determine whether the complex will undergo further reactions. These functions are defined as:

$$e(j, M) = \begin{cases} 1 & \text{if the matrix } M \text{ has a non-zero entry in the column representing} \\ & \text{the epitope the } j^{\text{th}} \text{ antibody binds to,} \\ 0 & \text{otherwise,} \end{cases} \quad (7.13)$$

$$p_k(j, M) = \begin{cases} 1 & \text{if the matrix } M \text{ includes one or more entry of the value } k \text{ in the} \\ & \text{row } j \text{ of } M, \\ 0 & \text{otherwise.} \end{cases} \quad (7.14)$$

The function  $e(j, M)$  tells us if a matrix represents a complex in which a new antibody of the  $j^{\text{th}}$  type can join. Notably, this will be either by an antibody of the  $j^{\text{th}}$  type binding to it or a complex containing an antibody of the  $j^{\text{th}}$  type binding to it. In either case, it will be through the paratope of the antibody of the  $j^{\text{th}}$  type binding to the associated antigen. This will be referred to as a binding reaction of the  $j^{\text{th}}$  type.

The function  $p_k(j, M)$  tells us if a certain paratope-epitope matrix represents a complex in which the  $j^{\text{th}}$  epitope or paratope is bound  $k$  times. Often the way a new antibody or antigen will

bind to a complex will be dependent on what is already bound. As a result it is also useful to introduce a vector valued function  $\mathbf{p}(j, M)$ , that takes as its components  $p_0(j, M)$ ,  $p_1(j, M)$  and  $p_2(j, M)$ .

Two epitope-paratope matrices that will occur a lot are given specific names -  $O$  and  $J$ .  $O$  is the epitope-paratope matrix representing unbound antigen,  $O$  has a 0 in every entry (earlier this was exemplified by the matrix  $A$ ).  $J_j$  is the epitope - paratope matrix representing an antigen bound to a single antibody of the  $j^{th}$  type. It is a matrix composed of zeros and a single one in the row  $j$ , in the column representing the epitope that antibody of the  $j^{th}$  type binds to. The matrix  $B$  is an example of such a matrix.

As we will generally see  $J_j$  in formulae that involve sums over  $j$  it is convenient to drop the subscript to reduce the amount of notation and simply write  $J$ .

There are two functions that will be regularly dot producted with  $\mathbf{p}(j, M)$ :  $\mathbf{C}(j, M)$  and  $\mathbf{D}(j, M)$ . They are defined as:

$$\mathbf{C}(j, M) = (2c_j(1 - e(j, M)), \beta_j x_O, 0), \tag{7.15}$$

$$\mathbf{D}(j) = (2c_j(1 - e(j, M)), \gamma_j, 0), \tag{7.16}$$

where  $c_j$  is the variable representing concentration of the  $j^{th}$  antibody at the chips surface,  $\beta_j$  is a constant representing the quotient of the binding rate constant for the  $j^{th}$  antibody already bound to one antigen, binding to a second antigen over the binding rate constant for an unbound antigen. It may be assumed that this ratio is the same regardless of the type of antigen and antibody involved with either reaction, but there is some variation between these quotients for the best parameter fits produced so far (see Table 6.6.1). The constant  $\gamma_j$  represents the quotient of the binding rate constant for antibody bound to a single antigen in a complex including a second antigen with the correct epitope free over the binding rate constant for an unbound antigen.

These dot products are useful because  $k_{aj}\mathbf{C}(j, M) \cdot \mathbf{p}(j, M)$  and  $k_{aj}\mathbf{D}(j, M) \cdot \mathbf{p}(j, M)$  are the rates at which a single complex represented by  $M$  will undergo binding of the  $j^{th}$  type to a monomer,

and as a result become the complex represented by  $M + J$ , in the two cases where  $M$  contains exclusively 0s and 1s, and where  $M$  contains one or more 2s but not a 2 in the column of every epitope.

In more detail these dot products give us:

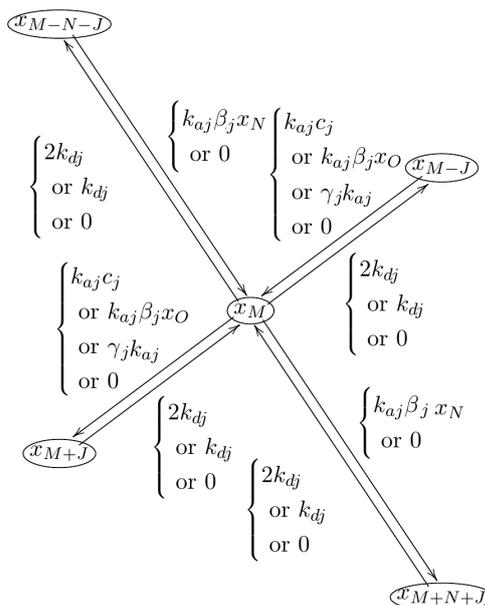
$$k_{aj}\mathbf{C}(j, M) \cdot \mathbf{p}(j, M) = \begin{cases} k_{aj}c_j & \text{if the } j^{\text{th}} \text{ antibody is not bound in the complex represented by} \\ & \text{the matrix } M, \text{ and the epitope it binds to is free,} \\ k_{aj}\beta_j x_O & \text{if the } j^{\text{th}} \text{ antibody is monovalently bound in the complex} \\ & \text{represented by the matrix } M, \\ 0 & \text{otherwise,} \end{cases} \quad (7.17)$$

$$k_{aj}\mathbf{D}(j, M) \cdot \mathbf{p}(j, M) = \begin{cases} 2k_{aj}c_j & \text{if the } j^{\text{th}} \text{ antibody is not bound in the complex represented by} \\ & \text{the matrix } M, \text{ and the epitope it binds to is free,} \\ k_{aj}\gamma_j & \text{if the } j^{\text{th}} \text{ antibody is monovalently bound in the complex} \\ & \text{represented by the matrix } M, \\ 0 & \text{otherwise,} \end{cases} \quad (7.18)$$

and these correspond with the reaction rates displayed in Fig. 7.4:

In Chapter ?? the bivalent effective rate constant model with spatial effects was fitted to data. For the antibodies investigated, the quotient  $\beta_j$  was around  $10^3 - 10^5 \text{M nm}^2 \text{pg}^{-1}$ . However, it varied with the density of antigen. This is because when the antibody is bound to a single antigen, it is constrained to a region within the experiment (see Fig. 7.3 (a)), and lower antigen density makes this spatial constraint more of a barrier to secondary binding.

It is less clear what kind of value  $\gamma_j$  might take. Figure 7.3 (b) shows how its free paratope is constrained to a very small region containing the epitope - however steric



**Figure 7.8:** Compartments and flows connected to  $x_M$ , where  $M$  has one or more entries taking the value of two;  $x_M$  represents the concentration of a specific complex containing two antigen and one or more antibody.

effects including the shape of the antigen and the locations of its epitopes may mean that this kind of binding happens at a much slower rate.

Due to the complexity and generality of this model a compartmental model diagram cannot be drawn, but the rates of flow for any complex may be divided into four categories. These are shown in Fig. 7.8, given a general complex represented by the matrix  $M$ , and the  $j^{th}$  antibody, there are four general categories of binding, binding of complexes to make the  $M^{th}$  complex, and associated unbinding (top left); binding of monomers to make the  $M^{th}$  complex, and associated unbinding (top right); binding of the complex with other complexes, and associated unbinding (bottom right); binding of the complex with single antigen and antibody, and associated unbinding (bottom left).

However, it is useful to present simplified compartmental diagrams for selected variables and discuss their flows before writing the model as a whole. These are shown in Fig.

7.9.

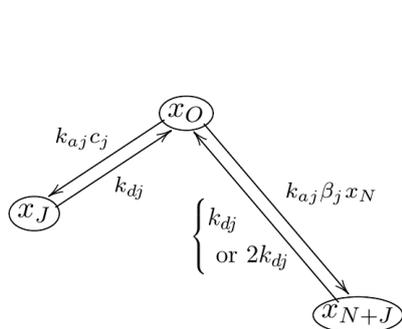
In this model, as illustrated by this diagram the unusual step of using matrices as subscripts for variables is taken. This is because it allows for a huge amount of information to be simplified into a single letter. In this particular model we are not just interested in what chemicals are in a complex, but how they are bound, and the epitope paratope matrix provides an easy way to do this.

### 7.2.2 Rates of change

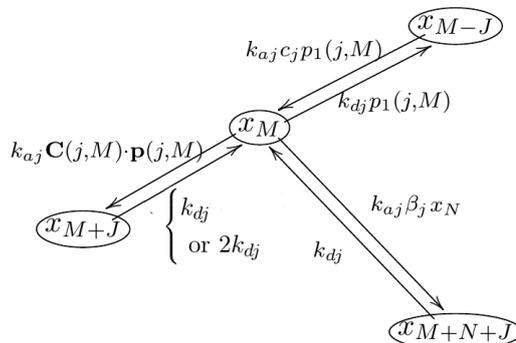
In the previous subsection, the paratope-epitope matrix was introduced, alongside related functions. In this section we will see how these can be used to simplify Fig. 7.8 into a set of diagrams that use these functions to determine rates of change, and then we will use these to write differential equations for the rates of change of the components of each complex. These diagrams, Fig. 7.9 are paired with illustrations showing examples of the complexes being considered in Fig. 7.10.

First we consider an antigen bound only to the reaction chip, with the paratope-epitope matrix  $O$  and concentration  $x_O$  (shown in 7.10).

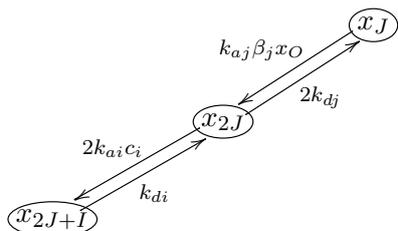
Figure 7.9 (a) shows the two different types of compartments and flows connected to  $x_O$ , the concentration of unbound antigen. Antigen that is unbound can bind through two different processes, either to a single antibody in flow or to a complex containing a free antibody. These two processes are illustrated by the two legs of the diagram, the shorter leg corresponds to a single antibody binding, and the longer corresponds to binding with a larger complex. In general these complexes dissociate back to this complex with the rate constant  $k_{dj}$ , but there is one special case, if the antigen binds to an antibody-antigen dimer (with the paratope-epitope matrix  $J$ ), then the antibody in the newly formed trimer (with the paratope-epitope matrix  $2J$ ) is bivalently bound and dissociation from either of



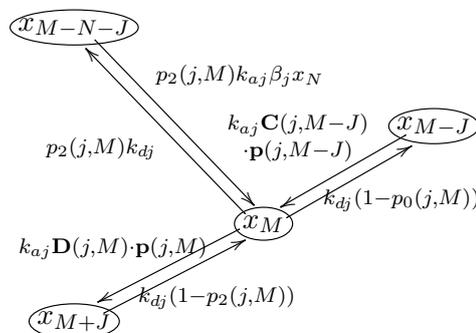
(a) Compartments and flows connected to  $x_O$ , the concentration of unbound antigen.



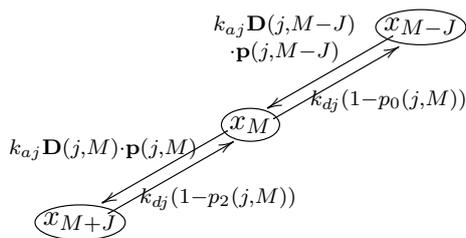
(b) Compartments and flows connected to  $x_M$ , where  $\text{Biv}(M) = 0$ .



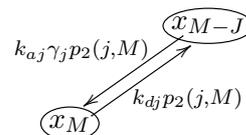
(c) Compartments and flows connected to  $x_{2J}$ .



(d) Compartments and flows connected to  $x_M$ , where  $\text{Biv}(M) = 1$ ,  $\text{Mono}(M) > 0$ .



(e) Compartments and flows connected to  $x_M$  where  $\text{Biv}(M) > 1$ .

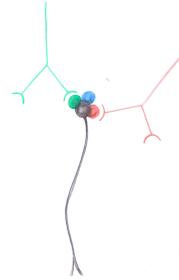


(f) Compartments and flows connected to  $x_M$ , where  $M$  has a two in every column.

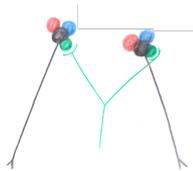
**Figure 7.9:** Simplified compartmental diagrams of generalised polyclonal model of antibody binding



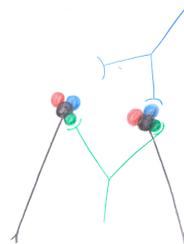
(a) A complex represented by the paratope-epitope matrix  $O$  - an unbound antigen.



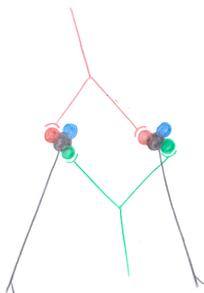
(b) A complex represented by the paratope-epitope matrix  $M$  where  $\text{Biv}(M) = 0$ ; a single antigen and one or more antibody.



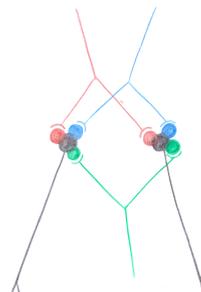
(c) A complex that could be represented by the paratope-epitope matrix  $M$  where  $\text{Biv}(M) = 2$ ,  $\text{Mono}(M) = 0$ ; two antigen linked by a single antibody.



(d) A complex that could be represented by the paratope-epitope matrix  $M$  where  $\text{Biv}(M) = 1$ ,  $\text{Mono}(M) > 0$ ; two antigen linked by a single antibody, with another antibody bound to one of the antigen.



(e) A complex that could be represented by the paratope-epitope matrix  $M$  where  $\text{Biv}(M) > 1$ ; two antigen linked by two antibody.



(f) A complex that could be represented by the paratope-epitope matrix  $M$  where  $M$  has a 2 in each column; two antigen linked by three antibody, leaving no epitopes on either available to bind with other antibody.

**Figure 7.10:** Illustrations of antibody- antigen complexes.

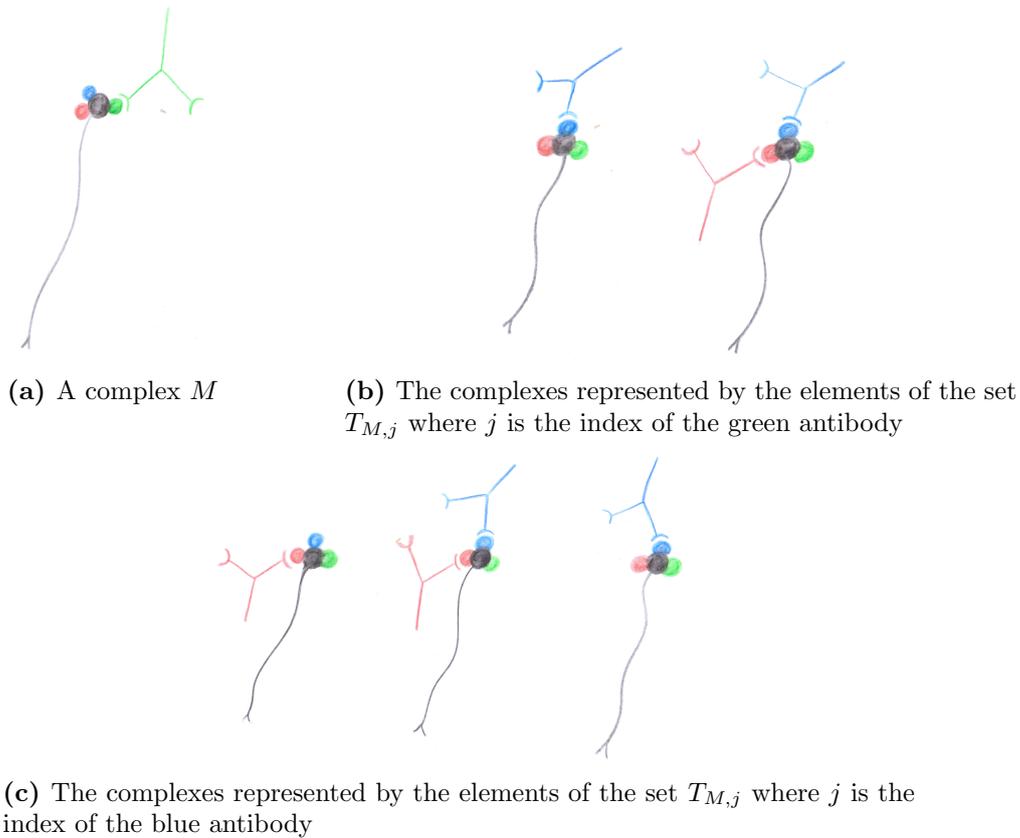
its paratope-epitope bonds will lead to one antigen being free. As a result the rate at which free antigen dissociate are formed by dissociation from this complex is  $2k_{dj}x_{2J}$ .

In a system with  $n$  different antibodies, there are  $n$  separate ways the antigen could form a complex with a single antibody and  $n$  complexes that could be formed; and the number of ways that the antigen could bind into a larger complex depends on which antibody compete for which epitopes. As a result the rate of change of this variable will include sums both over the  $n$  binding types, and over a set representing complexes that can be bound with. The rate of change of  $x_O$  is given by:

$$\begin{aligned} \frac{dx_O}{dt} = & \sum_{j=1}^n [-k_{a,j}x_Oc_j + k_{d,j}x_J] \\ & + \sum_{j=1}^n \sum_{N \in T_{O,j}} [-k_{a,j}\beta_j x_O x_N + k_{d,j}x_{N+J}] + \sum_{j=1}^n k_{dj}x_{2J} \end{aligned} \quad (7.19)$$

where the set  $T_{O,j}$  contains the matrices that represent non-monomeric complexes that can bind to free antigen through a binding reaction of the  $j^{th}$  type. As this contains  $J$  a final sum is added, to account for the difference between the rate free antigen is released by the complex represented by  $2J$  and other complexes with multiple antigen and only one bivalently bound antibody.

A more general version of this set  $T_{M,j}$  will be referred to frequently so it is useful to define it here:  $T_{M,j}$  is the set containing all epitope paratope matrices for the system with a single 1 on their  $j^{th}$  row if and only if  $M$  has no non-zero entries on its  $j^{th}$  row, that also do not have 2 as an entry, and that have a sum with  $M$  that does not have a 2 in any entry. The elements of this set represent the complexes that might bind with the complex  $M$  by through an antibody of type  $j$  binding to one of the free epitopes of  $M$ . This is illustrated in Fig. 7.11.



**Figure 7.11:** Illustrations to explain the set  $T_{M,j}$ .

The second case we shall consider is when  $M$  represents a complex containing one or more monovalently bound antibody and no bivalently bound antibody, as illustrated in Fig. 7.10 (b), that is where  $M$  has 1 as its maximum entry.

The corresponding diagram, Fig 7.9 (b), shows the three different types of compartments and flows connected to  $x_M$ .

The lower part of the diagram contains a longer and shorter leg, that represent the same binding processes as the longer and shorter leg of Fig. 7.9 (a). We see a single upper leg, representing this complex growing out of the addition of monomers to a smaller complex  $M - J$  (the complex has only one antigen, so it is too small to grow from dimers or larger complexes binding). On this upper leg the rates are multiplied by  $p_1(j, M)$ , the

multiplication by this function means that these flows are zero if the complex represented by  $M$  does not an antibody with the index  $j$ .

Another notable difference between this diagram and Fig. 7.9 (a) is that there is a dot product of vector valued functions in the flow between  $x_M$  and  $x_{M+J}$ . As we saw in Eq. (7.17) this gives us the correct rates for the various ways this complex may bind with free antibody and unbound antigen. Unlike the short upper leg in this diagram, there is no function like  $p_1(j, M)$  multiplying the dissociation rate constant, this is because the complex  $M + J$  represents is guaranteed to have an antibody of index  $j$ . Instead, like on the long leg of 7.9 (a) there are two possible dissociation rates, and this is for the same reason, if  $M = J$  then  $M + J = 2J$  which has a single antibody bound in two places, so it breaks down into the complexes  $O$  and  $J$  at twice the dissociation rate of a single antibody of index  $j$  bound to a single antigen.

The longer lower leg of the diagram is almost identical to that in Fig. 7.9 (a), except that the dissociation rate is exclusively  $k_{dj}$ . This is because there are no non-monomeric complexes that  $M$  can bind to that will produce a complex like  $2J$ .

As a result the rate of change of  $x_M$ , where  $\text{Biv}(M) = 0$  is:

$$\begin{aligned} \frac{dx_M}{dt} = & \sum_{j=1}^n p_1(j, M) [k_{aj}x_{M-J}c_j - k_{dj}x_M] \\ & + \sum_{j=1}^n [-k_{aj}x_M \mathbf{C}(j, M) \cdot \mathbf{p}(j, M) + k_{d,j}x_{M+J}] + \sum_{j=1}^n k_{dj}x_{2J} \\ & + \sum_{j=1}^n \sum_{N \in T_{M,j}} [-k_{a,j}\beta_j x_M x_N + k_{d,j}x_{M+N+J}]. \end{aligned} \quad (7.20)$$

The last two sums of this equation have the same role as the sums in Eq. (7.19), and correspond to the shorter and longer legs of the diagram Fig. 7.9. The first sum in this

equation represents the total rate at which smaller complexes become this complex. This sum was not in Eq. (7.19), as that was the rate of change of unbound antigen and there are no smaller complexes that can become a single antigen.

The third case we will consider is complexes with a single bivalently bound antibody, and no other antibody. Figure 7.9 (c) shows the two different types of compartments and flows connected to  $x_{2J}$ , as this the complex we are considering can not bind to any more antibody of the  $j^{th}$  type, it can't bind with other complexes (as we don't allow for complexes with more than two antigen), it has a very simple rate of change equation:

$$\frac{dx_{2J}}{dt} = [k_{aj}\beta_j x_0 x_J - 2k_{dj}x_{2J}] + \sum_{i \neq j} [-2k_{ai}c_i x_{2J} + k_{di}x_{2J+I}]. \quad (7.21)$$

The fourth case we shall consider is complexes with single bivalently bound antibody, but where not all epitopes of the antigen in the complex are attached to bivalently bound antibody, i.e. they are free or attached to monovalently bound antibody.

Figure 7.9 (d) shows the three different types of compartments and flows connected to  $x_M$ . This diagram appears quite similar to Fig. 7.9 (b) except that long arm goes down and right rather than left and up, and the flows on the shorter arms have changed.

The change in the long arm comes from the fact that  $x_M$  where  $M$  contains one 2 as an entry so represents a complex containing two antigen. This complex could be formed by multiple smaller complexes binding together, whereas the complexes represented by  $M$  in previous equations were too small to be formed this way, but could be formed into complexes like this. The flows for both association and dissociation on this arm are products of  $p_2(j, M)$ , this is because the complex  $X-M-N$  can only join another complex, through a binding reaction between the  $j^{th}$  antibody and its epitope if the complex  $M$  has a 2 in the that antibodies entry (see Fig. 7.5).

The short upper leg now has a dot product in its flow to  $x_M$ . that is because this complex is large enough so that it can be formed by the addition of free antigen as well as just antibody. As we have seen previously (Eq. (7.17), Fig. 7.4) we can use the dot product  $k_{aj}x_M\mathbf{C}(j, M) \cdot \mathbf{p}(j, M)$  to represent the total rate of change from both, except that as we are looking at flow from  $x_M - J$  we replace  $M$  with  $M - J$ .

The fact that  $\mathbf{D}(j, M)$  has replaced  $\mathbf{C}(j, M)$  on the lower arm, comes from the fact that the complex now includes multiple epitopes of each type - so antibody bound to one antigen can now bind to the second. Previously the short lower arm has never been able to have a flow of 0, that is because  $M + J$  by definition contains the  $j^{\text{th}}$  antibody which can dissociate. We only allow a single antibody of each type in a complex so if  $p_2(j, M) = 1$  no more antibody can bind to the complex, meaning that flow between  $x_M$  and  $x_{M+J}$  is by definition 0. As a result the rate at which a  $x_{M+J}$  dissociates to  $x_M$  is  $k_{dj}(1 - p_2(j, M))x_M$ .

The rate of change equations are:

$$\begin{aligned} \frac{dx_M}{dt} = & \sum_{j=1}^n p_2(j, M) \sum_{N \in T_{M-N-J, j}} [k_{a,j}\beta_j x_{M-N-J} x_N - k_{d,j} x_M] \\ & + \sum_{j=1}^n [k_{aj} x_{M-J} \mathbf{C}(j, M - J) \cdot \mathbf{p}(j, M - J) - k_{d,j}(1 - p_0(j, M))x_M] \\ & + \sum_{j=1}^n [-k_{aj} x_M \mathbf{D}(j, M) \cdot \mathbf{p}(j, M) + k_{d,j}(1 - p_2(j, M)) x_{M+J}]. \end{aligned} \quad (7.22)$$

The fifth and final case we will consider is that of complexes with multiple bivalently bound antigen. Figure 7.9 (e) shows a diagram like Fig. 7.9 (d), but simpler. as the complex has multiple bivalently bound antibody, it cannot bind with other complexes, but also there are no ways smaller complexes can bind together in a single reaction to form it, also as we

don't allow complexes with more than two antigen, it can't bind with other complexes. As a result there are no longer legs in the diagram.

There is only one change to the rates from Fig. 7.9 (d), the flow from  $X_{M-J}$  to  $X_M$ . The function  $\mathbf{C}(j, M - J)$  has been replaced by  $\mathbf{D}(j, M - J)$ . This is because any complex  $M - J$  has at least one bivalent antibody (see Eq. (7.17), Fig. 7.4).

As a result rate of change equations for complexes where  $\text{Biv}(M) > 2$  are:

$$\begin{aligned} \frac{dx_M}{dt} = & \sum_{j=1}^n [k_{aj}x_{M-J}\mathbf{D}(j, M - J) \cdot \mathbf{p}(j, M - J) - k_{d,j}(1 - p_0(j, M))x_M] \\ & + \sum_{j=1}^n [-k_{aj}x_M\mathbf{D}(j, M) \cdot \mathbf{p}(j, M) + k_{d,j}(1 - p_2(j, M))x_{M+J}]. \end{aligned} \quad (7.23)$$

Figure 7.9 (f) shows a special case of this, where because  $M$  has a 2 in every column the second sum in Eq. (7.23) is 0.

Equations (7.19)-(7.23) could be used by themselves as a state-space equation for a model. However in Chapter 5 the most successful model also incorporated the effects of isolated antigen and the varying concentration of antibody in flow through the ERC assumption.

Isolated antigen can only form much simpler complexes - those illustrated in 7.10 (a) and (b) as a result much simpler equations can be made for them. It is assumed that in the isolated antigen the same number of each epitope is available, and that the binding constants of antibody on a single antigen are not affected by each other. As a result if  $x_j$  is the amount of the  $j^{\text{th}}$  antibody bound to isolated antigen it has the rate of change:

$$\frac{dx_j}{dt} = k_{aj}c_j \left( R_m - \sum_{k \in e_j} x_k \right) - k_{dj}x_j, \quad (7.24)$$

for  $j$  between 1 and  $n$ , where  $e_j$  is the set of indices representing antibodies that bind to the same epitope as the  $j^{\text{th}}$  antibody, and  $R_m$  is the maximum density of analyte that can attach to isolated ligand to the sensor chip in  $\text{pg}^{-1} \text{nm}^2$ .

A transport equation for the unbound antibody may be derived by considering the concentration of antibody close to the chips surface:

$$\begin{aligned} \frac{dc_j}{dt} = & -k_{aj}c_j(R_m - x_j) + k_{dj}x_j - k_{aj}c_jx_O + k_{dj}x_J \\ & - \sum_{M \in \mathcal{M}_{j1}} [k_{aj}c_j + k_{dj}x_{M+J}] - \sum_{M \in \mathcal{M}_{j2}} [2k_{aj}c_j + k_{dj}x_{M+J}] \\ & + k_{mj}(I_j(t) - c_j), \end{aligned} \quad (7.25)$$

for  $j$  between 1 and  $n$ , where  $\mathcal{M}_{j1}$  is the set of matrices  $M$  representing complexes that a single unbound antibody of the  $j^{\text{th}}$  type could bind to, where  $\mathcal{M}_{j2}$  is the set of matrices  $M$  representing complexes that a two unbound antibody of the  $j^{\text{th}}$  type could bind to, and  $I_j(t)$  is the concentration of the  $j^{\text{th}}$  type of antibody in the polyclonal mixture being fed into the machine, and the the  $j^{\text{th}}$  component of the input function.

The variable  $c_j$  is assumed to be in quasi steady state meaning:

$$c_j = \frac{k_{dj} \left[ x_j + x_J + \sum_{M \in \mathcal{M}_{j1} \cup \mathcal{M}_{j2}} x_{M+J} \right] + k_{mj} I_j}{k_{aj} \left[ -(R_m - x_j) - x_O - \sum_{M \in \mathcal{M}_{j1}} x_M - 2 \sum_{M \in \mathcal{M}_{j2}} x_M \right] + k_{mj}}. \quad (7.26)$$

where  $\mathcal{M}_j$  is the set of matrices  $M$  representing complexes that an unbound antibody of the  $j^{\text{th}}$  type could bind to.

Alternately other assumptions about the concentration of antibody in contact with the ligand can be made. A Langmuir version of the model could be constructed where  $c_j = I_j(t)$ , or a Langmuir with transport version where Eq. (7.25) rather than Eq. (7.39) is part of the system equations.

The input function for this model is:

$$I(t) = \begin{cases} (C_1, \dots, C_n) & : t \in T_a \\ (0, \dots, 0) & : t \in T_d \end{cases} \quad (7.27)$$

The output equation for this model is  $\alpha$  multiplied by the total amount of antibody bound:

$$y = \alpha \sum_{j=1}^n \left[ x_j + (1 - p_0(j, m)) \sum_M x_M \right]. \quad (7.28)$$

The initial condition for this model is:

$$\begin{aligned} x_0 &= R_b \\ x_M &= 0 \text{ for all } M \\ x_j &= 0 \text{ for all } j, \end{aligned} \quad (7.29)$$

where  $R_b$  is the maximum density of analyte that can bind to ligand on the sensor chip, close enough to other ligand for the analyte to bind bivalently, in  $\text{pg}^{-1} \text{nm}^2$ . If it is assumed that all ligand are distributed so antibody can be bivalently bound  $R_b$  is replaced with  $R$ .

Together Eq. (7.19)-(7.24) and Eq. (7.26) -(7.29) make up the state, input and output equations for the model. For  $n$  antibody, its parameters are:

$$(k_{a1} \cdots k_{an}, k_{d1} \cdots k_{dn}, \beta_1 \cdots \beta_n, \gamma_1 \cdots \gamma_n, k_{m1} \cdots k_{mn}, C_1, \cdots C_n, R_m, R_b).$$

The BERCS model is a special case of this model, where the binding of a single antibody to a single antigen is the only reaction considered - as a result the epitope-paratope matrix has only a single element, and this element takes only the values 0, 1 and 2; as a result when the BERCS model is used on polyclonal data, it will be referred to as the *generalised polyclonal model for one antibody and one antigen* (GPM1X1).

The BERC and BL model models are also special cases that use these same epitope-paratope matrices. For both  $R_m = 0$  is assumed and in the case of the BL model Eq. (7.39) is replaced with  $c_j = I_j(t)$ .

There are two other special cases of this model that will be considered - two antibodies competing to bind for a single epitope, and two antibodies binding to two separate epitopes.

### 7.2.3 Two antibodies competing to bind for a single epitope

Previously we considered the chemical equation Eq. (7.7), but did not present a model for it. Now the generalised polyclonal model (GPM) has been developed we will use it to create a model specific to this interaction.

As this reaction now includes antibodies with two paratopes reacting with a single epitope on an antigen, the paratope-epitope matrix is now  $2 \times 1$ , so it can take the values:

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

This greatly simplifies the model, because there are now no complexes that contain both types of antibody. The rate of change of  $x_O$  (Eq. (7.19)) now becomes:

$$\begin{aligned} \frac{dx_O}{dt} &= \sum_{j=1}^n [-k_{a,j}x_Oc_j + k_{d,j}x_J] \\ &\quad + \sum_{j=1}^n [-k_{a,j}\beta x_Ox_J + k_{d,j}x_{2J}], \end{aligned} \quad (7.30)$$

where  $\beta = \beta_1 = \beta_2$ , allowing to further reduce the number of parameters. This reduction of complexity is because  $T_{O,j} = \{J\}$ , where  $J$  is the epitope - paratope matrix representing an antigen bound to a single antibody of the  $j^{\text{th}}$  type. For this model it will be a vector with a 1 in the  $j^{\text{th}}$  entry and a 0 in the other entry.

The only matrices  $M$  that contain only 0s and 1s are the  $J$  matrices,  $T_{J,j}$  is the empty set, and  $p_1(k, J)$  is 1 if  $k = j$  and 0 otherwise. As a result Eq. (7.20) simplifies to:

$$\frac{dx_J}{dt} = k_{a,j}x_Oc_j - k_{d,j}x_J - k_{a,j}x_J\beta x_O + k_{d,j}x_{2J}. \quad (7.31)$$

The only matrices  $M$  that contain one or more 2s are the  $2J$  matrices,  $T_{2J,j}$  is the empty set, and  $p_1(k, 2J) = 0$  for all  $k$ . As a result Eq. (7.22) simplifies to:

$$\frac{dx_{2J}}{dt} = k_{a,j}\beta x_Jx_O - k_{d,j}x_{2J}. \quad (7.32)$$

The paratope-epitope matrix does not have multiple columns so (7.23) is redundant in this case.

The formula for  $c_j$  is simplified, as isolated antigen haven't been considered:

$$c_j = \frac{k_{dj} \left[ x_J + \sum_{M \in \mathcal{M}_{j1}} x_{M+J} \right] + k_m I_j}{k_{aj} \left[ x_O + \sum_{M \in \mathcal{M}_{j1}} x_M \right] + k_m}. \quad (7.33)$$

where  $k_m = k_{m1} = k_{m2}$ . Equations (7.30) - (7.33) therefore are the system equations of the model. The output equation and initial condition are:

$$y = \alpha \sum_{j=1}^2 [x_j + x_J + x_{2J}], \quad (7.34)$$

$$x_M = 0 \text{ for all } M. \quad (7.35)$$

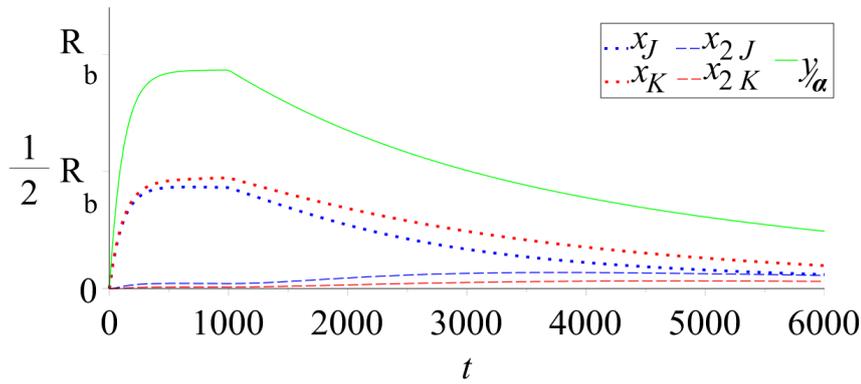
It has the input:

$$I(t) = \begin{cases} (C_1, C_2) & : t \in T_a \\ (0, 0) & : t \in T_d. \end{cases} \quad (7.36)$$

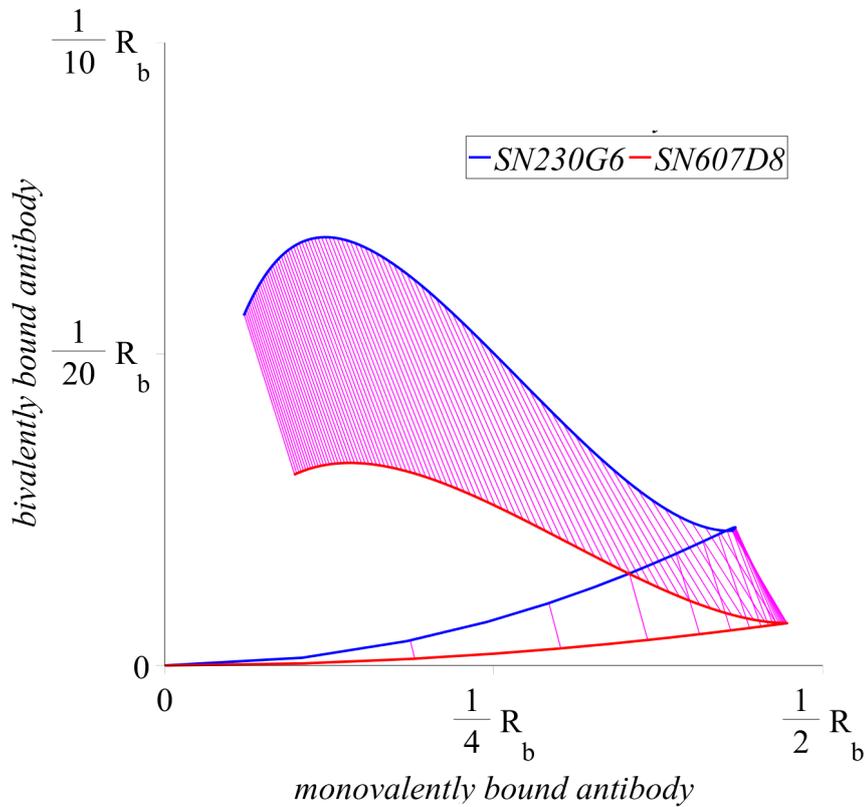
The parameters for this model shown in Table 7.2, together with (7.30) - (7.33) these define the *generalised polyclonal model for two antibodies competing over a single epitope* (GPM2X1).

Table 7.2 also shows the values that are used in Fig. 7.12 to show the models states and output for an input where  $T_a = [0, 1000)$  and  $T_b = [1000, 6000)$ .

Fig. 7.12 shows the result of a simulated experiment using this model with the parameters taken from those established for the SN230G6 and SN607D8 antibody in interaction with the A2 antigen shown in Table 6.6.1 averaged over both replicates for either experiments. These resulting parameters are shown in Table 7.2.



(a) A visualisation of the state and output variables, against time.



(b) A visualisation of the state variables, with magenta bars linking the concentrations of the two antibody in either configuration at 60 second intervals.

**Figure 7.12:** Graphs of the state and output variables for the GPE 2 antibody competitive model with parameters as shown in Table 7.2, where SN230G6 is antibody 1 and SN607D8 is antibody 2 and their epitope-paratope matrices for dimers are  $J$  and  $K$  respectively.

Parameter	$k_{a1}$	$k_{a2}$	$\beta$	$k_{d1}$	$k_{d2}$
	$M^{-1}s^{-1}$	$M^{-1}s^{-1}$	$Mnm^2pg^{-1}$	$s^{-1}$	$s^{-1}$
Value	4.3E+5	2.2E+6	3.8E+2	5.3E-3	1.3E-3

Parameter	$R_b$	$R_m$	$C_1$	$C_2$	$k_m$
	$pg\ nm^{-2}$	$pg\ nm^{-2}$	nM	nM	$pgnm^{-2}\ M^{-1}s^{-1}$
Value	5.0E+2	0	50	50	$10^{-4}$

**Table 7.2:** Values taken by parameters in Fig.7.12 and Fig. 7.13 .

The behaviour of this model is very similar to both the BL and the BERC models - except that there are now two species of antibody competing for the antigen. As a result Fig. 7.12 has qualitative similarities to Fig. 5.10 and Fig 5.13. Particularly high initial binding of monovalently bound antibody and antibody shifting to bivalent binding as well as dissociating in the dissociation phase.

If the dissociation phase is run for an arbitrarily large period of time the model predicts that the amount of antibody bound to isolated antigen and the amount of unbound antibody ready to bind become insignificant. This means the behaviour of the model becomes similar to that of the bivalent Langmuir model.

If  $j$  is the index of the antibody that dissociates slower, all of the elements of the models state vector except  $x_j$  and  $x_{2j}$ , become negligible. Similar to how Eq. (5.11) was derived,  $x_j$  can be put in a quasi steady state and the resulting equation can be used to eliminate  $x_j$  from Eq. (7.37). The resulting equation can be then expanded as a Taylor series about the origin to give:

$$\frac{dx_{2j}}{dt} = - \left( \frac{k_{dj}}{R_b \beta k_{a1} + k_{dj}} \right) x_{2j} + O(x_{2j}^2). \quad (7.37)$$

As the quadratic and higher order terms tend to zero faster than  $x_{2j}$ , when the dissociation phase is arbitrarily long, the model behaves increasingly like that given by the single

configuration Langmuir model, but with a modified dissociation constant.

#### 7.2.4 Two antibodies binding to two separate epitopes

A second chemical equation that was considered without a model being presented for it was Eq. (7.9). Now the generalised polyclonal model (GPM) has been developed we will also use it to create a model specific to that interaction.

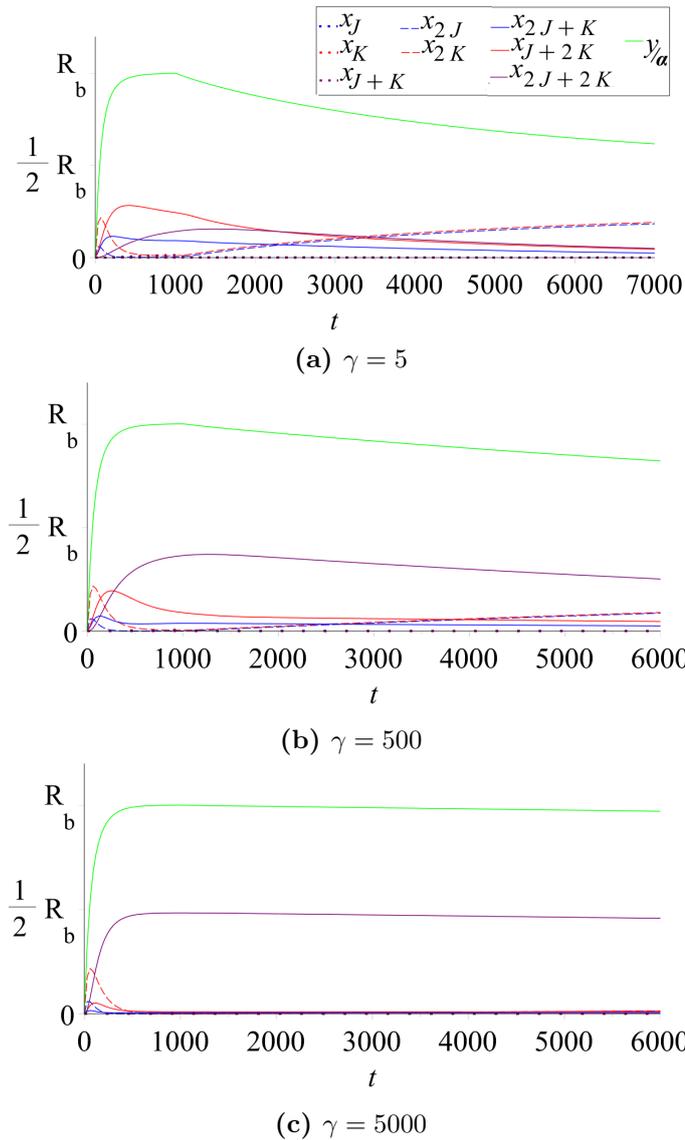
As this reaction now includes antibodies with two paratopes reacting with two separate epitopes on an antigen, the paratope-epitope matrix is now  $2 \times 2$ , so it can take the nine values:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}.$$

Unlike the previous form of the model, this does not lead to a simplification to its system equations. So its equations are (7.19) - (7.29). Its parameters are shown in Table 7.2 except  $\gamma$ . Together these define the *generalised polyclonal model for two antibodies competing over two epitopes* (GPM2X2).

Fig. 7.13 shows the result of a simulated experiment using this model with the parameters taken from those established for the SN230G6 and SN607D8 antibody in interaction with the A2 antigen shown in Table 6.6.1 averaged over both replicates for either experiment as well as the input function shown in Eq. (7.36) for  $T_a = [0, 1000)$  and  $T_b = [1000, 7000)$ .

This model introduces one extra parameter  $\gamma$ , which when multiplied by  $k_a j x_{2K+J}$  gives the rate at which antibody monovalently bound antibody of the  $j^{th}$  type in a tetramer will bind to the free epitope, creating a ring-like structure where all antibody are bound



**Figure 7.13:** Graphs of the state and output variables for the GPE 2 antibody non-competitive model, where SN230G6 is antibody  $j$  and SN607D8 is antibody  $k$ .

to all antigen (illustrated in Fig. 7.4) . As a result graphs are made for a variety of values for this parameter.

To determine the values the parameter should be considered, we consider Fig. 7.3. In this figure we see that the free paratope of an antibody monovalently bound into a complex

has a much smaller volume to move in relative to the its free target epitope in the same complex, than an antibody monovalently bound to a single antigen relative to another antigen on the same chip. We might expect that  $\gamma$  would have a value much greater than  $\beta$ , however it also has to incorporate factors like the flexibility of the proteins involved.

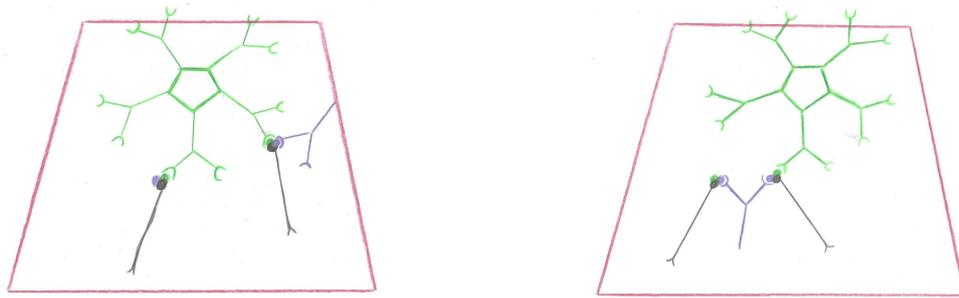
This problem is simplified by the fact that for the other parameters used (see Table 7.2) the states predicted by the model, within 6000s become insensitive to changes in  $\gamma$  if  $\gamma > 5000\text{Mpg}^{-1}\text{nm}^{-2}$ . As a result simulations are shown with  $\gamma$  from 5-5000.

Even for the lowest value of  $\gamma$  considered, the output remains much higher for much longer than for the competitive model 7.12. For higher values of  $\gamma$  we see the amount of dissociation becoming negligible, just like the graphs with extremely shallow dissociation curves in the high affinity plasma effluent data set (Fig. 4.14).

Interestingly another feature of this model could explain the higher rates of binding seen when monoclonals were mixed to simulate polyclonal serum. In Fig. 4.16 and Fig. 4.17 we see that mixtures of the two antibody have higher total amounts of binding when they are mixed. In all previous models the maximum amount of body that could bind has been either  $R$  or when we've divided the antigen into isolated and non isolated  $R_b + R_m$ ; whereas for the two antibody non competitive sub-case of the GPC model the maximum amount of antibody that can bind of either type is  $2R$  as two antibody can bind to each antigen; and if we were to consider the spacial sub-case of this model  $2R_b + 2R_m$ .

### 7.2.5 Adapting the GPM for interactions between antibodies of different size

Whilst the GPM was being fit to data two new hypotheses were considered. Both centred on the idea of the kinetic rates changing depending on what complexes are involved in a reaction. The motivation for this was the data presented in Section 4.9 particularly



(a) If the IgG antibody is bivalently bound the IgM molecule is large enough to bind bivalently.

(b) If the IgM antibody is bivalently bound the IgG molecule may not be large enough to bind bivalently, as a result the complex may have to go through a sequence of binding and unbinding reactions before both can be bivalently bound.

**Figure 7.14:** Comparison between the tetramers with IgG and IgM bivalently bound.

sensograms like Fig. 4.24 (a) where the initial rate of change in the dissociation phase is very slow, and doesn't appear to change much over the phase. Sensorgrams like this suggest that even antibody bound to isolated antigen is binding very strongly. One explanation for this could be that a larger antibody molecule is large enough to reach from one of these isolated ligands to another, for example an IgM pentamer.

This makes sense given the sizes of the reactants. The linker that binds the antigen to the chip is  $54\text{\AA}$ , and the distance between the two binding domains of the IgG antibody are expected to be at most  $70\text{\AA}$ , whereas on the IgM molecule they are normally separated by  $350\text{\AA}$  (Saber *et al.*, 2011). As a result IgM antibody would be able to bind bivalently to antigen in these experiments even if IgG could not (see Fig. 7.14).

The simplest way of modelling larger antibody was to allow bivalent binding to the isolated antigen for one antibody type. For a system where IgM is the antibody with the index 1

Eq. (7.24) is replaced with two equations:

$$\begin{aligned} \frac{dx_1}{dt} &= k_{a1}c_1 \left( R_m - 2x_{11} - \sum_{k \in e_1} x_k \right) - k_{d1}x_1 \\ &\quad - k_{a1}\beta_1x_1 \left( R_m - 2x_{11} - \sum_{k \in e_1} x_k \right) + 2k_{d1}x_2 \\ \frac{dx_{11}}{dt} &= k_{a1}\beta_1x_1 \left( R_m - 2x_{11} - \sum_{k \in e_1} x_k \right) - 2k_{d1}x_2, \end{aligned} \quad (7.38)$$

where  $e_j$  is the set of indices representing antibodies that bind to the same epitope as the  $j^{\text{th}}$  antibody. This change results in the change of the quasi-steady-state for the concentration of the antibody:

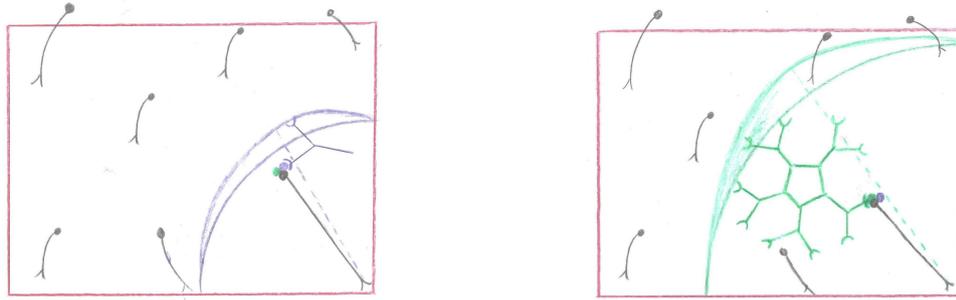
$$c_j = \frac{k_{dj} \left[ x_j + x_J + \sum_{M \in \mathcal{M}_{j1} \cup \mathcal{M}_{j2}} x_{M+J} \right] + k_{mj}I_j}{k_{aj} \left[ -(R_m - x_j - 2x_{jj}) - x_O - \sum_{M \in \mathcal{M}_{j1}} x_M - 2 \sum_{M \in \mathcal{M}_{j2}} x_M \right] + k_{mj}}, \quad (7.39)$$

for  $j = 1$ . The output equation for this model is the total concentration of antibody bound, multiplied by  $\alpha$ . Generally

$$y = \sum_{j=1}^n \alpha \left[ x_j + (1 - p_0(j, m)) \sum_M x_M \right] + \alpha x_{11}. \quad (7.40)$$

The initial condition for this model is:

$$\begin{aligned} x_0 &= R_b \\ x_M &= 0 \text{ for all } M \\ x_j &= 0 \text{ for all } j \\ x_{11} &= 0. \end{aligned} \quad (7.41)$$



(a) If the IgG antibody is bound to isolated ligand it is unable to bind bivalently.

(b) If the IgM antibody is bound to isolated ligand it may be able to bind bivalently.

**Figure 7.15:** Comparison between the IgG and IgM bound on isolated antigen.

Together Eq. (7.19)-(7.24) and Eq. (7.26) for indexes other than 1, Eq. (7.38)- (7.41) make up the system and output equations for the model. This model is referred to as the *generic polyclonal model for antibody of separate sizes (GPMSS)*. The form of it that will be applied to data, later in this thesis will be the GPMSS for two antibody and two antigen (2X2).

Fig. 7.15 shows why for this model we would not assume  $\gamma_i$  is the same for all antibodies.

## 7.3 Identifiability and indistinguishability

### 7.3.1 Langmuir model

The **two ligand** model has the parameter vector  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_1, C_2, R)$ .

This model was analysed with the output equation method and the output limit method.

Together these methods gave that if  $\bar{\mathbf{p}} \sim \mathbf{p}$  then  $\bar{\mathbf{p}}$  takes one of two forms:

$$\bar{\mathbf{p}} = \left( \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{C}_1, \bar{C}_2, R \right) \text{ or } \left( \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{C}_1, \bar{C}_2, R \right).$$

That is the model is structurally globally unidentifiable, because  $C_1, C_2, k_{a1}, k_{a2}$  are unidentifiable. The ORC method was used to investigate the models indistinguishable state-space solutions. The two indistinguishable parameterisations of the model are mapped by the system mapping to two indistinguishable state space solutions - which are reflections in the line  $x_1 = x_2$ .

Although the state-space of this model is different to all models previously discussed, the relationship between its indistinguishable solutions is the same as that for the two concentration Langmuir and ERC models. If  $\mathbf{x} \sim \bar{\mathbf{x}}$  then  $\bar{\mathbf{x}}(t, \mathbf{p}) = \lambda(\mathbf{x}(t, \mathbf{p}))$  where

$$\lambda(\mathbf{x}(t, \mathbf{p})) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \mathbf{x}(t, \mathbf{p}) \text{ or } \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{x}(t, \mathbf{p}).$$

The indistinguishability of this model was analysed with the generic Taylor series method. It was found that the model was structurally indistinguishable from all previously discussed models.

The **two concentration model** shares its parameter vector with the Langmuir two ligand model. That is, its parameter vector is  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_1, C_2, R)$ .

This model is simply structurally indistinguishable from the two configuration Langmuir model. That is for a generic input  $I$ , and the parameter vector  $\mathbf{p}$  the state-space solution is the same as the two configuration Langmuir model with the input  $I$  and the parameter vector  $\bar{\mathbf{p}}$ . Where  $\bar{\mathbf{p}} = (\frac{k_{a1}C_1}{C}, k_{d1}, \frac{k_{a2}C_2}{C}, k_{d2}, \bar{C}, R)$ . As a result the two models are also structurally indistinguishable.

For a parameterisations  $\bar{\mathbf{p}}$  of this model to have the same output as the parameterisation  $\mathbf{p}$  then:

$$\bar{\mathbf{p}} = \left( \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{C}_1, \bar{C}_2, R \right) \text{ or } \left( \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{C}_1, \bar{C}_2, R \right).$$

For two solutions to be indistinguishable they must have the same relationship as those of the two configuration Langmuir model, meaning that if  $\mathbf{x} \sim \bar{\mathbf{x}}$  then  $\bar{\mathbf{x}}(t, \mathbf{p}) = \lambda(\mathbf{x}(t, \mathbf{p}))$  where Eq. (7.3.1) defines  $\lambda$ .

### 7.3.2 Effective rate constant approximation model

The **two concentration model** has the parameter vector is:

$$\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_1, C_2, k_m, R).$$

This model is simply structurally indistinguishable from the two configuration ERC model. That is for a generic input  $I$ , and the parameter vector  $\mathbf{p}$  the state-space solution is the same as the two configuration Langmuir model with the input  $I$  and the parameter vector  $\bar{\mathbf{p}}$ , where  $\bar{\mathbf{p}} = (\frac{k_{a1}C_1}{C}, k_{d1}, \frac{k_{a2}C_2}{C}, k_{d2}, \bar{C}, k_m, R)$ . As a result the two models are also structurally indistinguishable.

For a parameterisations  $\bar{\mathbf{p}}$  of this model to have the same output as the parameterisation  $\mathbf{p}$  then:

$$\begin{aligned} \bar{\mathbf{p}} &= \left( \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{C}_1, \bar{C}_2, \frac{\bar{k}_{a1}}{k_{a1}} k_m, R \right) \\ &\text{or } \left( \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{C}_1, \bar{C}_2, \frac{\bar{k}_{a1}}{k_{a1}} k_m, R \right). \end{aligned} \tag{7.42}$$

For two solutions to be indistinguishable they must have the same relationship as those of

the two configuration ERC model, meaning that if  $\mathbf{x} \sim \bar{\mathbf{x}}$  then  $\bar{\mathbf{x}}(t, \mathbf{p}) = \lambda(\mathbf{x}(t, \mathbf{p}))$  where Eq. (7.3.1) defines  $\lambda$ .

### 7.3.3 General model of polyclonal antibody binding

Most forms of this model have too many variables, and non-linearities in its state-space equations to allow for an analysis of the identifiability of its parameters. However the state-space method is capable of analysing the structural identifiability of the model, and in the simplest case of the model the identifiability of the model's parameters can be analysed with the output limit method.

Regardless of the number of antibodies the model in its Langmuir and ERC form is simply structurally unidentifiable. This can be seen from observing that there are parameters that only appear in quotients in the state-space equations, with other parameters that appear in those quotients, i.e.  $k_{aj}$ ,  $C_j$  and  $k_{mj}$  only appear together. As a result regardless of the output these parameters will be structurally unidentifiable.

The structural identifiability of the Langmuir form of the model for two antibodies competing over a single epitope can be analysed with the output limit method. The model has the parameter vector  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_1, C_2, \beta, R)$ . For a second parameter vector  $\bar{\mathbf{p}}$ , the relationship  $\bar{\mathbf{p}} \sim \mathbf{p}$  holds if and only if:

$$\bar{\mathbf{p}} = \left( \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{C}_1, \bar{C}_2, \beta, R \right) \text{ or } \left( \bar{k}_{a2} \frac{\bar{C}_2}{C_2}, k_{d2}, \bar{k}_{a1} \frac{\bar{C}_1}{C_1}, k_{d1}, \bar{C}_1, \bar{C}_2, \beta, R \right).$$

## 7.4 Model selection and discussion

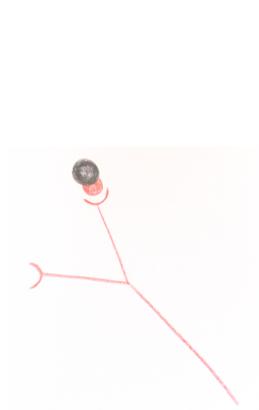
The GMP is a general model with special cases that allow any number of antibody binding with an antigen at any number of binding sites. We expect clinical data not to include the numbers of antibody binding or their epitope targets on the antigen. As a result, for accurate parameter estimation, we want to determine these details, or at least estimate them.

There are two approaches that can be taken to determining the number of clones of antibody in a sample, statistical, and experimental. In this thesis the concept of AIC has been repeatedly used to compare the relative value of models in fitting the data. As a result, one could simply fit a variety of special cases of the GPM models to each experiment and then select the one with the smallest AIC.

Alternatively each experiment could be replicated with changes in concentration and antigen density. In this way the repeatability of parameter estimates could be tested. If one model was to give consistently the lowest AIC for a specific experiment and its replicate, as well as similar estimates of rate constants, this would suggest that that from of the model was right for the sample.

However, it is also useful to bolster this theoretical work with additional experimental techniques. It is beyond the scope of this chapter, and even this thesis to discuss them fully but there are a number of techniques that could be used to assess the way antibodies in a sample interact with antigen.

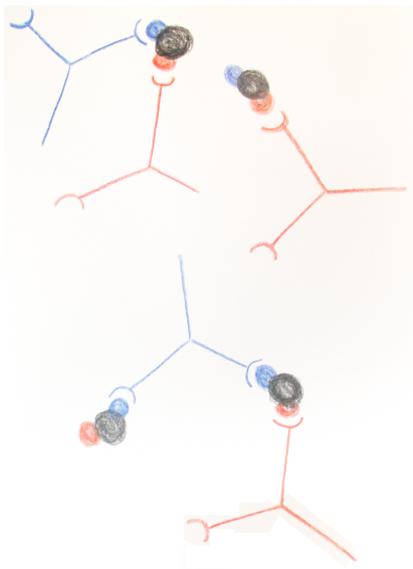
It is possible to measure the distribution of the sizes of protein complexes in a sample with a technique called size exclusion chromatography. This may be done for a patient serum sample. If antigen is added to the sample, change in the distribution of molecular weights will indicate the way in which antibody binds.



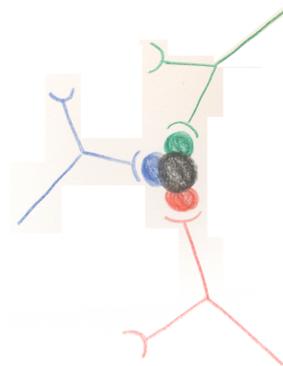
(a) A single IgG antibody binds to a single antigen.



(b) A single IgM antibody binds to an antigen



(c) Two IgG antibody bind to antigen in a number of configurations.



(d) three IgG antibody bind to antigen in a single configuration

**Figure 7.16:** Illustrations of the kinds of complex that can be made when various kinds of antigen and antibody molecules react

Some of the ways antibody can bind to antigen are shown in Fig. 7.16. If IgG bind to the antigen, the amount of protein with a molecular weight equal to that of the sum of a single IgG and antigen will increase, and the amount of protein with weight close to that of an IgG will decrease. Similarly, if IgM molecules will react, there will be a similar change in the distribution of the weights of protein complexes.

If there are two or more antibody that bind to the antigen at separate epitopes, much larger complexes are possible three are shown in Fig 7.16 (c). Chains of antibody and antigen can form, where many antibody are bound to matching epitopes on two antigen (or more if IgM molecules are involved), and many antigen have antibody at all of the epitopes able to react to the antibody.

If experimentally it was demonstrated that these complexes could occur, it would be a strong support of the hypotheses that underly the GPM. As well as being a useful tool for aiding in the choice of SPR model it would also be a useful tool for separating patients by their immune response. Particularly if these kinds of crosslinked interactions do create lower effective dissociation rate constants, and affinity is an indicator of immune response.

# Chapter 8

## Model and parameter fitting to SPR experiments with polyclonal antibody

Whilst SPR experiments with monoclonal antibody flowing across immobilised antigen were being conducted, and models for them developed and fitted, a similar process was being conducted for antibody taken from patients. As each patient is expected to have large numbers of hypermutated B lymphocytes producing the antibody clones in the sample it is expected to be *polyclonal*, rather than monoclonal (see Section 1.2).

This process of fitting models to sensorgrams made with polyclonal antibody began with sensorgrams provided by Dr A. Bentall (discussed in Section 4.5) which was used to compare the single concentration and the two concentration version of the ERC model. The ERC model was chosen for this because it had already been demonstrated in the monoclonal fitting to give superior fits to the Langmuir model (see Section 6.4.1).

When experiments with plasma effluent and serum samples were initially conducted, they were fitted with the two concentration Langmuir and ERC models.

Whilst this fitting process continued models that included more complex features of antibody binding were developed for monoclonal antibody (such as bivalence and the effects of

spatially isolated ligand, see Subsections 5.1.5 - 5.1.6). These improved models were shown to give more consistent estimates of kinetic constants and better fits (Subsections 6.4.2-6.6). The success of these models raised the question of whether the rate constant estimates made so far for polyclonal antibody were influenced by the same kinds of errors seen in work with equivalent models on monoclonal antibody (see Section 6.4.1). This problem was heightened by the fact that the data that had been originally used for polyclonal experiments did not include replicates of experiments (Section 4.5) so distributions of estimates of kinetic constants could not be compared in the same way as they were for the monoclonal experiments (see Fig. 6.6, Fig. 6.10, Fig. 6.14, Fig. 6.20(b), Fig.6.23(b)).

Additionally, further experiments were conducted and they showed very slow rates of change of signal in the dissociation phase, which lead to the Proteon manager software giving estimates of  $K_D$  from 8.7E-12 to 3.8E-24 (Subsection 4.7), vastly higher than what is thought possible for antibody binding.

These sensorgrams motivated the development of a more complex model, which combined the assumptions of the models that best explained the behaviour of the monoclonal antibody with a new concept - that of complexes containing multiple-antigen and multiple-antibody - to explain the new behaviour observed in the polyclonal experiments, the GPM (see Subsection 7.2).

To validate these assumptions further experiments were conducted with mixtures of monoclonal antibody being used to simulate the polyclonal mixtures from patient samples (Subsection 4.8). These experiments served multiple purposes. Firstly they are very different in design to any of the previous monoclonal experiments. If they produce similar estimates of the kinetic constants of the antibody-antigen interaction this would help validate the model previously used for the monoclonal experiments as well as the new model for the polyclonal experiments.

Secondly, they could validate whether these low rates of change of output in the dissociation phase could be explained well by the interactions of the antibody themselves - rather than some other molecule in the samples. Thirdly if these lower rates of change of output were observed this would suggest the presence of multi-antigen, multiple antibody complexes (as pictured in Fig. 7.10).

This result would be significant for pharmaceutical reasons - particularly suggesting that there are ways to employ monoclonal antibodies that create complexes from which antibody dissociate at a much slower rate.

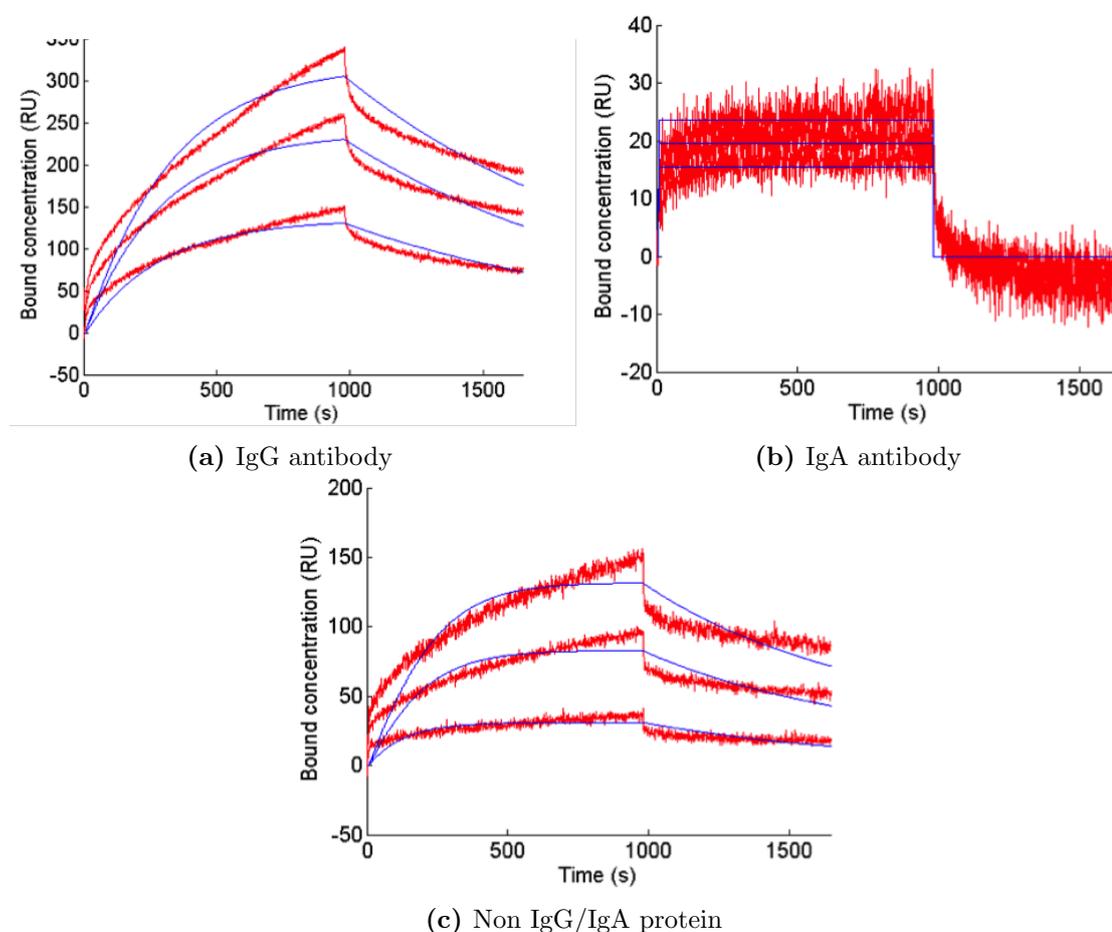
Finally, sensorgrams taken from experiments with sensitised patients and antigen they were sensitised to were analysed with the GPM model.

## 8.1 Fitting models to sensorgrams from purified samples of IgG, IgA and the remaining proteins

The first polyclonal data set was presented in Section 4.5, it features three samples of protein taken from a single patient. IgG and IgA were purified from the initial sample and used as SPR analytes, and the remaining protein, which is expected to include IgM was used as a third analyte.

This section presents fits with an established model - the single concentration ERC model (ERC1, Eq. (5.6)) and a model that had recently been published (Evans *et al.*, 2013), the two concentration version of the ERC model (ERC2C, Eq. (7.6)).

The fit of the ERC1 model to sensorgrams from this data set is presented in Fig 8.1. It is a poor visual match for the data, particularly in parts (a) and (c). In these the bound concentration in the early part of the association phase is underestimated, the mid part is overshot, and the last part it tends to a value much lower than seen in the experiments.



**Figure 8.1:** Fits of the single concentration ERC model (blue lines) to data from parallel experiments (red lines) with trisaccharide linker densities of 0.3mg/ml, 0.2mg/ml and 0.1mg/ml.

In the early part of the dissociation phase, the bound concentration from the experiments reduces rapidly, but the model output does not, and predicts higher than observed bound concentrations in each channel, and less than observed in the final part of the dissociation phase.

The ERC1 model gives a better visual match for the data in part (b). It appears that the experiments reached saturation in each channel at a much lower bound concentration than the experiments in parts (a) and (c), and as a result has a low signal to noise ratio. In the early part of the association phase the model bound concentration increases

	IgG	IgA	non-IgG/A protein
$k_{a1}$	$2.7 \times 10^{-5}(0.01)$	$3.1 \times 10^{-5}(\text{NA})$	$2.1 \times 10^4(\text{NA})$
$k_{d1}$	$8.8 \times 10^{-4}(\text{NA})$	$1.1 \times 10^{-4}(\text{NA})$	$2.2 \times 10^5(0.03)$
$K_{M1}$	$9.5 \times 10^{-3}(0.48)$	$7.1 \times 10^{-6}(\text{NA})$	$3.6 \times 10^{-6}(0.01)$
$K_{M2}$	$2.0 \times 10^{-3}(\text{NA})$	$1.2 \times 10^{-2}(0.08)$	$2.1 \times 10^{-6}(0.01)$
$K_{M3}$	$1.5 \times 10^{-4}(\text{NA})$	$6.3 \times 10^{-6}(0.06)$	$1.3 \times 10^{-6}(0.02)$
$R_1$	$4.1 \times 10^{-1}(0.01)$	$8.8 \times 10^{-1}(0.09)$	$1.4 \times 10^{-1}(0.06)$
$R_2$	$3.1 \times 10^{-1}(0.02)$	$7.4 \times 10^{-1}(0.09)$	$9.2 \times 10^{-2}(0.02)$
$R_3$	$1.8 \times 10^{-1}(0.01)$	$6.0 \times 10^{-1}(0.09)$	$3.4 \times 10^{-2}(0.06)$
RSS	$7.4 \times 10^5$	$4.2 \times 10^4$	$6.4 \times 10^4$

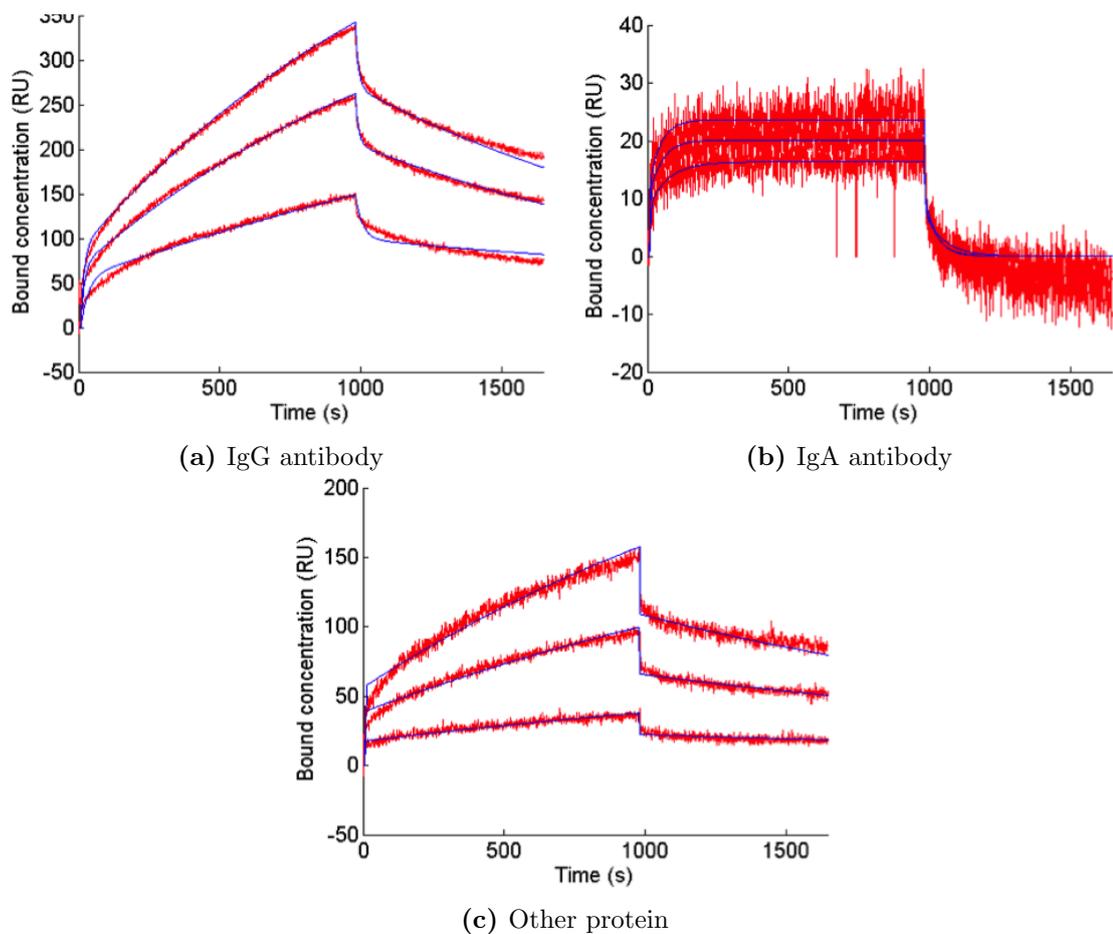
**Table 8.1:** Parameter estimates from the single concentration ERC model with SDLNs in brackets

rapidly until it reaches the saturation point where this increase abruptly stops; in the experiment a similar increase is seen except that the rate of increase decreases as the bound concentration approaches saturation. In the early part of the dissociation phase the model bound concentration falls to zero much faster than the experimental bound concentration.

Estimated parameter values are presented in 8.1 along with a corresponding standard deviation of their natural logarithm (SDLN) in brackets for each of these parameters. FACSIMILE works in terms of internal parameters that are the natural logarithm of the parameters used here and presents these as a measure of confidence in its estimates.

Some parameters are given NA rather than a SDLN value. This signifies that the parameter is not well determined by the data, and has an SDLN of greater than 0.2. If this is the case the parameter value given was chosen to minimise residual sum of squares (RSS).

The fit of the ERC2C model to the data is presented in Fig 8.2. Whilst it offers a dramatically better fit it also shows some disparities similar, to the homogeneous model. The model fits much closer to the data in parts (a) and (c). In these two it accurately estimates the bound concentration throughout the association phase, except in the very early part where it overshoots. The experimental bound concentration in the dissociation



**Figure 8.2:** Fits of the two concentration ERC model (blue lines) to data from parallel experiments (red lines) with trisaccharide linker densities of 0.3mg/ml, 0.2mg/ml and 0.1mg/ml.

phase is reproduced accurately in part (c), but in part (a) the model appears to be tending asymptotically towards different values than the experiments.

In part (b) where there were saturation problems, this model gives a very similar estimate to that provided by the homogeneous model. This model, however, predicts the slowing of the increase in bound concentration as it reaches saturation in the association phase, and again as it reduces to zero in the dissociation phase.

	IgG	IgA,	non-IgG/A protein
$k_{a1}$	$5.2 \times 10^{-5}$ (0.06)	$1.6 \times 10^{-4}$ (NA)	$3.7 \times 10^1$ (0.09)
$k_{a2}$	$4.1 \times 10^{-5}$ (0.07)	$9.4 \times 10^{-6}$ (NA)	$4.5 \times 10^{-3}$ (0.03)
$k_{d1}$	$7.8 \times 10^{-4}$ (0.01)	$6.0 \times 10^{-1}$ (NA)	$8.4 \times 10^{-1}$ (0.04)
$k_{d2}$	$5.4 \times 10^{-2}$ (0.01)	$1.9 \times 10^{-2}$ (0.06)	$6.7 \times 10^2$ (0.04)
$K_{M1}$	$1.6 \times 10^{-4}$ (0.07)	$2.7 \times 10^{-4}$ (0.31)	$1.7 \times 10^0$ (0.02)
$K_{M2}$	$1.5 \times 10^{-4}$ (0.07)	$1.7 \times 10^{-4}$ (0.27)	$2.1 \times 10^{-3}$ (0.03)
$K_{M3}$	$8.9 \times 10^{-4}$ (0.06)	$9.7 \times 10^{-5}$ (0.23)	$1.0 \times 10^{-5}$ (0.02)
$R_1$	$1.4 \times 10^0$ (0.02)	$1.3 \times 10^0$ (0.08)	$2.3 \times 10^{-1}$ (0.01)
$R_2$	$1.0 \times 10^0$ (0.02)	$1.1 \times 10^0$ (0.08)	$1.2 \times 10^{-1}$ (0.01)
$R_3$	$5.3 \times 10^{-1}$ (0.02)	$8.8 \times 10^{-1}$ (0.08)	$4.1 \times 10^{-2}$ (0.08)
$C_1$	$8.0 \times 10^0$ (0.06)	$2.2 \times 10^1$ (0.03)	$6.9 \times 10^0$ (0.02)
$C_2$	$9.2 \times 10^1$ (-)	$7.8 \times 10^1$ (-)	$1.0 \times 10^2$ (-)
RSS	$2.5 \times 10^3$	$3.5 \times 10^4$	$1.0 \times 10^4$

**Table 8.2:** Parameter estimates from the single concentration ERC model with SDLNs in brackets

Estimated parameters are presented in Table ??.

The differences in the quality of the fit produced by these models is reflected in their RSS values.

### 8.1.1 Comparison

The simulated output of the single concentration model captures some of the dynamics of the data (Fig 8.1), although it was unable to match the rapid changes in bound concentration at the start of either phase for the IgG experiment Fig 8.1 (a) or the non-IgG/IgA experiment Fig 8.1 (c), and overestimated the speed that that the experiment would reach saturation in part Fig 8.1 (b).

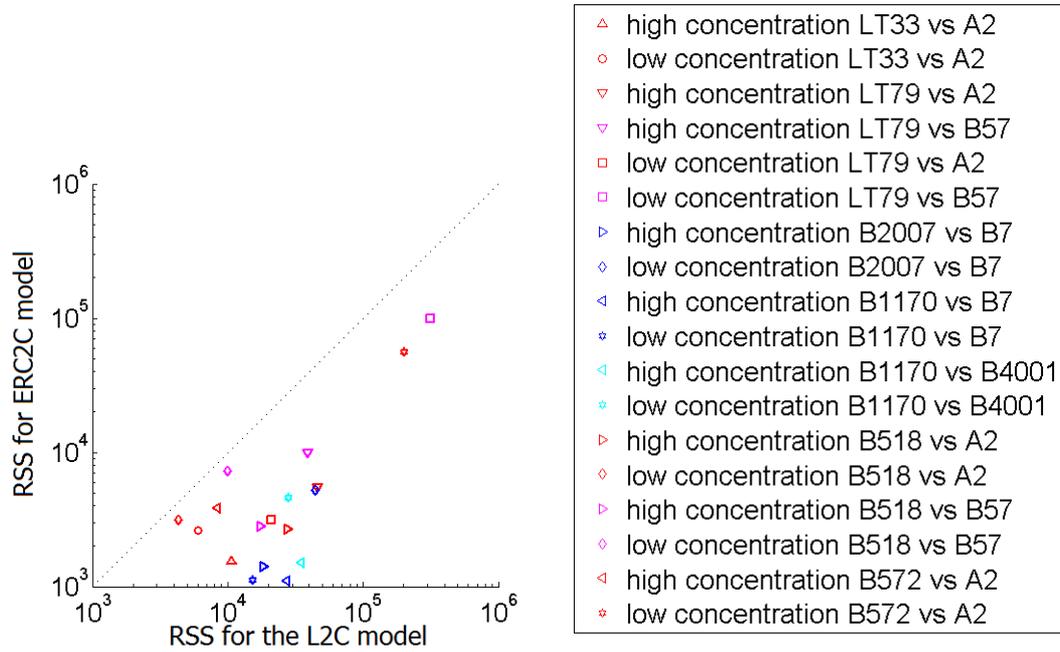
The two concentration model (Fig 8.2) dramatically improved on this. It was better able to match the rapid changes in bound concentration at the start of the phases, although it overestimated this rapid change in the IgG experiment Fig 8.2 (a) or the non-IgG/IgA experiment Fig 8.2 (c). However fits with it were vastly inferior to those achieved with monoclonal antibody in Section 6.6, indicating that the ERC2 model may still be missing some vital components of polyclonal antibody-antigen binding.

The binding affinity estimates for both models are separated by orders of magnitude. This is to be expected, ERC1 assumes that all of the analyte in the experiment binds with the same affinity, as a result it will not pick up a small amount of analyte that binds with higher affinity than the rest. This is a problem even if it is used to fit data from experiment with antibody of one isotype, because the antibody will be polyclonal and contain a variety of affinities. The ERC2C model assumes that there are two binding types within the analyte and as a result is better able to deal with experiments with polyclonal data. This is reflected by the difference in RSS between the models.

Although visually the best fits were not as good as those presented in Chapter 6, they were a considerable improvement on those from the single concentration model. This can be seen from the residual sum of squares of the models. The ERC2C model has a RSS of 0.05% of the RSS for the ERC1 model for the IgG experiment, 83% for the IgA experiment, and 16% for the non IgG/IgA experiment. This suggests that the ERC2C model is an improvement as a model for each experiment, though this is much more prominent for IgG. This improvement is also reflected in the reduction in RSS for the model summed across all experiments; the total RSS for the heterogeneous model is 6% of the total RSS for the homogeneous model.

These results were peer reviewed and published in H.A.J.Moyse *et al.* (2012), a copy of this paper is included in Appendix B.2.

The result that the ERC2C version of the model is significantly better for modelling antibody from a single class, and the kinetic constants of antibody from a single class differ so widely suggests there may be complications in creating models for plasma effluent and patient serum. If each antibody subclass's binding was well modelled by a single concentration of antibody then a model that only considers a mixture of antibody subclasses may be sufficient to explain the dynamics of binding in these complex mixtures.



**Figure 8.3:** Scatter graph of the RSS for the Fits of the the ERC2C and L2C models to the sensorgrams from Section 4.6

## 8.2 Plasma effluent and serum samples

The two concentration from of the ERC model was also tested against the two concentration form of the Langmuir model, to establish whether the ERC effect was as significant to modelling patient antibody interactions as those for monoclonal antibody.

A scatter graph of the RSS of these fits is shown in Fig. 8.3, the two configuration from of the ERC model shows great improvements in RSS when compared to the two configuration Langmuir model. This motivated the inclusion of the ERC assumption in the GPM model (see (7.26)).

	LT33 vs A2	LT33 vs A2	LT79 vs A2	LT79 vs B57	LT79 vs A2	LT79 vs B57
	Hc	Lc	Hc	Hc	Lc	Lc
$k_{a1}(M^{-1}s^{-1})$	1.1E-03	1.6E-04	1.5E-03	7.5E-04	4.4E-04	3.6E-05
$k_{a2}(M^{-1}s^{-1})$	3.2E-05*	7.7E-09*	2.5E-08*	5.7E-05*	1.7E-05*	3.2E-05*
$k_{d1}(s^{-1})$	1.2E-04	1.8E-04	1.9E-04	2.9E-04	2.3E-04	2.7E-04
$k_{d2}(s^{-1})$	6.9E-03*	1.8E-04*	9.3E-05*	1.6E-02*	1.1E-02*	6.9E-03*
$R_1(\text{pgnm}^{-2})$	6.3E-01	5.3E-01	6.1E-01	1.3E-01	3.5E-01	2.0E-01
$R_2(\text{pgnm}^{-2})$	5.9E-01	4.8E-01	5.4E-01	1.0E-01	3.1E-01	1.9E-01
$R_3(\text{pgnm}^{-2})$	5.4E-01	4.1E-01	4.8E-01	9.6E-02	2.7E-01	2.6E-01

	B2007 vs B7	B2007 vs B7	B1170 vs B7	B1170 vs B7	B1170 vs B4001	B1170 vs B4001
	Hc	Lc	Hc	Lc	Hc	Lc
$k_{a1}(M^{-1}s^{-1})$	6.9E-04	1.4E-04	2.0E-03	4.9E-04	6.1E-04	1.8E-04
$k_{a2}(M^{-1}s^{-1})$	3.9E+10*	3.9E+10*	3.9E+10*	3.9E+10*	3.9E+10*	3.9E+10*
$k_{d1}(s^{-1})$	2.5E-04	4.3E-04	8.1E-05	1.5E-04	1.3E-04	2.0E-04
$k_{d2}(s^{-1})$	2.1E+01*	2.1E+01*	2.1E+01*	2.1E+01*	2.1E+01*	2.1E+01*
$R_1(\text{pgnm}^{-2})$	4.4E-01	3.0E-01	7.6E-01	5.6E-01	7.7E-01	3.7E-01
$R_2(\text{pgnm}^{-2})$	3.9E-01	2.8E-01	6.8E-01	4.8E-01	6.9E-01	3.4E-01
$R_3(\text{pgnm}^{-2})$	3.8E-01	2.3E-01	5.8E-01	4.6E-01	6.5E-01	3.1E-01

	B518 vs A2	B518 vs A2	B518 vs B57	B518 vs B57	B572 vs A2	B572 vs A2
	Hc	Lc	Hc	Lc	Hc	Lc
$k_{a1}(M^{-1}s^{-1})$	6.1E-04	1.2E-04	8.7E-04	1.9E-04	1.7E-04	6.5E-05*
$k_{a2}(M^{-1}s^{-1})$	4.7E-04*	5.0E-04*	5.0E-04*	4.2E-05*	5.0E-04*	4.2E-05*
$k_{d1}(s^{-1})$	1.7E-04	2.2E-04	2.8E-04	2.7E-04	2.4E-04	5.6E-04
$k_{d2}(s^{-1})$	1.2E-02*	7.4E-03*	7.4E-03*	1.3E-02*	7.4E-03*	1.3E-02*
$R_1(\text{pgnm}^{-2})$	9.1E-01	6.2E-01	4.0E-01	2.3E-01	5.3E-01	3.7E+01
$R_2(\text{pgnm}^{-2})$	7.9E-01	5.8E-01	3.6E-01	2.2E-01	3.7E-01	2.9E+01
$R_3(\text{pgnm}^{-2})$	7.6E-01	5.3E-01	3.2E-01	2.0E-01	4.7E-01	2.1E+01

(a) L2C model parameter estimates.

	LT33 vs A2	LT33 vs A2	LT79 vs A2	LT79 vs A2	LT79 vs B57	LT79 vs B57
	Hc	Lc	Hc	Hc	Lc	Lc
$k_{a1}(M^{-1}s^{-1})$	2.3E-02*	1.7E-04	1.1E-02	1.1E+17*	4.4E+00*	6.5E-02
$k_{a2}(M^{-1}s^{-1})$	6.0E-04	1.3E+02*	2.1E-03	5.9E-04	2.6E-04	2.9E-03
$k_{d1}(s^{-1})$	1.2E-04	5.0E-04	1.7E-04	5.1E-04*	8.4E-03	5.8E-04
$k_{d2}(s^{-1})$	6.9E-03	8.9E-06*	1.5E-02	2.7E-03	1.2E-04	1.1E-01
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	6.5E-01	3.4E-01	7.0E-01	1.3E-01	3.9E-01	5.5E-02
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	6.1E-01	2.9E-01	6.2E-01	1.0E-01	3.7E-01	2.5E-02
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	5.5E-01	7.1E-01	6.0E-01	8.3E-02	3.5E-01	2.2E-02
$R_1(\text{pgnm}^{-2})$	8.5E-03*	7.7E-04	2.5E-03	2.0E+04	1.1E+01*	4.4E-05
$R_2(\text{pgnm}^{-2})$	2.0E-02*	8.5E-04	8.0E+04*	3.1E+04	1.6E+01*	5.0E-05
$R_3(\text{pgnm}^{-2})$	3.4E-02*	1.2E-04	1.0E+00*	4.9E+04*	3.3E+01*	5.5E-05

	B2007 vs B7	B2007 vs B7	B1170 vs B7	B1170 vs B7	B1170 vs B4001	B1170 vs B4001
	Hc	Lc	Hc	Lc	Hc	Lc
$k_{a1}(M^{-1}s^{-1})$	2.3E-02*	2.1E+07*	1.7E+00	3.3E-02	1.1E-01	1.7E-02
$k_{a2}(M^{-1}s^{-1})$	4.6E-04	1.3E-04	2.1E-03	2.4E-04	6.5E-04	5.8E+04
$k_{d1}(s^{-1})$	2.1E-04	1.4E-05*	9.1E-05	8.9E-05	8.9E-05	3.1E-01
$k_{d2}(s^{-1})$	8.6E-03	9.6E-04	1.3E-02	7.7E-03	1.4E-02	5.2E+03
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.6E-01	2.2E-01	7.9E-01	6.0E-01	8.2E-01	3.3E-01
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.1E-01	2.2E-01	7.0E-01	5.0E-01	7.2E-01	2.1E-01
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.0E-01	3.9E-01	6.0E-01	4.8E-01	6.4E-01	1.2E-01
$R_1(\text{pgnm}^{-2})$	8.2E-03*	4.4E-03*	3.6E-01	3.8E-02	5.3E-02*	6.6E-05
$R_2(\text{pgnm}^{-2})$	2.4E-02*	4.4E-03*	1.4E+00	2.2E-01*	2.0E-01	9.0E-05
$R_3(\text{pgnm}^{-2})$	5.0E-02*	1.2E-04	6.4E+01*	5.7E+01*	9.7E+01*	1.0E-04

	B518 vs A2	B518 vs A2	B518 vs B57	B518 vs B57	B572 vs A2	B572 vs A2
	Hc	Lc	Hc	Lc	Hc	Lc
$k_{a1}(M^{-1}s^{-1})$	5.5E-04	3.3E-13*	6.0E-04	1.3E-04	4.5E+01*	7.9E-16
$k_{a2}(M^{-1}s^{-1})$	4.7E-04	1.3E-04	5.0E-04	4.2E-05	2.2E-04	1.1E-07
$k_{d1}(s^{-1})$	1.4E-04	1.1E-02*	2.3E-04	2.3E-04	2.1E+08*	7.6E-03
$k_{d2}(s^{-1})$	1.2E-02	1.9E-04	7.4E-03	1.3E-02	2.4E-04	1.2E-03
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	9.6E-01	5.8E-01	4.3E-01	2.8E-01	5.0E-01	7.4E+02
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	8.3E-01	5.3E-01	4.0E-01	2.8E-01	3.1E-01	3.3E+02
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	8.1E-01	4.8E-01	3.8E-01	2.7E-01	4.2E-01	1.5E+02
$R_1(\text{pgnm}^{-2})$	5.4E-04	6.3E+06*	4.2E-04	1.0E+04*	3.2E-04	2.0E-05
$R_2(\text{pgnm}^{-2})$	2.8E-03*	5.8E+02*	1.6E-03	1.1E+04*	1.9E-02	2.8E-05
$R_3(\text{pgnm}^{-2})$	1.8E-02*	1.4E+05*	7.9E+00*	1.0E+04*	4.5E-04	4.6E-03

(b) ERC2C model parameter estimates.

Table 8.3: Parameter estimates.

### 8.3 Mixtures of monoclonal antibody

After the plasma effluent and serum samples had been fit, the high-affinity data presented in Section 4.7 was obtained. Affinity estimates ranging from  $8.7\text{E-}12\text{M}$  to  $3.8\text{E-}24\text{M}$  prompted the development of the models designed to explain the slower dissociation of these analytes by the creation of larger, longer lasting multi-antibody multi-antigen complexes (Section 7.2).

These hypothetical complexes explained the slower dissociation of an analyte because when one antibody in the complex dissociates from an antigen within the complex, they are held close together by the rest of the complex, and as a result are much more likely to re-associate (this was presented visually in Fig. 7.3).

As the existence of these complexes and their effect on SPR sensorgrams was a new hypothesis, it was necessary to conduct further experiments to test this it. These were the mixed monoclonal experiments presented in Section 4.8.

These experiments consisted of mixtures in varying concentrations of SN607D8 and SN230G6 being passed over the A2 antigen- an antigen where both antibodies bind to separate epitopes. The goal was to measure differences in the rate of total antibody dissociation in the dissociation phase between the various mixtures.

In the high-affinity sensorgram data set, the dissociation phases had been characterised by extremely shallow curves, either showing a very small change in total concentration of bound antibody over the dissociation phases, or a comparatively high initial rate of total dissociation in the early part of that phase, followed by a very low rate of total dissociation in the later part of the phase. These features lead to the XPR software estimating giving extreme estimates for affinity ( $8.7\text{E-}12\text{M}$  to  $3.8\text{E-}24\text{M}$ ).

The mixed monoclonal sensorgrams are shown in full in Section 4.8. Visually their

dissociation curves are not near as shallow as those from the high-affinity serum samples data set. As a result, the question as to whether there was any difference in total dissociation rate between the different mixtures of antibody was asked.

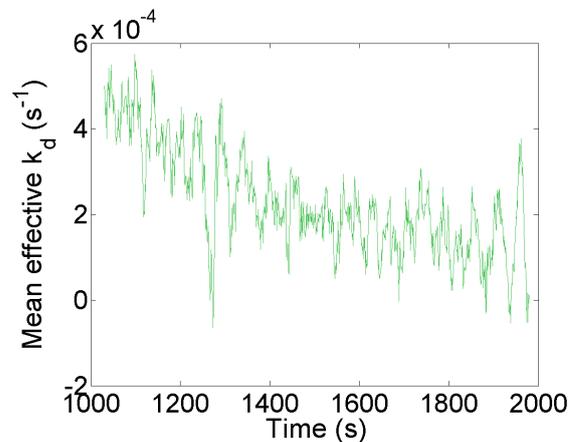
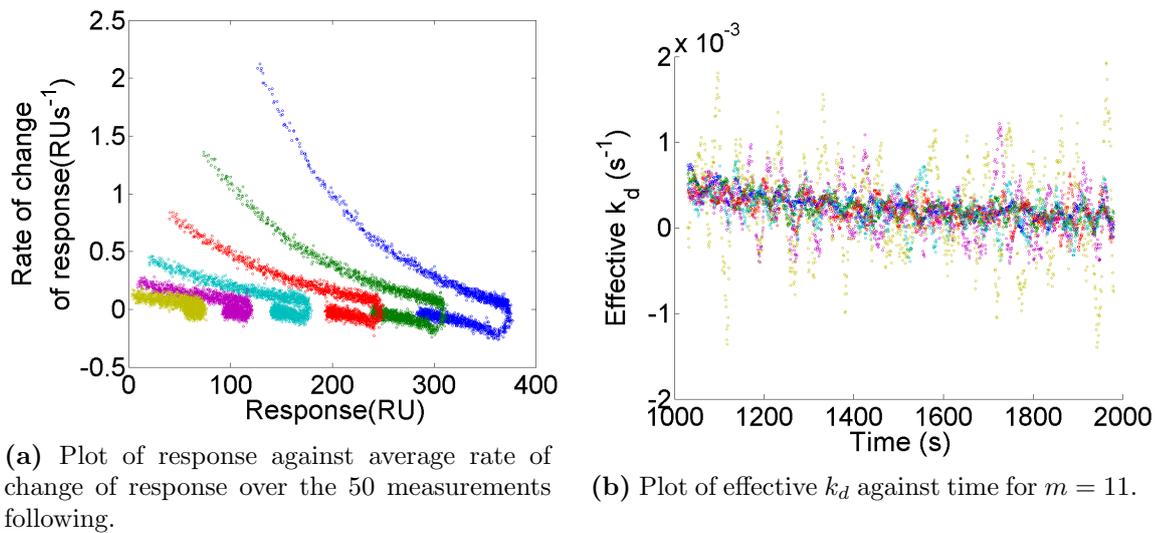
To answer this question effective  $k_d$  for a period of time  $[t_1, t_n]$  was defined. This was done first by calculating effective  $k_d$  at a point and then averaging it throughout the time interval.

The concept of effective  $k_d$  comes from the fact that the Langmuir model predicts that the output in the dissociation phase of an experiment is an exponential decay i.e.  $\dot{y} = -k_d y$ . The effective dissociation rate at a time point,  $k_d(t_i)$ , was defined as the quotient  $\frac{\Delta y(t_i, m)}{y(t_i)}$ , where  $\Delta y(t_i, m)$  is an approximation of the rate of change at  $y(t_i, m)$  defined as  $\frac{(y(t_{i+m}) - y(t_{i-m}))}{t_{i+m} - t_{i-m}}$ , where  $m$  is a parameter chosen by the user.

Figure 8.4 illustrates this process for a single sensorgram. Figure 8.4 (a) illustrates the rates of change for the 6 curves that make up the sensorgram for both phases. Figure 8.4 (b) shows how effective  $k_d$ , for  $m = 11$ , varies throughout the dissociation phase, and for each of the sensorgram curves. Whilst effective  $k_d$  values for each of the curves are affected by the noise of the SPR measurements, they clearly are clustered around a curve.

Averaging the effective  $k_d$  estimates for each curve in a sensorgram at a single point in time allows us to compare effective  $k_d$  at different time-points for the 5 antibody mixtures. Figure 8.5 shows this. They also show that the sensorgrams with mixed antibody have lower effective  $k_d$  than those with a single antibody only. However as a large value of  $m$  was chosen (200), so that the effective  $k_d$  curve for each sensorgram would be smooth in appearance and SPR measurement error would generally not cause the curves to overlap, these curves do not tell us about effective  $k_d$  at the end of the phase - particularly which mixtures give the lowest final  $k_d$ .

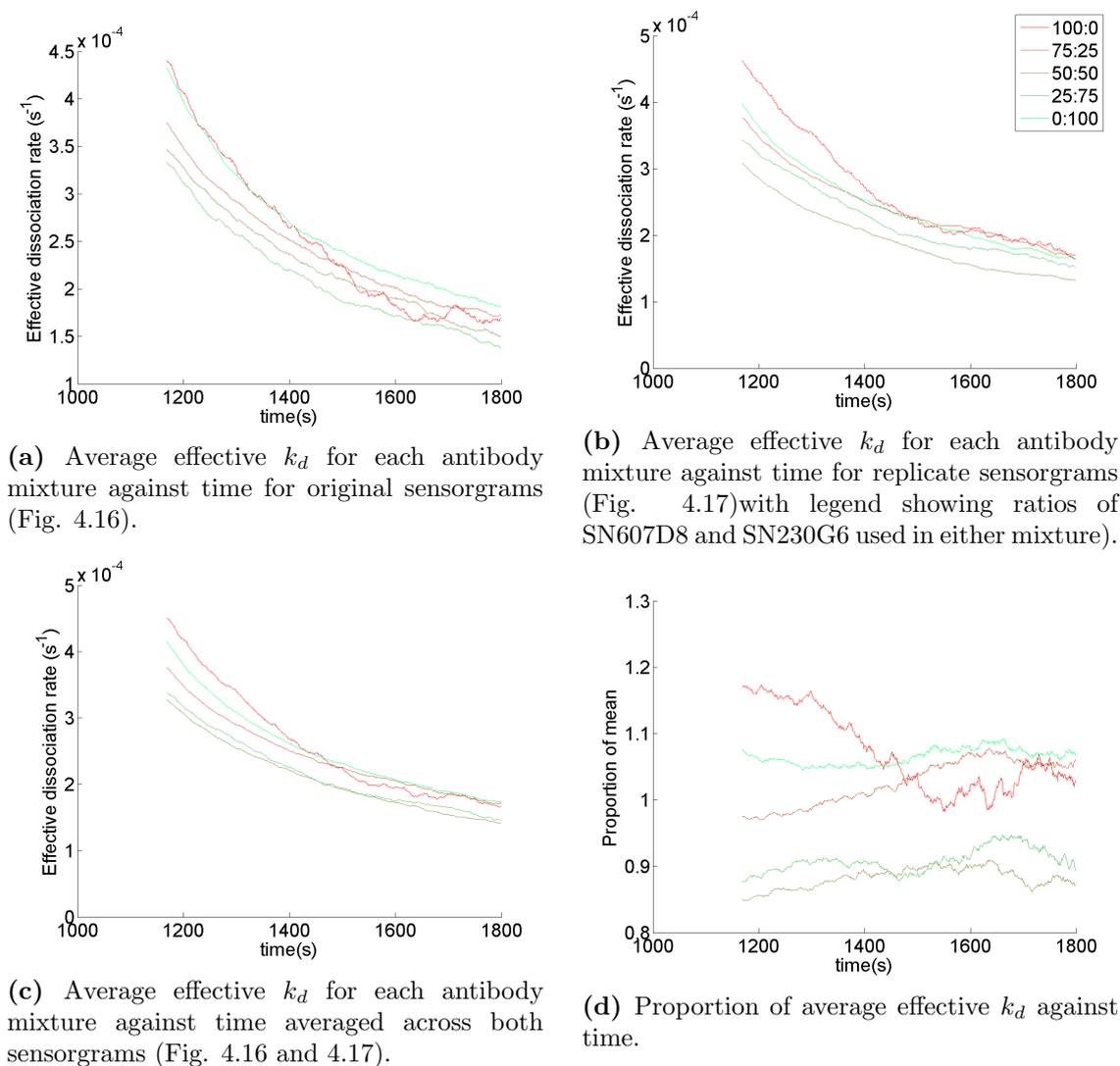
To more clearly show the difference between the sensorgram curves for each concentration



**Figure 8.4:** Alternate visualisations of the data from a single sensorgram Fig. 4.16(c).

Fig. 8.5 (c) was made. It shows the effective  $k_d$  at each point averaged across the two replicates. As a result the curve representing SN607D8 antibody exclusively is smoother. To better evaluate the proportional difference between these 5 curves, the average of them was found at each time-point, and the values the curves take at each time point was divided by this. This is shown in Fig. 8.5 (d).

To allow for more detailed comparisons between the curves in 8.5 (c), their mean effective



**Figure 8.5:** Average effective  $k_d$  for each mixed antibody sensorgram (Fig. 4.16 and 4.17) for  $m = 200$  against time, and related graphs.

$k_d$  across all lanes and across all time-points was calculated and they were scaled by it, to produce Fig. 8.5 (d). This shows that whilst initially the 50:50 mixed lane effective  $k_d$  values differ by over 30% of the mean, from the SN607D8 lane, this drops over the first half of the phase to only 20% which stays close to for the later half of the dissociation phase. Interestingly whilst the 50:50 mixture and 25:75 mixture have effective  $k_d$  values that remain close to each other through out the phase the 75:25 mixture remains closer

to the unmixed monoclonals.

Box-plots were made of the distribution of effective  $k_d$  for  $m=11$  over the last 60 measurements, for which effective  $k_d$  could be calculated at  $m=11$ . These are shown in Fig. 8.6.

This value of  $m$  was chosen to be small enough that effective  $k_d$  calculations could be made close to the end of the phase (only excluding the last 11 measurements) and large enough that after the effective  $k_d$  values for each time point were averaged across replicates the whiskers of all box-plots (representing the mean plus-minus 1.5 standard deviations) did not go below 0.

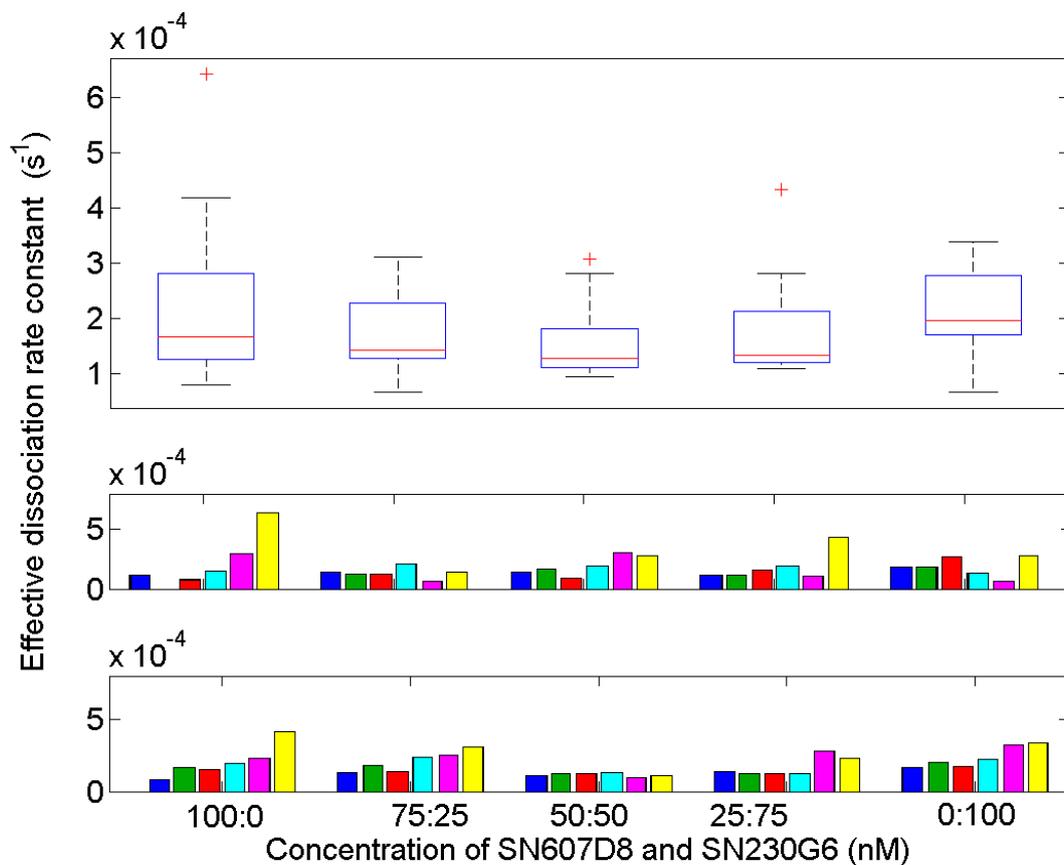
Figure 8.6 also shows mean effective  $k_d$  for the last 60 usable measurements for each of the sensorgram curves, so they can be compared.

Figure 8.6 also shows how the distribution of effective  $k_d$  values varies as the mixture of the two antibodies vary. Notably, the effective dissociation rate constants are higher for the single antibodies, and lowest for the 50:50 mixture. Like Fig. 8.5 this suggests that the two antibodies are in some way slowing each other's dissociation.

The one-way ANOVA (Howell, 2002, pages 324-325) test was used to determine if there was a significant difference between the means of the effective  $k_d$  values for each mixture. The test rejected the null hypothesis, that there was no significant difference between the means of these groups, with the p-value 2.8E-6.

This feature of the data can be seen in the sensorgrams (Fig. 4.16 and 4.17), by comparing the lanes with equal total concentration. Notably, the mixed lanes, especially the 50:50 mixed lanes have a similar slope at the end of their dissociation phases but a higher overall response. As effective  $k_d$  is a measure of the quotient of the rate of change of response over response, this means they have lower effective  $k_d$ s.

This supports the hypothesis that underlies the GPM (as presented in Subsection 7.2);



**Figure 8.6:** Boxplots of average effective  $k_d$  at the last 91s of the dissociation phase in the mixed monoclonal experiments (top) and bar graphs average effective  $k_d$  for each lane in initial experiments (middle) and replicates (bottom)

that is that multiple antibodies against the same antigen, with separate epitope targets, form into high avidity multi-antibody multi-antigen complexes, which lead to slower overall antibody dissociation. The next toward validating the model was using it for parameter estimation.

### 8.3.1 Using the GPM2X2 to estimate concentration of antibody in a mixture

The first fits of the GPM model were intended to gauge to what degree it could estimate the concentration of antibody in a mixture given the association and dissociation rate constants of the antibody in that mixture.

In the Section 8.2 it was established that the ERC2C model gave better fits to polyclonal data than the L2C model. As a result an ERC version of the GPM was used. The experiments were conducted with two antibody which bound to separate epitopes of the same antigen, so the form of the model introduced in Subsection 7.2.4 was used, the GPM2X2.

As in Chapter 6 differential evolution was coupled with FACSIMILE to fit parameters. These fits assumed the kinetic rate concentrations of SN607D8 and SN230G6 as estimated in Section 6.6, and the known total concentrations of antibody, and fit all other parameters, including proportion of antibody of each type.

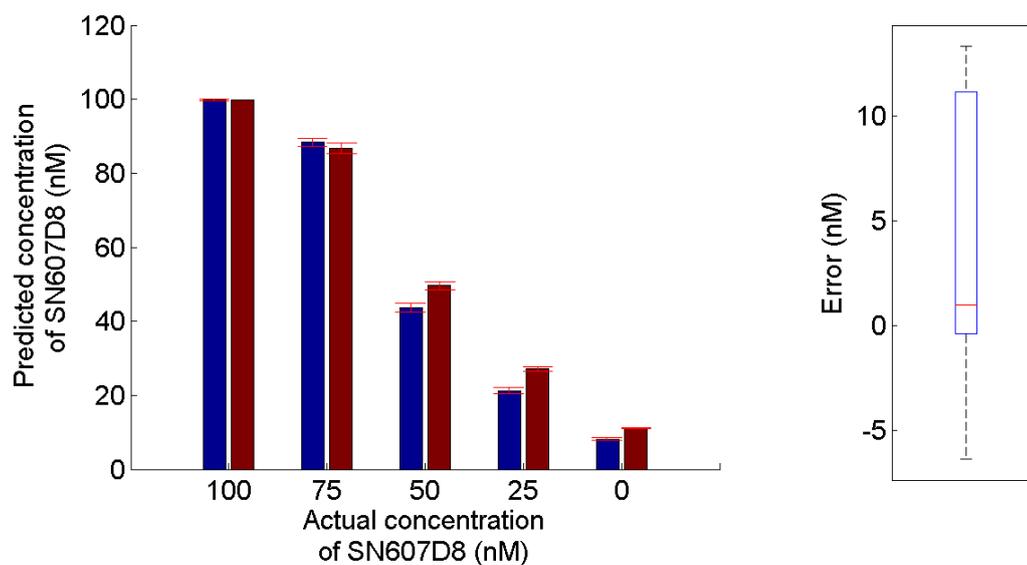
As FACSIMILE does not allow for the variation of a parameter to be constrained to an interval, a new unitless parameter was introduced,  $D$ , to allow FACSIMILE to vary the proportion of each antibody where:

$$C_1 = \frac{100}{1 + D} \quad (8.1)$$

$$C_2 = 100 - C_1 \quad (8.2)$$

Figure 8.7 shows the estimates made of SN607 concentration and the error of these estimates. The actual concentration of antibody is correlated with the estimated concentration of antibody and the average error of the estimates for SN607 is 1.0nM.

This suggests that the GPM2X2 can accurately estimate the concentration of antibody



(a) Bar graphs showing each estimate, and confidence intervals for (b) boxplot of errors of each estimate

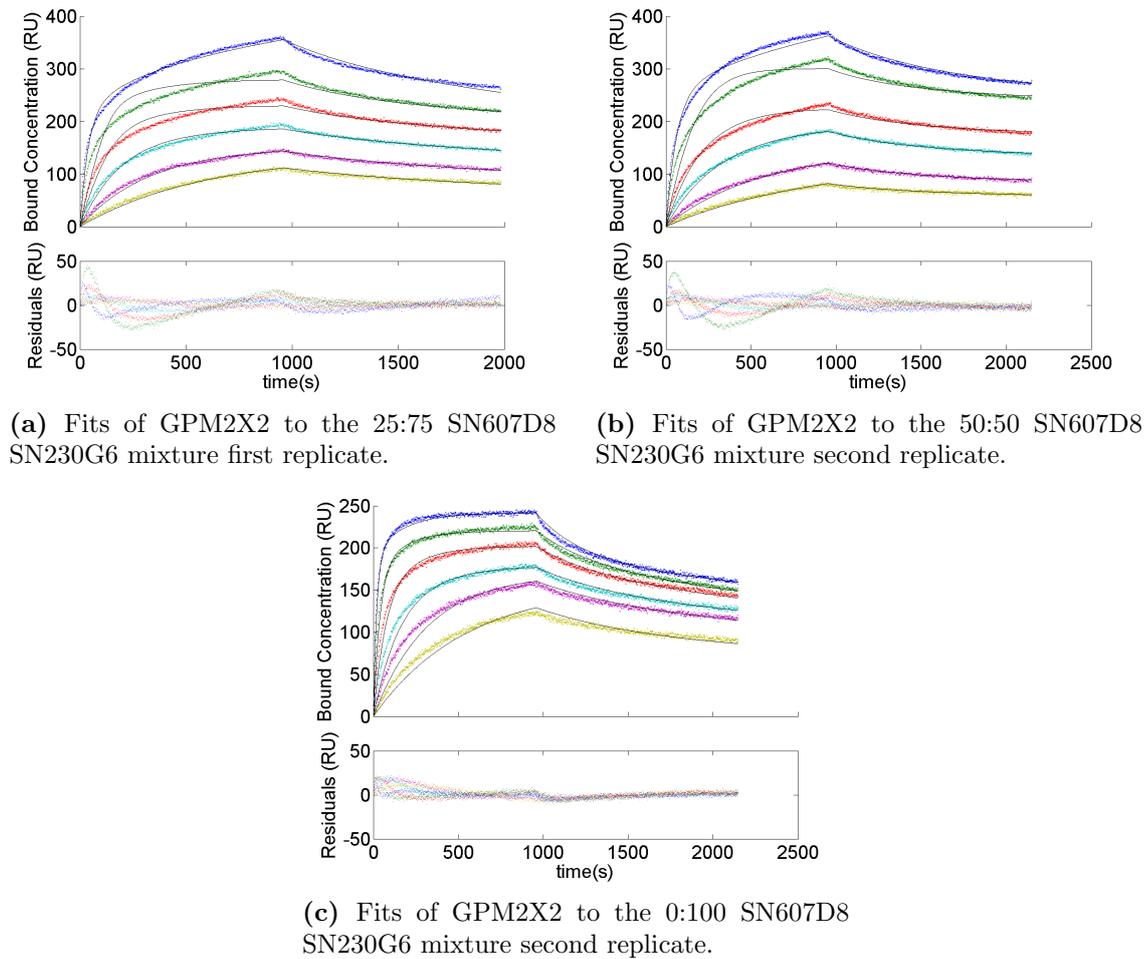
**Figure 8.7:** Boxplot of the error of these estimates.

in a mixture given the association and dissociation rate constants of the antibody in that mixture.

However, the sensorgram fits were generally not as good as those produced with the BERCS model to the monoclonal data (see section 6.6). Figure 8.8 shows two typical fits, in both cases the model generally fits well- especially at the lowest three concentrations (3.1nM-12.5nM), but worse on the higher concentrations (25nM-100nM). At these concentrations the model underestimates the output in the early part of the association phase, overestimates it in the mid part of the phase, and underestimates it at the end. As the model predicts the wrong output at the start of the dissociation phase, the early part is typically badly fit, although the latter part is well fit.

These problems with fitting lead to large residuals, much larger than those produced with the BERCS model to the monoclonal data (see section 6.6).

The exception to this was sensorgrams where the mixture include only one of the two



**Figure 8.8:** GPM2X2 fits to sensorgram data of mixed antibody from Fig. 4.16 and 4.17.

antibodies, such as Fig. 8.8 (c), where the sensorgram fits and residuals were of similar quality to those produced with the BERCS model to the monoclonal data (see section 6.6).

It was unclear what was causing these problems with the fits. One explanation was that the larger number of variables (11 per lane, 66 per sensorgram) was leading to longer integration time, and thus the correct parameters were not being found, another competing explanation was that there were interactions between SN607D8, SN230G6 and HLA-A2 that were not included in the model.

### 8.3.2 Fitting the GPM to sensorgrams with mixtures of monoclonal antibody reacted against a single antigen

To reduce the number of variables in each numerical integration process that FACSIMILE would conduct datasets were composed of the interactions where the antibody had the same total concentration, rather than where antibody was in the same proportion (see Fig. 8.10). This meant that rather than having 10 datasets with 6 interaction spots we now had 12 with 5.

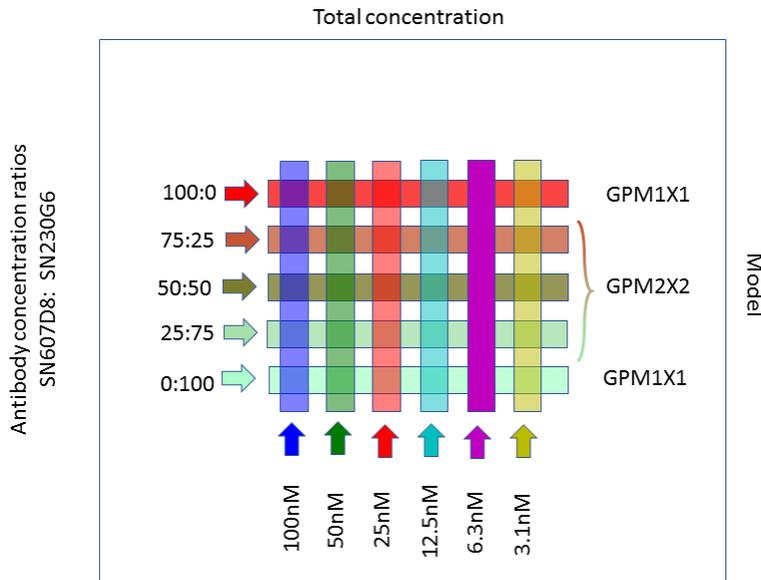
Additionally Each dataset now included output from two interaction spots where only one antibody was interacting with the ligand. The GPM1X1 model was used for these spots. As a result the number of variables FACSIMILE would have to integrate per dataset was reduced from  $6 \times 11$  to  $3 \times 11 + 2 \times 3$ . This allowed for faster numerical integration and thus a faster search of parameter space.

As one of these datasets was missing the output from an interactions spot it was removed before fitting (see Fig. 8.11 (a) and (b)).

Unlike the Fits of the previous section, these were intended to see if good visual fits could be made with known concentration and association constants as a result, these were not varied.

Four of the eleven fits are shown in Fig. 8.11. These have generally small residuals, nearly as small as those produced with the BERCS model to the monoclonal data (see section 6.6). The model lines in nearly all cases remain within the data. The biggest exception to this is the 100nM concentration lane with a 75:25 mixture of the two antibodies in the late association phase.

The parameter estimates from these fits are shown in Table 8.5. The vast majority of parameters are well determined, with only 123 of the 682 being marked as NWD. However



**Figure 8.9:** Total antibody concentration of 100nM

**Figure 8.10:** Diagram of the ways the outputs from interaction spots may be grouped for modelling

there is a great variety in the estimates of  $\gamma$  and in the majority of cases  $\beta_1$  and  $\beta_2$  are NWD.

Interestingly Fig. 8.11 also highlights something that can be seen in the bar graphs of maximum bound concentration in Fig. 4.16 (a) and Fig.4.17 (a), i.e. that for the lower total concentrations of antibody (6.2nM-3.1nM) in either replicate the pure SN230G6 has a higher maximum response than the 25:75 mixture.

The hypothesis of antibody-antigen tetramers being formed explains this, because once concentration is low enough, it is unlikely that a single antigen will bind to both antibodies, so the sensorgrams maximum output decreases.

These parameter estimates allow us to evaluate to what degree the parameters used in previous simulations reflect those estimated by the fitting process. In Fig.7.13 it was shown

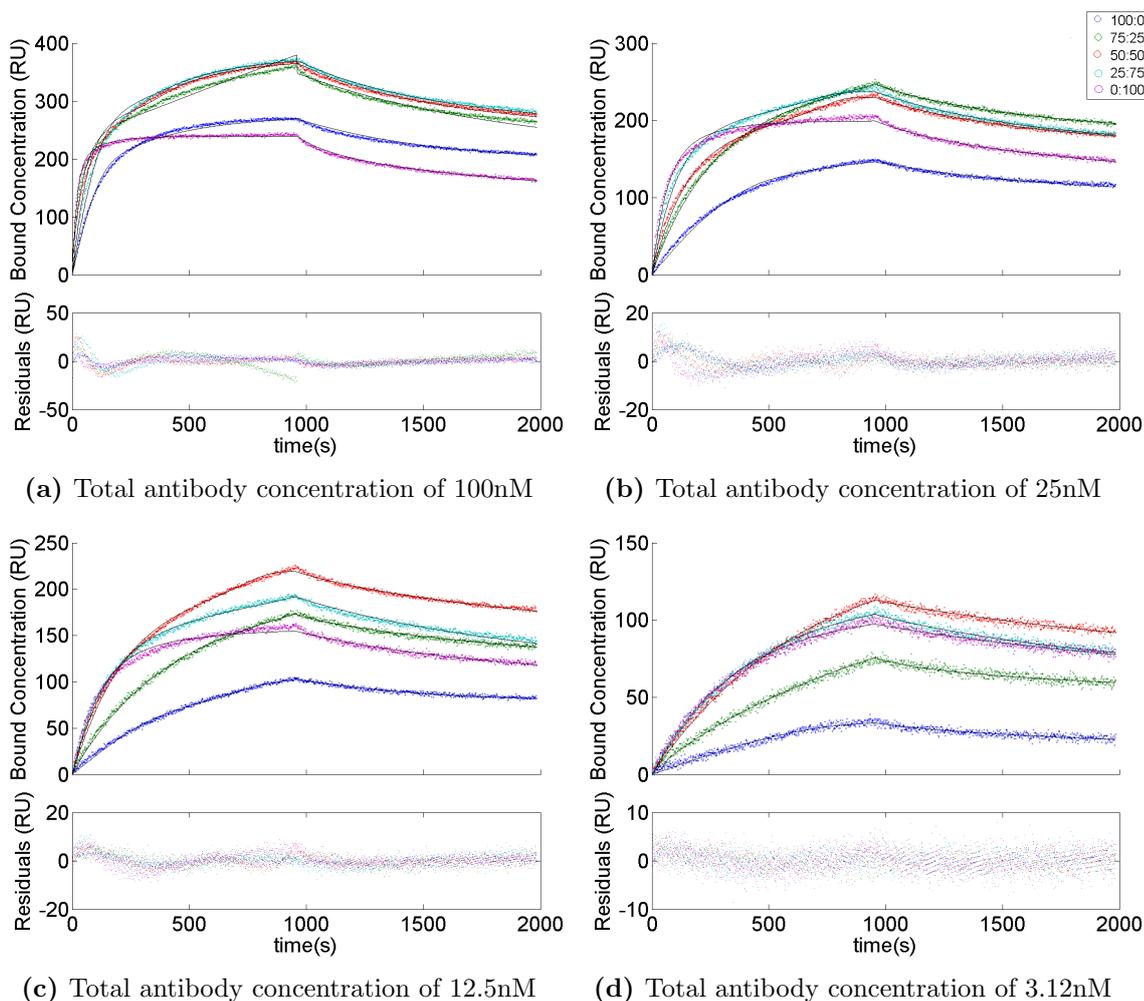
$\beta_1$ (Mnm <sup>-2</sup> pg <sup>-1</sup> )	100nM	50nM	25nM	12.5nM	6.3nM	3.1nM
$\beta_2$ (Mnm <sup>-2</sup> pg <sup>-1</sup> )	1.1E+2*	7.1E+1*	2.0E+1*	7.6E+1*	1.0E+2*	2.1E+1
$\gamma$ (Mpg <sup>-1</sup> nm <sup>2</sup> )	6.6E+1*	8.4E+1*	5.8E+1*	3.0E+2*	1.1E+2*	9.2E+2
$R_1$ (pgnm <sup>-2</sup> )	2.2E-3	4.0E+2*	1.6E-3*	8.6E+5*	1.2E+1*	1.2E+5
$R_2$ (pgnm <sup>-2</sup> )	5.3E+0*	2.2E-1	1.4E+0	1.9E+0*	1.9E-1	1.7E-1
$R_3$ (pgnm <sup>-2</sup> )	3.3E-1	1.8E+0*	1.1E+0*	3.1E-1	2.1E-1	2.8E-1
$R_4$ (pgnm <sup>-2</sup> )	4.4E-1	4.9E-1	4.0E-1	3.1E-1	2.4E-1	3.1E-1
$R_5$ (pgnm <sup>-2</sup> )	3.5E-1	4.1E-1	3.1E-1	2.5E-1	1.9E-1	2.2E-1
$k_{m11}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	1.9E+1	2.4E-1	3.2E+0	3.2E+0	3.7E-1	2.2E-1
$k_{m21}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	2.9E-3*	1.3E-4	1.2E-3	2.8E-4*	7.2E-3*	2.5E-5
$k_{m31}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	5.6E-4	5.5E-3*	2.8E-3*	2.9E-5	1.4E-1*	8.6E-6
$k_{m41}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	7.0E-4	5.3E-4	3.5E-4	4.9E-5	4.2E+0*	2.3E-6
$k_{m22}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	5.2E-3	1.0E+1*	8.0E-4	5.8E-5	1.3E+2*	1.5E-5
$k_{m32}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	4.4E-5	1.2E-4	6.9E-5	1.3E-1*	1.4E-1*	2.0E-3
$k_{m42}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	1.3E-4	1.5E-2*	1.3E-1*	4.0E-4	4.0E+0*	1.2E-3
$k_{m52}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	6.9E-5	1.2E-7*	8.8E-2*	8.4E-4	1.3E+2*	8.8E-4
$T_{1b}$	9.6E-3	2.2E-2*	2.6E-3	6.6E-3	4.2E+3*	4.5E-2
$T_{2b}$	1.8E+1*	3.8E-6*	8.4E+0	1.2E+1*	7.0E-1*	3.3E+0
$T_{3b}$	1.8E+0	1.2E-1*	2.0E-1*	1.8E-1	2.3E-1	1.4E-1
$T_{4b}$	2.9E-1	1.9E-1	5.9E-2	1.9E-1	3.5E-1	1.6E-1
$T_{5b}$	2.9E-1	3.2E-1	7.5E-2	3.7E-1	4.4E-1	1.7E-1
	9.0E+1	4.4E-2*	1.9E+1*	1.2E+1*	7.3E-1*	8.5E-2

(a) 1<sup>st</sup> replicate

$\beta_1$ (Mnm <sup>-2</sup> pg <sup>-1</sup> )	100nM	25nM	12.5nM	6.3nM	3.1nM
$\beta_2$ (Mnm <sup>-2</sup> pg <sup>-1</sup> )	5.8E+1*	2.2E+1*	1.8E+1*	4.6E+0*	4.8E+1*
$\gamma$ (Mpg <sup>-1</sup> nm <sup>2</sup> )	4.0E+2*	5.4E+2*	5.7E+2*	5.6E+1*	1.5E+6*
$R_1$ (pgnm <sup>-2</sup> )	1.8E-4	1.4E+1	5.9E+3	8.0E+5*	1.1E+3*
$R_2$ (pgnm <sup>-2</sup> )	3.4E+0	3.9E-1	1.1E+0*	3.6E+0	1.6E-1
$R_3$ (pgnm <sup>-2</sup> )	2.4E+0	4.2E-1	3.4E-1	7.3E+0	2.9E-1
$R_4$ (pgnm <sup>-2</sup> )	4.9E-1	3.1E-1	2.5E-1	1.1E+0	2.3E-1
$R_5$ (pgnm <sup>-2</sup> )	5.0E-1	3.3E-1	2.6E-1	1.1E+0	3.4E-1
$k_{m11}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	4.1E-1	2.5E+0	3.8E-1	5.0E+0	3.9E+0
$k_{m21}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	1.7E-3	8.0E-5	1.6E-4*	7.4E-5	2.3E+3*
$k_{m31}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	7.8E-3	3.5E-5	2.6E-5	1.5E-4	5.5E-2*
$k_{m41}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	1.7E-4	5.1E-5	4.4E-5	4.2E-3*	1.7E-1*
$k_{m22}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	3.3E-4	1.2E-4	1.0E-4	5.5E-1*	3.1E+0*
$k_{m32}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	3.5E-5	5.7E-2*	3.5E-3*	1.4E-4*	1.2E-2*
$k_{m42}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	1.3E-3	9.2E-4	1.0E-3	8.4E-3*	1.5E-3
$k_{m52}$ (pgnm <sup>-2</sup> M <sup>-1</sup> s <sup>-1</sup> )	3.0E-4	8.2E-4	2.2E-3*	3.3E-1*	5.5E-1*
$T_{1b}$	9.1E-2*	5.9E-2*	1.3E-1*	3.4E+1*	7.9E+1*
$T_{2b}$	8.1E+0	1.2E+0	5.2E+0*	3.5E-3*	8.5E-1*
$T_{3b}$	1.0E-1	1.7E-1	1.4E-1	8.4E+1*	1.9E-1
$T_{4b}$	4.8E-2	2.1E-1	2.0E-1	3.1E-3*	4.4E-1
$T_{5b}$	5.7E-2	1.8E-1	2.0E-1	1.3E-1	5.5E-1
	1.6E-2*	6.2E+0*	2.7E-1*	4.1E+0*	1.3E+1*

(b) 2<sup>nd</sup> replicate

**Table 8.4:** Parameter estimates with the GPN model ran on the mixed monoclonal dataset.



**Figure 8.11:** Fits of the GPM to sensorgrams with mixtures of SN607D8 and SN230G6 in the mixtures 100:0 blue, 75:25 green, 50:50 red, 25:75 cyan, and 0:100 purple

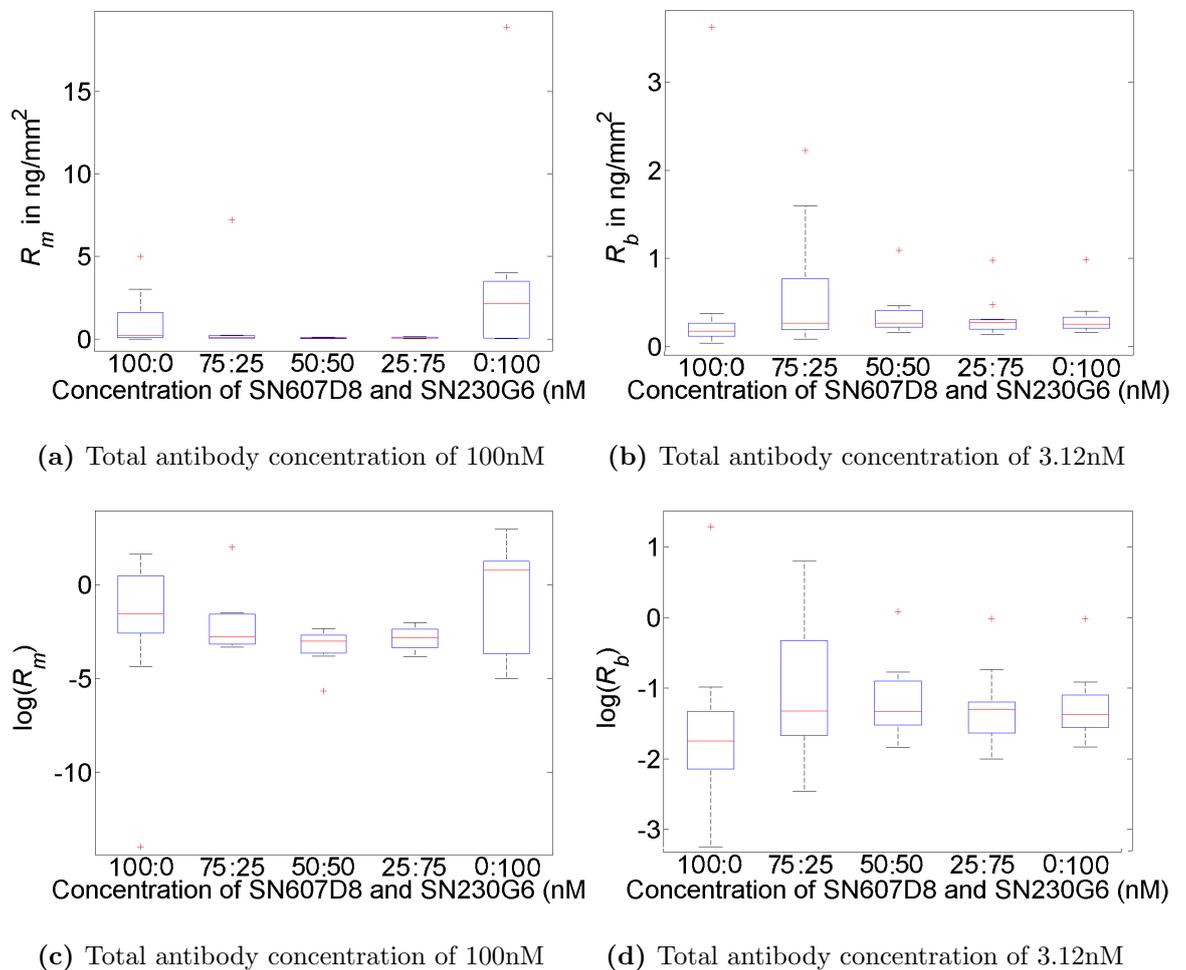
that for  $\gamma > 5$  antibody bound to non-isolated antigen dissociate very slowly. In 8 of the 13 estimates  $\gamma$  is in this range. However these sensorgrams do not look visually similar to those simulations.

This is largely for two reasons, firstly that the parameters used for that figure used  $R = R_b, R_m = 0$  to illustrate the interactions of antibody on antigen able to form multi-antigen complexes (Table 7.2). Secondly a much larger value for  $\beta$  was used, which was taken by averaging the quotients  $k_{a1}/k_{a2}$  obtained in the fits for SN230G6 and SN607D8 with the

BERCs model (Table 6.6.1).

The values of  $\beta$  estimated for each antibody here are much lower - suggesting that something about the interaction is reducing the rate that multi-antigen complexes are being formed.

The distribution of estimates for  $R$  and  $T_b$  appear to be dependant on the mixtures of antibody. As a result the parameters they determine  $R_b$  and  $R_m$  were plotted in Fig. 8.12.



**Figure 8.12:** Plots of the distribution of estimates of  $R_b$  and  $R_m$  and their logs in the mixed monoclonal fits

These graphs show that there is little variation between the average estimates of amount of antigen to which antibody is only able to bind monovalently (Fig 8.12 a) for the mixed experiments, but huge differences between those estimates and the estimates made where the mixture contains only one antibody. When the logs of these estimates are compared (Fig 8.12 c) it becomes apparent that the estimates are generally lower for the 50:50 mix but only very slightly.

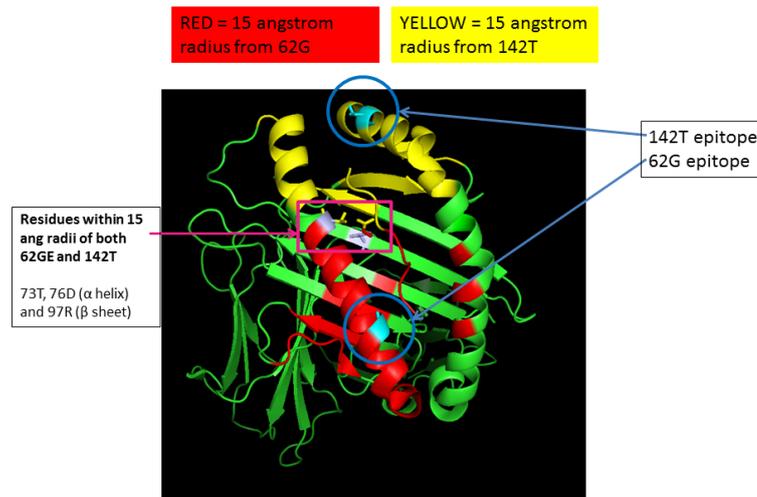
There is much less variation between the distribution of estimates of antigen available that antibody can bind bivalently on (Fig. 8.12(b) and (d)). This is interesting because whilst we would expect there to be variation in the proportion of antibody available to bind in either way, due to variations in the structure of the surface of the chip at its interaction sites, but we would not expect it to be this much.

It is possible that the model fits the data wrongly because it misses a feature of the interaction between the antibody and antigen. Intuitively explanations for this could be either that slight epitope overlap or that there was some kind of allosteric effect.

The possibility of slight epitope overlap or allosteric effect was investigated, and it was found that because of the width of the binding arm of the antibody (15Å (Duquesnoy, 2006, 2007, 2014)), that the two antibodies interfere with each other's binding.

This is illustrated in Fig. 8.13, taken from Jones *et al.* (2016, personal communication). On the HLA-A2 antigen the epitopes the antibodies bind to, 142T and the 62G, are within 30Å. As a result, when one antibody binds to one, it is expected to interfere with another binding to the other.

This is a significant result because it shows that the GPM  $2 \times 2$  so well captures the dynamics of antibody antigen interactions, that where discrepancies in parameter estimates exist they can be used to detect unusual features of the binding reactions- such as this steric effect.



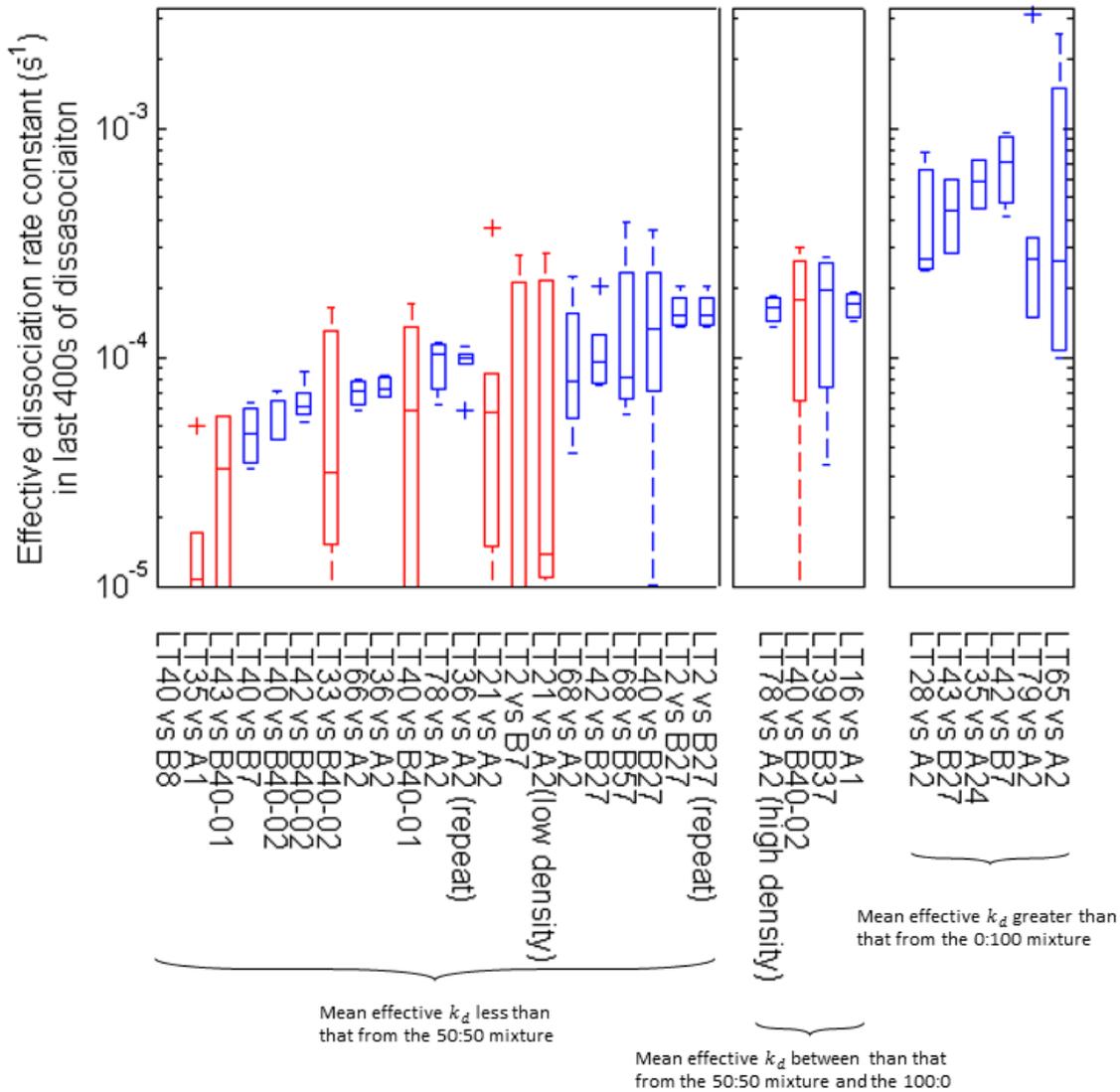
**Figure 8.13:** Diagram of the antibody footprint of SN230G6 (yellow) and SN607D8 (red) and their overlap (purple) on HLA-A2 taken from Jones *et al.* (2016, personal communication)

## 8.4 Sensitized patient study

In Section 4.7 it was noted that the dissociation curves in the high affinity polyclonal dataset, looked different to those of the previously taken sensorgrams (Sections 4.3-4.6). The curves appeared shallower with very low amounts of dissociation throughout the phase.

In Section 8.3 a tool for analysing this, effective  $k_d$  was introduced. This can now be applied to the sensorgrams from the sensitized patient study (Section 4.9), to evaluate whether the effective  $k_d$  values from it differ with those derived for mixtures of monoclonal antibody in Section 4.9, and if the high affinity subset (Section 4.7) differs from the others.

The result of this analysis is shown in Fig. 8.14. As the sensorgrams produced were much shallower a longer period could be used (averaging effective  $k_d$  at  $m = 11$  for the



**Figure 8.14:** Boxplots of average effective  $k_d$  at the last 400s of the dissociation phase in the polyclonal experiments

last 400s per curve). However some sensorgrams had abnormalities, such as increases in response in the dissociation phase or negative responses, these lead to negative effective  $k_d$  values, that could not be printed on the graph. Where an effective  $k_d$  was negative the boxplot was marked in red.

In Fig. 8.14 sensorgrams are sorted into three categories: those for which the mean effective

$k_d$  is lower than the lowest of the mixed monoclonal reactions, the 50:50 sensorgram (21/31); those for which the mean effective  $k_d$  is between that and the lowest of the reactions with a single monoclonal, the 100:0 mixture (4/31); and as there are no reactions with mean effective  $k_d$  values between the single monoclonal values, those for which the mean effective  $k_d$  is greater than that of any of the monoclonal mixtures (6/31).

This suggests that either the same kind of interactions that lead the mixed monoclonal antibody to have higher maximum outputs and lower effective  $k_d$  values may be happening in a large number of the sensorgrams, or individual patient antibody with extreme affinity are involved in the interactions.

Whilst the Proteon Manager affinity estimates (Table 4.5) that originally motivated this study might not be physically meaningful, Fig. 8.14 shows that there is something unusual about the sensorgram outputs from 4.9 especially those in Fig. 4.14.

## 8.5 Modelling sensitized patient sensorgrams with variations of the GPM model.

The initial hypothesis that was held when this data was analysed was that the 1X1 and 2X1 form of the GPM (Section 5.1.6 and Section 7.2.3) would provide the good fits for the sensorgrams with higher effective  $k_d$  values, but would either give bad fits or poor parameter estimates for the lower effective  $k_d$  values.

Conversely, the 2X2 form of the GPM (Section 7.2.4) was expected to fit the data well with lower effective  $k_d$  values, but offer only marginal improvement to the data with higher effective  $k_d$  values.

It was noted that in some cases the sensorgrams don't just have very slow rates of change of output at the end of the dissociation phase but throughout the dissociation phase (Fig.

4.23 (b)-(d), Fig. 4.24, Fig. 4.28). One explanation for this could be larger antibody molecules binding bivalently to an antigen that would normally be isolated.

Notably, Fig 4.10 (c) shows the binding curve for the patient protein after IgG and IgA have been removed and it reaches close to half the maximum response of the IgG sample (Fig 4.10 (a)), suggesting that much larger IgM antibody should be expected in at least some patient samples (see also Section 1.2).

To investigate the possibility of the presence of multiple sizes of antibody acting causing this sensorgram feature the GPMSS was developed, so that its 2X2 version could be applied to the data (Section 7.2.5).

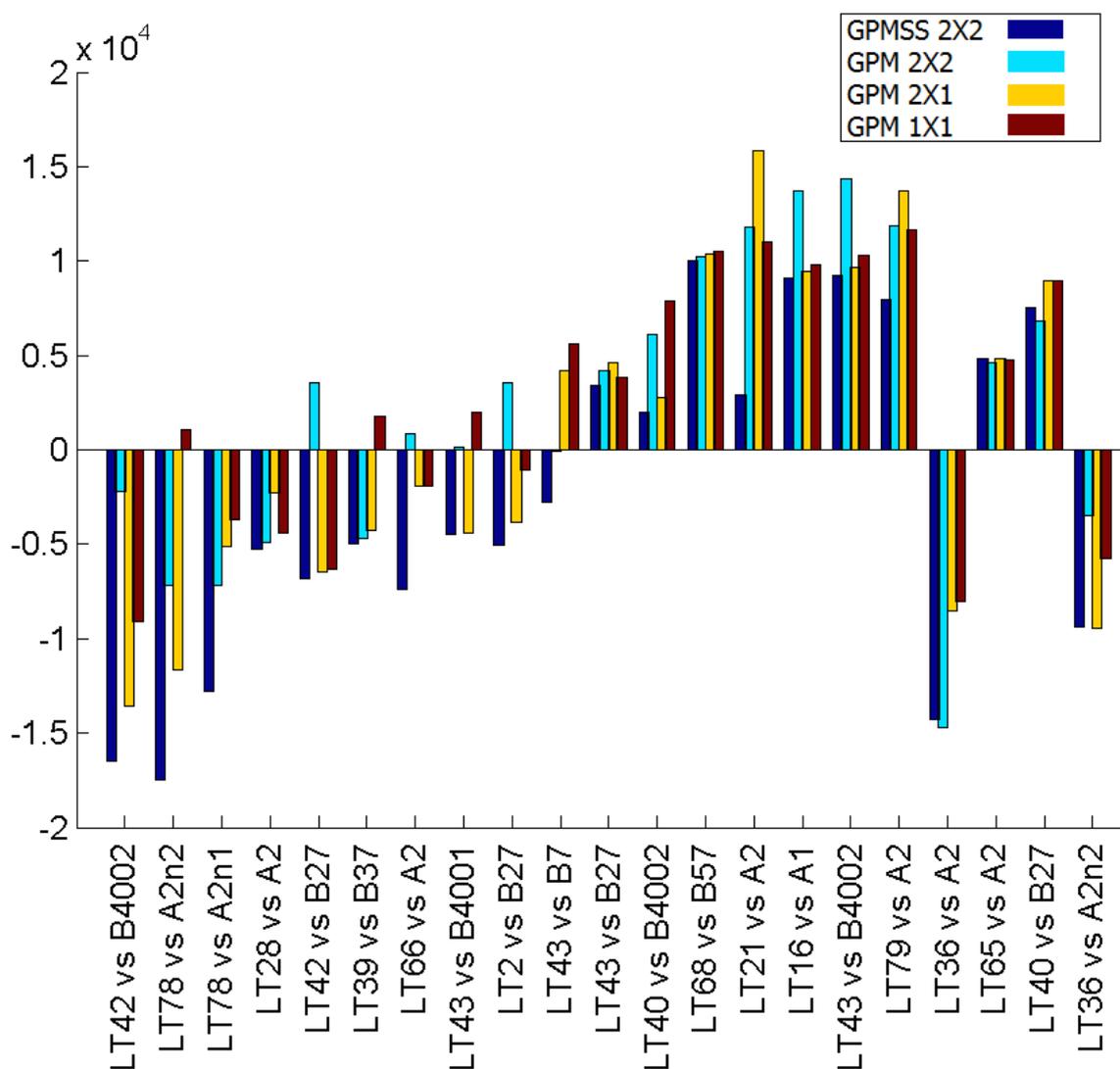
Lanes including unusual behaviour- negative response or sustained positive rates of change of output in the dissociation phase were removed from the data and these four models, the GPM 1X1, 2X1, 2X2 and GPMSS 2x2 were fit.

As in Section 6.4.5 simplifications of the BERCs model had lead to faster parameter fitting, it was hypothesised that making the new assumption, that  $k_m$  is the same for both kinds of antibody in a single lane, would lead to better fits. As a result the tables of parameter estimates (Table 8.5-Table 8.8) have at most six  $k_{mi}$  parameters, one for each lane.

To allow for varying concentrations of either antibody the parameter  $D$  and Eq. (8.2) were again used for the models that assumed multiple antibody were binding GPM 2X1, 2X2 and GPMSS 2x2. In 6 of the 21 datasets the wrong parameter was passed to the command in FACSIMILE that leads it to be varied, as a result the parameters  $S_{ib}$  were varied instead of  $T_{ib}$  (see Table 8.5).

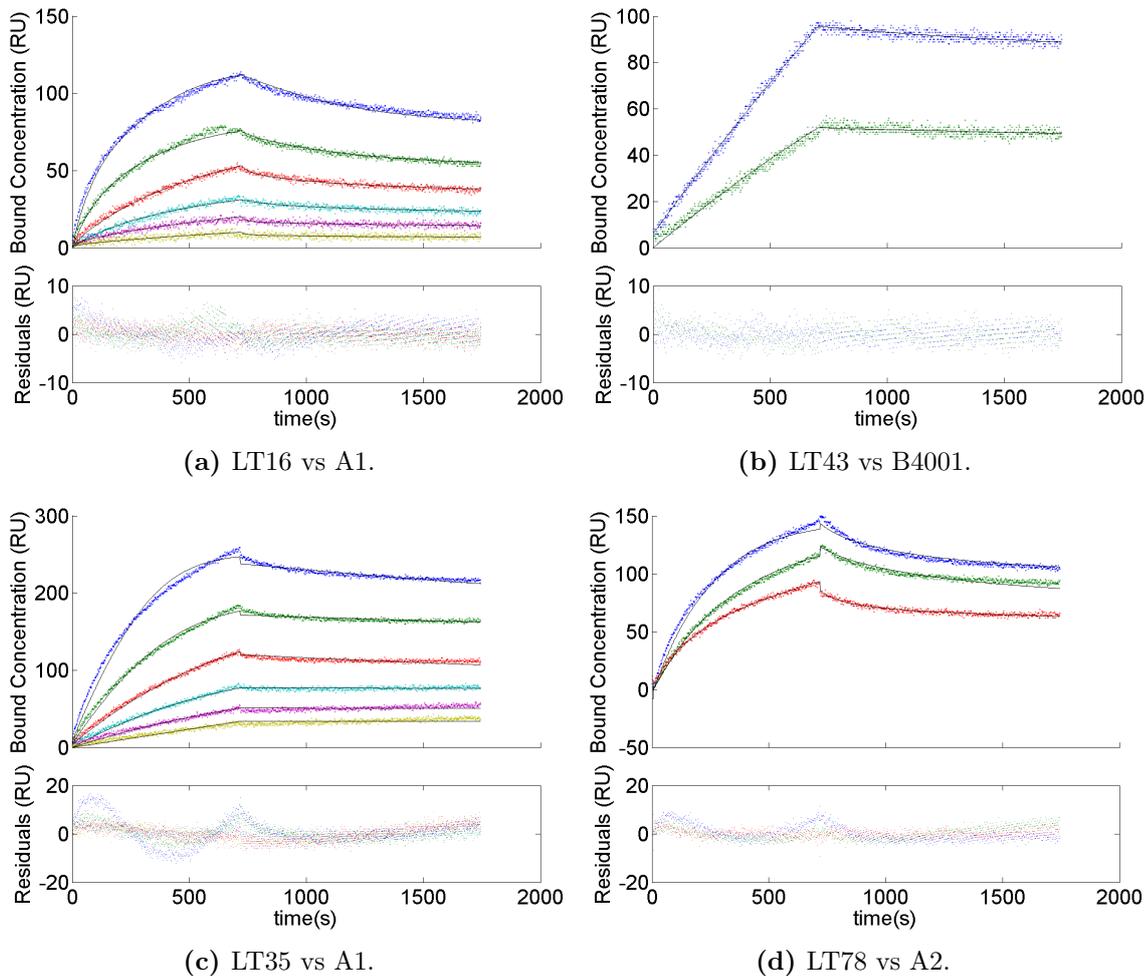
The AICs of their fits are presented in Fig. 8.15. The GPMSS 2X2 gave consistently the best AICs (17 out of 21 sensorgrams), the GPM 2x2 gave the best fit to the remaining four sensorgrams.

Whilst in 6 cases the inferior fit of the GPM 1X1 is influenced by the wrong parameter



**Figure 8.15:** AICs for the fits to the data with the four models, 2 antibody of separate size non-competitive (dark blue), 2 antibody same size non-competitive (cyan), 2 antibody competitive (yellow), BERCs model (red)

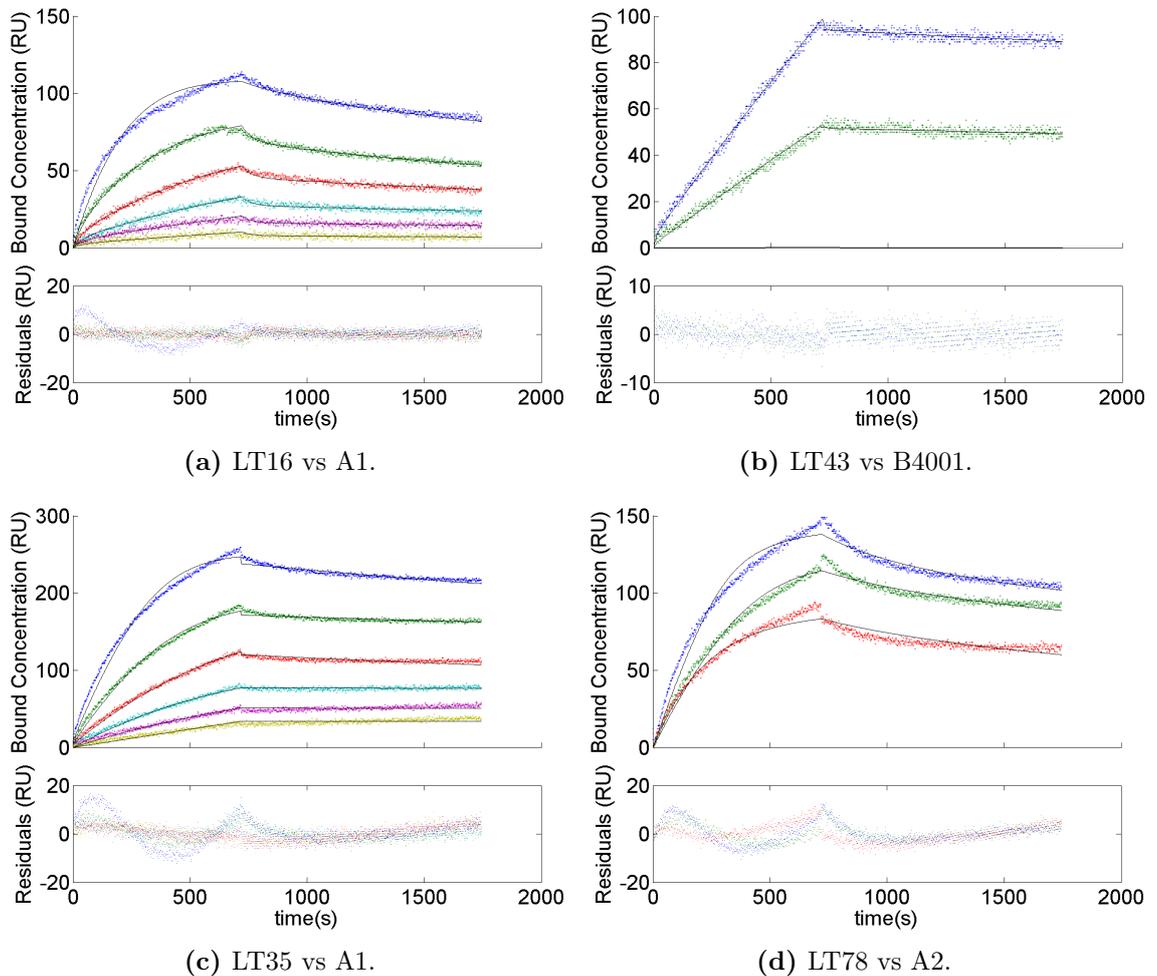
being varied, as all the outputs the model can give for a given input are contained by the set of outputs the GPM2X1 can make for the same input. It is demonstrated that the fits are consistently inferior to the GPMSS 2X2 and GPM 2x2 because the GPM 2X1 fits also consistently have worse AICs than the 2x2 models.



**Figure 8.16:** Plots of fits of the GPMSS 2X2 to patient sensorgrams.

The fits of the separate size model are amongst the best in thesis (see Fig. 8.16). Particularly they fit the very slow dissociation in the early parts of the dissociation phase well.

In contrast to these the fits from the GPM1x1 have in many cases the model lines outside of the data (Fig. 8.17). Interestingly these fits of the GPM1X1 often show strange features at the beginning of the dissociation phase where there is a rapid dissociation of a small amount of antibody predicted by the model, but not in the data (Fig. 8.17 (b)), cross referencing this with the tables of parameter estimates (Table 8.5) this is reflected in very



**Figure 8.17:** Plots of fits of the GPM 1X1 to patient sensorgrams.

high values for the  $T_{ib}$  parameters (see Eq. (6.5)-(6.6)), meaning that this model fits the data in many cases by taking extremely low values of  $R_{ib}$  to make the model dissociation curves as shallow as they are in the data.

Interestingly the fits from the GPM1x1 are consistently worse than those for the BERCs model to the monoclonal data (see Fig. 6.27) even though the models are indistinguishable, and rely on the same assumptions - one antibody binding to one antigen, ERC concentration and some antigen being isolated. This illustrates that neither is able to fit data that does not come from experiments where these assumptions are violated.

Parameters from these fits are shown in Tables 8.5-8.8. Unlike previous usages of the same models (the GPM 2X2 model in Subsection 8.3.2, and the BERCS model, equivalent to the GPM 1X1 model in Section 6.6) there are large numbers of parameters marked with an asterisk for being not well determined (NWD): 116 of 351 parameters are marked as NWD for the GPM1X1, 216 of 414 parameters are marked as NWD for the GPM2X1, 214 of 456 parameters are marked as NWD for the GPM2X2 and 192 of 456 parameters are marked as NWD for the GPMSS2X2.

The key parameters relating to association are graphed in the Fig. 8.18 and Fig. 8.19. As concentrations of the analytes were unknown, the association rate constants  $k_{ai}$  could not be directly estimated, instead the association rate concentration product  $k_{ai}C_i$  was. This is shown in Fig 8.18, and the corresponding  $k_{ds}$  are shown in Fig. 8.19. The distribution of both is shown in Fig. 8.20.

Interesting comparisons can now be made from the distribution of the kinetic rates both between the models and between these and the monoclonal models. Fig. 8.20 shows the way the distribution of the estimated constants varies between models.

Estimates of dissociation rate constants,  $k_d$ , in the monoclonal study ranged from  $10^{-4}$  to  $10^{-2}$  (see Fig. 6.28) for all models we see dissociation rates being estimated outside of this range. Interestingly the most extreme estimates come from the GPM 2X2 and 2X1, but not the 1X1, which is equivalent to the BERCS model. This interesting - because it shows that the GPM 1X1 is able to explain these high affinity interactions - without taking extreme dissociation constants like those estimated by the XPR fits (see Table. 4.6) . This dovetails with the discussion of Fig. 8.17, where it was noted that the GPM 1X1 model deals with the very small rates of total dissociation throughout the dissociation phase of the patient data by giving extremely high values for  $T_{ib}$ .

Both the competitive (2X1) and non-competitive (2X2) version of the GPM generally received better AICs (Fig 8.15). Some of their fits gave more unrealistic parameter

	LT79 vs A2	LT43 vs B7	LT36 vs A2	LT36 vs A2 (repeat)	LT21 vs A2	LT16 vs A1	LT40 vs B40:02
$k_{a1}(M^{-1}s^{-1})$	2.6E-5	2.2E-4	5.3E-4	6.7E-4	1.5E-4	1.2E-4	1.3E-4
$k_{d1}(s^{-1})$	5.9E+2*	2.2E-3*	9.6E-2	6.6E-1	1.7E+4*	7.7E+1	2.8E+1
$\beta \text{ Mnm}^{-2} \text{ pg}^{-1}$	2.3E+7*	1.0E+1*	1.8E+2	9.8E+2	1.1E+8*	6.2E+5	2.2E+5
$R_1(\text{pgnm}^{-2})$	8.0E+2*	3.7E+2	1.7E+3	5.6E+2	4.7E+3*	2.4E+2	2.1E+2
$R_2(\text{pgnm}^{-2})$	1.5E+3*	3.3E+2	1.4E+3	4.8E+2	5.6E+2*	3.7E+2	2.7E+2
$R_3(\text{pgnm}^{-2})$	1.2E+3	2.9E+2	1.2E+3	4.1E+2	1.1E+3*	4.5E+2	3.0E+2
$R_4(\text{pgnm}^{-2})$	3.6E+3*	3.2E+2	1.0E+3	3.7E+2	9.6E+2*	5.8E+2	3.8E+2
$R_5(\text{pgnm}^{-2})$	7.4E+1*	3.5E+2	9.5E+2	3.6E+2	8.3E+2*	8.5E+2	4.4E+2
$R_6(\text{pgnm}^{-2})$	-	3.7E+2	9.7E+2	3.8E+2	4.5E+3*	9.0E+2	-
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.3E-3*	2.3E-4	1.3E-3	4.1E-4	2.1E-5	4.1E-4*	4.2E+2*
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	8.0E-1*	1.0E-3	2.1E-3	9.2E-4	8.7E-5	3.7E-3*	1.2E+0*
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.6E+1*	6.1E+1*	3.1E-3	1.8E-3	2.3E-2*	8.7E+2*	1.7E+8*
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	7.0E+3*	1.2E+5*	5.8E-2*	2.7E+2*	5.7E+0*	2.1E+8*	4.3E+7*
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.1E+6*	6.9E+7*	1.1E-1*	7.7E+9*	6.8E-2*	7.0E+8*	3.2E+7*
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	-	5.5E+9*	5.5E-1*	7.1E+14*	1.9E+0*	6.6E+14*	-
$T_{1b}$	2.5E+0*	2.8E+0*	1.8E-3*	2.0E-4*	1.9E+1*	6.7E-6*	5.8E-1
$T_{2b}$	5.2E+0	5.3E+0*	3.3E-2	2.6E-3*	3.0E+0*	1.1E+0	1.8E+0
$T_{3b}$	6.9E+0	4.7E-1*	8.2E-2	5.2E-2	1.5E+1*	1.8E+0	1.9E+0
$T_{4b}$	1.8E+1	5.7E-1*	1.4E-1	6.3E-2	9.4E+0*	2.7E+0	2.2E+0
$T_{5b}$	1.2E+0*	6.5E-1*	2.3E-1	3.5E-1	8.7E+0*	4.1E+0	1.7E+0
$T_{6b}$	-	9.8E-1*	3.4E-1	2.6E-1	4.0E+1*	4.8E+0	-
$S_{1b}$	-	-	-	-	-	-	-
$S_{2b}$	-	-	-	-	-	-	-
$S_{3b}$	-	-	-	-	-	-	-

	LT40 vs B27	LT43 vs B40:01	LT43 vs B40:02	LT43 vs B27	LT42 vs B27	LT42 vs B40:02	LT2 vs B27
$k_{a1}(M^{-1}s^{-1})$	1.6E-4	4.5E-6*	1.5E-7*	5.6E-8*	1.4E-4	2.4E-4	2.1E-4
$k_{d1}(s^{-1})$	5.4E+2	3.5E+2*	1.5E+0*	1.6E+0*	3.6E-1	2.5E-1	5.2E+0
$\beta \text{ Mnm}^{-2} \text{ pg}^{-1}$	3.5E+6	7.7E+7*	9.8E+6*	2.9E+7*	2.6E+3	1.0E+3	2.5E+4
$R_1(\text{pgnm}^{-2})$	1.3E+2	7.6E+3*	1.2E+5*	1.4E+6*	1.3E+3	2.6E+3	8.9E+2
$R_2(\text{pgnm}^{-2})$	1.9E+2	5.4E+3*	3.9E+5*	3.2E+6*	1.2E+3	2.4E+3	5.3E+2
$R_3(\text{pgnm}^{-2})$	1.7E+2	6.4E+1*	1.6E+4*	3.9E+0*	1.2E+3	2.2E+3	7.3E+2
$R_4(\text{pgnm}^{-2})$	3.6E+2	-	-	-	1.3E+3	2.3E+3	9.8E+2
$R_5(\text{pgnm}^{-2})$	7.6E+2	-	-	-	1.1E+3	2.3E+3	8.8E+2
$R_6(\text{pgnm}^{-2})$	-	-	-	-	1.2E+3	2.5E+3	-
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.8E+1*	7.6E-3*	1.1E-5*	1.3E-2*	8.1E-3*	2.2E-3	2.2E+1*
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.6E+5*	4.9E-2*	8.8E-6	9.6E-6	1.3E-2*	6.1E-3	2.4E-2*
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.3E+8*	1.7E-1*	1.7E+0*	4.3E-4*	6.1E+0*	7.9E-1*	1.4E-2*
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.1E+11*	-	-	-	9.1E+0*	3.8E+0*	2.9E-3*
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.5E+14*	-	-	-	1.2E+1*	6.0E+9*	1.2E-2*
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	-	-	-	-	8.1E+9*	1.2E+7*	-
$T_{1b}$	3.8E-1	3.9E+0*	-	2.8E+1	2.7E-4*	4.5E-6*	1.3E-2*
$T_{2b}$	2.2E+0	2.2E+0*	-	1.4E+2	9.2E-2	4.2E-2	3.7E-2*
$T_{3b}$	1.8E+0	1.0E+0*	-	1.4E+0*	6.7E-2	8.5E-2	2.7E-1
$T_{4b}$	4.9E+0	-	-	-	1.4E-1	1.7E-1	1.2E+0
$T_{5b}$	8.3E+0	-	-	-	3.3E-1	1.7E-1	9.6E-1
$T_{6b}$	-	-	-	-	3.8E-1	2.5E-1	-
$S_{1b}$	-	-	1.0E+0*	-	-	-	-
$S_{2b}$	-	-	1.0E+0*	-	-	-	-
$S_{3b}$	-	-	1.0E+0*	-	-	-	-

**Table 8.5:** Parameter estimates with the GPN 1X1 model ran on the sensitized patient sensorgrams.

estimates - that is by giving one antibody in a fit such a high dissociation rate as to make it an unimportant component in the reaction. This is interesting because it suggests that maybe with longer fitting times the single concentration version of the model could have done equally well.

The size differentiated version of the model (GPMSS2X2) produced two distinct groups of

	LT39 vs B37	LT65 vs A2	LT68 vs B57	LT78 vs A2	LT78 vs A2	LT66 vs A2 (repeat)	LT28 vs A2 vs A2
$k_{a1}(M^{-1}s^{-1})$	9.8E-5	5.8E-6*	8.9E-5	4.4E-4	4.2E-4	2.5E-4	1.3E-3*
$k_{d1}(s^{-1})$	1.0E+2*	1.2E+2*	2.3E+2*	3.5E-1	1.3E-1	6.5E+1*	4.9E-3*
$\beta \text{ Mnm}^{-2} \text{ pg}^{-1}$	1.0E+6*	2.1E+7*	2.6E+6*	7.8E+2	3.0E+2	2.6E+5*	3.7E+0*
$R_1(\text{pgnm}^{-2})$	4.7E+2	3.5E+3*	3.0E+2	2.3E+2	5.5E+2	8.3E+2	1.5E+3
$R_2(\text{pgnm}^{-2})$	2.2E+2	4.6E+4*	2.5E+2	2.2E+2	4.4E+2	3.7E+2	8.0E+2
$R_3(\text{pgnm}^{-2})$	1.8E+2	1.9E+3*	3.6E+2	1.7E+2	3.7E+2	1.3E+2	1.2E+3
$R_4(\text{pgnm}^{-2})$	-	2.4E+4	3.1E+2	-	-	-	-
$R_5(\text{pgnm}^{-2})$	-	1.9E+1*	3.8E+3*	-	-	-	-
$R_6(\text{pgnm}^{-2})$	-	-	-	-	-	-	-
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	5.7E-5	1.5E-5	2.4E-2*	5.6E-5	1.3E-4	1.3E-4	5.4E-6
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.2E-4	1.7E-5	9.5E-1*	8.5E-5	2.4E-4	5.0E-4	4.0E-6
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.7E-3*	2.2E-5	3.1E+1*	4.0E-2*	2.0E-3*	1.8E-2*	7.5E-6
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	-	1.8E+2*	7.4E+3*	-	-	-	-
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	-	7.0E+0*	1.1E+6*	-	-	-	-
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	-	-	-	-	-	-	-
$T_{1b}$	-	1.7E-1*	4.5E-2*	-	-	-	-
$T_{2b}$	-	1.6E+1*	3.1E-3*	-	-	-	-
$T_{3b}$	-	1.2E-2*	1.2E-1	-	-	-	-
$T_{4b}$	-	2.1E+1*	1.4E+0	-	-	-	-
$T_{5b}$	-	1.4E-2*	5.3E-1*	-	-	-	-
$T_{6b}$	-	-	-	-	-	-	-
$S_{1b}$	1.0E+0*	-	-	1.0E+0*	1.0E+0*	1.0E+0*	1.0E+0*
$S_{2b}$	1.0E+0*	-	-	1.0E+0*	1.0E+0*	1.0E+0*	1.0E+0*
$S_{3b}$	1.0E+0*	-	-	1.0E+0*	1.0E+0*	1.0E+0*	1.0E+0*

Table 8.5: Parameter estimates from the GPN 1X1 model ran on the sensitized patient sensorgrams (continued).

	LT79 vs A2	LT43 vs B7	LT36 vs A2	LT36 vs A2 (repeat)	LT21 vs A2	LT16 vs A1	LT40 vs B40:02
$k_{a1}(M^{-1}s^{-1})$	9.9E-1*	2.6E-1*	1.1E-3	7.5E-4	1.1E-1*	5.3E-1	1.1E-3
$k_{a2}(M^{-1}s^{-1})$	1.0E+2*	2.1E-2*	5.4E-1*	1.9E-1	1.9E+0*	2.3E-4	5.5E-3
$k_{d1}(s^{-1})$	1.8E+0*	1.8E+1*	6.6E-5	2.0E-2	2.0E-4*	3.5E-1	1.9E-2
$k_{d2}(s^{-1})$	8.2E-3	3.8E-5	1.0E-4	7.8E-4	2.5E+1*	9.3E-3	1.7E-3
$\beta \text{ Mnm}^{-2} \text{ pg}^{-1}$	7.8E-4*	6.8E+0*	5.4E-9*	1.2E+0*	4.9E-1*	5.9E+4	2.0E+2
$R_1(\text{pgnm}^{-2})$	2.1E+4*	4.3E+2	9.7E+2	5.3E+2	3.8E+5*	2.4E+2	1.1E+2
$R_2(\text{pgnm}^{-2})$	1.9E+4*	4.3E+2	7.8E+2	4.9E+2	8.6E+1	3.0E+2	1.6E+2
$R_3(\text{pgnm}^{-2})$	1.2E+3*	3.9E+2	6.4E+2	4.2E+2	3.5E+1	2.2E+2	1.0E+2
$R_4(\text{pgnm}^{-2})$	1.3E+1	4.5E+2	5.6E+2	4.0E+2	6.4E+1	3.6E+2	9.4E+1
$R_5(\text{pgnm}^{-2})$	3.3E+3*	5.0E+2	5.0E+2	4.1E+2	1.6E+3*	5.9E+2	9.5E+1
$R_6(\text{pgnm}^{-2})$	0.0E+0	5.4E+2	5.1E+2	4.6E+2	1.6E+2*	6.5E+2	0.0E+0
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.6E-4	2.4E-2*	6.7E+1*	1.4E-1	1.5E-5	2.6E+0*	9.5E-4*
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.7E-4*	4.8E-2*	2.7E-2*	7.9E-2	3.5E-3*	4.4E-2*	1.5E-4
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.6E-4	1.2E+0*	2.6E-1*	2.5E-1	3.3E-1*	8.6E+2*	3.1E-4
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.0E-4	2.6E-1*	1.7E+0*	2.3E+1*	4.2E-3*	4.9E+5*	6.1E-4*
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.0E-5*	4.9E-1*	4.0E+4*	1.8E+2*	2.3E-3*	1.7E+4*	8.1E-4*
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	3.1E-1*	5.9E+4*	1.8E+0*	3.9E-3*	3.8E+6*	0.0E+0
$T_{1b}$	2.8E+0	1.2E+1*	5.0E-2*	4.2E-2*	1.6E-4*	4.2E-4*	2.5E-1
$T_{2b}$	1.9E+0	4.5E+0*	8.8E+1*	7.4E-3*	1.7E-1*	8.2E-1	1.2E+0
$T_{3b}$	2.5E-1*	3.9E+0*	1.7E-1*	1.3E-1	1.3E+0*	8.0E-1	4.9E-1
$T_{4b}$	3.6E-2	4.7E+0*	6.9E-1	1.8E-1	7.7E-2*	2.5E+0	3.5E-1
$T_{5b}$	3.5E-1*	8.4E-1	9.3E-1	3.2E-1	9.6E-2*	4.9E+0	6.5E-2*
$T_{6b}$	0.0E+0	1.3E+0*	3.8E-2*	2.2E-1	1.1E-2*	6.5E+0	0.0E+0
$D$	5.3E-2	1.4E-2*	1.6E-3*	5.3E-3	1.0E+2*	2.2E+3*	1.6E-1

Table 8.6: Parameter estimates from the GPN 2X1 model ran on the sensitized patient sensorgrams.

	LT40 vs B27	LT43 vs B40:01	LT43 vs B40:02	LT43 vs B27	LT42 vs B27	LT42 vs B40:02	LT2 vs B27
$k_{a1}(M^{-1}s^{-1})$	9.0E-3	1.1E-2*	9.7E-1*	5.0E-3*	3.4E-3*	4.6E-4	4.4E-4
$k_{a2}(M^{-1}s^{-1})$	7.8E-3	2.9E-5	8.0E-8*	1.1E+0*	5.1E-4	2.0E-3	8.1E-3*
$k_{d1}(s^{-1})$	3.3E-2	1.1E-8*	2.2E-4	1.2E-3	3.3E-2*	8.4E-3	1.3E-2
$k_{d2}(s^{-1})$	1.9E-4	4.6E-5	1.1E-3	2.4E-1*	1.2E-4	3.8E-5	5.9E-4
$\beta \text{ Mnm}^{-2} \text{ pg}^{-1}$	4.3E-7*	6.0E-5*	1.4E-5*	3.4E-4*	6.4E-3*	4.9E-7*	1.1E+1*
$R_1(\text{pgnm}^{-2})$	4.3E+1	5.0E+2	1.9E+6*	6.4E+2*	6.7E+2	1.4E+3	7.6E+2
$R_2(\text{pgnm}^{-2})$	5.9E+1	2.3E+2	2.4E+2*	2.0E+3*	6.7E+2	1.3E+3	4.7E+2
$R_3(\text{pgnm}^{-2})$	4.0E+1	1.6E+3*	1.1E+2*	5.7E+2*	7.8E+2	1.4E+3	6.2E+2
$R_4(\text{pgnm}^{-2})$	9.2E+3*	0.0E+0	0.0E+0	0.0E+0	8.1E+2	1.5E+3	4.8E+2
$R_5(\text{pgnm}^{-2})$	2.5E+1	0.0E+0	0.0E+0	0.0E+0	7.2E+2	1.5E+3	4.7E+2
$R_6(\text{pgnm}^{-2})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	6.4E+2	1.7E+3	0.0E+0
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	5.0E-4*	5.6E+1*	9.1E-6	5.3E-6	8.4E-2*	2.9E+2*	1.4E+0*
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.2E-5	8.3E-4*	2.1E-3	1.5E-4*	2.3E-2*	4.6E+3*	5.3E-3*
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	8.4E-5	1.7E+4*	1.3E-3*	2.5E-4*	1.6E-3	7.0E+4*	8.9E-3*
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	7.7E-5	0.0E+0	0.0E+0	0.0E+0	2.6E-3	1.2E+6*	3.1E+0*
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	6.5E-3*	0.0E+0	0.0E+0	0.0E+0	1.6E-3*	1.8E+7*	8.3E-2*
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	1.7E+0*	2.7E+8*	0.0E+0
$T_{1b}$	1.2E+2*	7.6E-1*	1.3E-2*	2.4E-3*	8.1E-1*	5.4E+1*	1.1E-1*
$T_{2b}$	4.0E-1	3.0E-2*	3.2E-2*	3.8E+2*	3.4E-1*	3.5E+0*	3.0E-1*
$T_{3b}$	3.5E+0*	1.1E+0*	1.2E+2*	1.1E+0*	1.0E-1*	9.7E-1	2.5E-1
$T_{4b}$	3.7E+0*	0.0E+0	0.0E+0	0.0E+0	3.1E-2*	4.4E+0*	5.3E-1
$T_{5b}$	2.3E-3*	0.0E+0	0.0E+0	0.0E+0	5.6E+0*	1.8E+0*	3.4E-1
$T_{6b}$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	2.4E+0*	9.0E+0*	0.0E+0
$D$	2.2E-1	7.3E+1*	2.7E+2	2.8E+1*	1.4E+1*	4.8E-1*	6.1E-2*

	LT39 vs B37	LT65 vs A2	LT68 vs B57	LT78 vs A2	LT78 vs A2	LT66 vs A2 (repeat)	LT28 vs A2
$k_{a1}(M^{-1}s^{-1})$	6.0E-4*	1.7E+1*	9.9E+0*	2.5E-3*	4.0E-4	6.4E-4	2.2E-2
$k_{a2}(M^{-1}s^{-1})$	1.8E-4	1.8E-5*	3.9E-4	5.2E-4	9.4E-1	9.3E-4	1.2E+1
$k_{d1}(s^{-1})$	1.1E-3	1.5E+3*	2.6E-4	1.2E-2	8.8E-3	3.5E-3	4.2E-2
$k_{d2}(s^{-1})$	5.7E-5	6.5E-4	2.3E-5*	2.0E-4	9.8E-3	4.3E-5	2.6E-4
$\beta \text{ Mnm}^{-2} \text{ pg}^{-1}$	1.8E-4*	4.2E-5*	4.6E-7*	2.9E-3*	1.2E+0*	2.8E-6*	1.2E-7*
$R_1(\text{pgnm}^{-2})$	2.7E+2	7.0E+4*	1.4E+2	1.5E+2	5.6E+3	4.3E+2	3.4E+1
$R_2(\text{pgnm}^{-2})$	1.7E+2	2.4E+3*	1.1E+2	1.7E+2	6.9E+2	2.1E+2	3.2E+1*
$R_3(\text{pgnm}^{-2})$	1.5E+2	1.3E+3*	7.6E+1	1.9E+2	9.1E+2	7.7E+1	1.9E+1
$R_4(\text{pgnm}^{-2})$	0.0E+0	8.4E+2*	3.2E+1	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$R_5(\text{pgnm}^{-2})$	0.0E+0	2.9E+2*	3.1E+3*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$R_6(\text{pgnm}^{-2})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.9E+0*	1.4E-5	3.3E-4*	8.2E-2*	7.4E-4	2.7E-4	6.5E-4
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.5E-4*	2.1E-5	1.6E-2*	2.7E-4	3.7E-4	4.9E-1*	4.4E-6
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.5E-2*	3.6E-5	5.0E-1	2.1E+0*	6.2E-4	1.2E+2*	1.3E-3
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	2.6E+3*	1.3E+1*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	1.5E+5*	8.3E+2*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$T_{1b}$	2.7E-3*	4.4E-1	3.9E+1*	9.5E+1*	4.2E+0	1.6E+1*	5.5E+2*
$T_{2b}$	3.7E+0*	4.1E-5*	2.3E-4*	5.6E-1	2.8E-1	3.0E+1*	8.5E-3*
$T_{3b}$	4.4E-1*	2.0E-2*	1.8E-1*	3.0E-3*	6.0E-1	1.9E+1*	9.2E+2*
$T_{4b}$	0.0E+0	2.6E+1*	1.6E+1*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$T_{5b}$	0.0E+0	9.9E+0*	4.4E-1*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$T_{6b}$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$D$	8.2E+0*	4.0E+0*	6.4E+3	3.8E+0*	2.5E-1	1.9E+0*	8.4E-3

**Table 8.6:** Parameter estimates from the GPN 2X1 model ran on the sensitized patient sensorgrams (continued).

$k_d$  estimates. Interestingly the  $k_d$  estimates for the larger molecules are generally higher than the lower molecules. Intuitively this makes sense because we would expect the larger molecules to be IgM which have lower affinity than IgG.

The distribution of  $k_d$  estimates for the smaller molecules - which presumably would be IgG also take a similar range to those from the monoclonal study (see Fig. 6.28).

	LT79 vs A2	LT43 vs B7	LT36 vs A2	LT36 vs A2 (repeat)	LT21 vs A2	LT16 vs A1	LT40 vs B40:02
$k_{a1}(M^{-1}s^{-1})$	2.3E-04	6.8E-03*	1.4E-03*	9.9E-04	1.3E-01*	7.0E-03	1.8E-03
$k_{a2}(M^{-1}s^{-1})$	2.3E-03	3.1E-04	4.5E-04	1.6E-02	4.1E-04	8.0E-04*	3.7E-04
$k_{d1}(s^{-1})$	9.3E+01	8.4E-02*	1.8E-02	3.7E-03	1.7E-01	8.7E-02*	1.0E-01
$k_{d2}(s^{-1})$	8.8E+07	9.1E-05	4.8E-04	7.8E-02	5.3E-04*	2.1E-02	4.5E-03
$\beta_1 Mnm^{-2}pg^{-1}$	5.4E+02*	4.5E+02*	3.1E+06*	6.0E+03*	6.3E-04*	3.6E+06*	3.7E+03*
$\beta_2 Mnm^{-2}pg^{-1}$	7.5E+02*	2.5E+04*	1.2E+01*	3.6E-01*	9.9E+03*	3.8E+02*	6.3E+03*
$\gamma Mpg^{-1}nm^2$	8.8E+07*	1.2E+02*	8.1E+03*	1.6E+03*	4.1E+02*	3.5E+03*	5.1E+04*
$R_1(pgnm^{-2})$	1.1E+02	4.5E+02	1.7E+03	1.0E+03	2.0E+02	2.4E+02	1.7E+02
$R_2(pgnm^{-2})$	8.5E+01	4.5E+02	1.5E+03	5.1E+02	1.2E+02	1.8E+02	1.7E+02
$R_3(pgnm^{-2})$	4.1E+01	4.4E+02	1.5E+03	4.7E+02	6.1E+01	1.9E+02	1.7E+02
$R_4(pgnm^{-2})$	2.6E+01	5.0E+02	1.6E+03	4.6E+02	7.9E+01	1.7E+02	1.9E+02
$R_5(pgnm^{-2})$	3.2E+01*	4.9E+02	1.3E+03	5.2E+02	7.5E+01	1.8E+02	1.8E+02
$R_6(pgnm^{-2})$	0.0E+00	5.2E+02	1.4E+03	6.1E+02	9.1E+01	1.5E+02	0.0E+00
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	3.4E-05	4.2E-04*	1.2E-01*	1.2E-03	3.0E-05	1.0E-04	4.7E-02*
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	3.8E-05	5.6E-04*	1.1E-02	1.5E-03	1.3E-04*	1.8E-04	1.7E+02*
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	2.9E-05	9.4E-04*	7.3E-01*	2.7E-03	2.8E-04*	1.9E-04	4.9E+02*
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	4.4E-03*	6.9E-04*	4.4E+05*	1.3E-02*	3.0E-04*	3.1E-04	1.5E+04*
$k_{m5}(pgnm^{-2}M^{-1}s^{-1})$	1.2E-05*	1.0E-01*	2.5E+07*	2.3E-01*	2.1E-04*	6.2E-04*	4.0E+06*
$k_{m6}(pgnm^{-2}M^{-1}s^{-1})$	0.0E+00	1.7E-01*	2.7E+11*	5.7E-03*	4.2E-01*	1.3E-01*	0.0E+00
$T_{1b}$	6.8E-01	2.7E+00*	3.5E-01	7.6E-01	7.4E-01*	8.4E-02*	7.4E-01
$T_{2b}$	3.6E-01*	1.4E+00*	2.4E-01	7.2E-02	2.2E-01*	2.7E-01*	1.3E+00
$T_{3b}$	4.6E-01*	5.3E-01*	2.9E-01	2.0E-01	4.4E-01*	7.4E-01*	8.2E-01
$T_{4b}$	1.5E-01*	3.3E-03*	3.4E-01	2.1E-01	8.1E-04*	1.1E+00	5.8E-01
$T_{5b}$	2.0E+00*	1.0E+00	6.2E-04*	4.8E-01	1.6E-03*	2.3E+00	5.8E-02*
$T_{6b}$	0.0E+00	2.7E+00*	1.1E-01	4.1E-01	6.8E-03*	3.3E+00	0.0E+00
$D$	5.4E-01	2.2E+01*	1.8E+00*	7.3E-02	7.5E+00*	3.6E+00	1.9E+00*
	LT40 vs B27	LT43 vs B40:01	LT43 vs B40:02	LT43 vs B27	LT42 vs B27	LT42 vs B40:02	LT2 vs B27
$k_{a1}(M^{-1}s^{-1})$	2.5E-03*	2.5E-04	2.3E-04	2.0E-04	2.3E-01	6.4E-04	3.0E-03*
$k_{a2}(M^{-1}s^{-1})$	6.7E-04*	1.8E-03	2.4E-03	2.5E-03	3.3E-04	1.8E-01	4.3E-04
$k_{d1}(s^{-1})$	1.3E+02	1.4E+03	3.1E+05	7.4E+01	5.5E-03	8.0E-03	2.0E-02
$k_{d2}(s^{-1})$	7.7E-03	2.3E+03	4.0E+04	2.9E+05	3.8E-03	8.1E-03	8.7E-04
$\beta_1 Mnm^{-2}pg^{-1}$	5.2E+09*	4.1E+03*	3.2E+06*	6.9E+00*	1.7E+03*	3.5E+01*	7.2E+03*
$\beta_2 Mnm^{-2}pg^{-1}$	2.6E+04*	5.4E+02*	4.4E+02*	4.0E-01*	8.9E+01*	6.4E+01*	2.1E+01*
$\gamma Mpg^{-1}nm^2$	4.4E+08*	2.3E+03*	4.0E+04*	2.9E+05*	3.5E+04*	3.4E+03*	8.4E+03*
$R_1(pgnm^{-2})$	1.2E+02	5.3E+02*	2.7E+02	6.6E+04*	1.3E+03	2.5E+03	7.5E+02
$R_2(pgnm^{-2})$	1.1E+02	2.8E+02	8.2E+04*	4.1E+01	9.3E+02	2.2E+03	4.5E+02
$R_3(pgnm^{-2})$	9.9E+01	1.0E+03*	3.2E+02	3.5E+03*	7.2E+02	1.8E+03	5.7E+02
$R_4(pgnm^{-2})$	1.4E+02	0.0E+00	0.0E+00	0.0E+00	4.6E+02	1.2E+03	4.5E+02
$R_5(pgnm^{-2})$	1.9E+02	0.0E+00	0.0E+00	0.0E+00	2.0E+02	7.4E+02	3.2E+02
$R_6(pgnm^{-2})$	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.1E+02	4.5E+02	0.0E+00
$k_{m1}(pgnm^{-2}M^{-1}s^{-1})$	1.8E-02*	2.2E-05*	2.0E-05	7.8E-06	1.1E-03	2.2E-03	7.0E-03*
$k_{m2}(pgnm^{-2}M^{-1}s^{-1})$	2.3E+03*	1.9E-05	7.8E-06	1.3E-04	1.2E-03	3.3E-03	7.3E-04*
$k_{m3}(pgnm^{-2}M^{-1}s^{-1})$	8.0E+06*	6.5E-05*	6.6E-06	1.7E+00*	1.5E-03	4.9E-03	1.0E-03*
$k_{m4}(pgnm^{-2}M^{-1}s^{-1})$	2.2E+07*	0.0E+00	0.0E+00	0.0E+00	2.0E-03	7.0E-03	2.7E-03*
$k_{m5}(pgnm^{-2}M^{-1}s^{-1})$	4.3E+13*	0.0E+00	0.0E+00	0.0E+00	1.9E-03	8.4E-03	7.3E-02*
$k_{m6}(pgnm^{-2}M^{-1}s^{-1})$	0.0E+00	0.0E+00	0.0E+00	0.0E+00	2.3E-03	1.1E-02	0.0E+00
$T_{1b}$	6.4E-01	3.2E+00*	2.2E+02*	5.9E-01	1.5E-01	5.8E-02	4.9E-01
$T_{2b}$	9.9E-01	1.0E-02*	1.1E-02*	3.0E-01*	6.5E-02	2.4E-02	3.3E-01
$T_{3b}$	5.6E-01	4.8E-01*	1.7E+02*	5.5E+00*	5.9E-02	1.1E-02	3.7E-01
$T_{4b}$	1.0E+00	0.0E+00	0.0E+00	0.0E+00	4.7E-02	2.5E-06*	6.5E-01
$T_{5b}$	9.1E-01	0.0E+00	0.0E+00	0.0E+00	1.1E-01	6.5E-03*	2.1E-01*
$T_{6b}$	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.2E-01	8.8E-03*	0.0E+00
$D$	5.3E-01*	1.3E+00*	7.5E-01*	7.1E-02	5.6E+00	1.4E-01	2.8E+00*

**Table 8.7:** Parameter estimates from the GPN 2X2 model ran on the sensitized patient sensorgrams.

	LT39 vs B37	LT65 vs A2	LT68 vs B57	LT78 vs A2	LT78 vs A2	LT66 vs A2 (repeat)	LT28 vs A2 vs A2
$k_{a1}(M^{-1}s^{-1})$	2.3E-04	5.8E-06*	2.3E-04	1.4E-04	1.7E-04	5.2E-04	5.0E-02
$k_{a2}(M^{-1}s^{-1})$	1.9E-03	2.6E-05*	2.0E-03	2.1E-03*	9.3E-05	6.8E-02*	4.8E-03*
$k_{d1}(s^{-1})$	5.6E+02	2.5E-01*	1.3E+04	1.2E-02	3.3E-02*	3.6E-05*	2.6E+01*
$k_{d2}(s^{-1})$	5.3E+03	2.4E-01*	1.1E-01	3.7E-02	7.6E-03	1.3E-01*	4.6E-04
$\beta_1 \text{ Mnm}^{-2} \text{ pg}^{-1}$	7.9E+00*	8.2E+06*	1.2E+08*	5.5E+02*	8.8E+02*	9.0E-01*	2.1E+18*
$\beta_2 \text{ Mnm}^{-2} \text{ pg}^{-1}$	1.5E+02*	2.6E+03*	8.2E+05*	1.1E+05*	4.9E+03*	2.5E+05*	1.7E+01*
$\gamma \text{ Mpg}^{-1} \text{ nm}^2$	5.3E+03*	1.0E+01*	1.1E-01*	5.6E+02	4.8E+01*	1.8E+02*	2.1E+11*
$R_1(\text{pgnm}^{-2})$	3.8E+02	2.6E+04*	3.8E+02	7.1E+02	1.5E+03	1.4E+03	1.5E+02*
$R_2(\text{pgnm}^{-2})$	1.8E+02	4.3E+04*	1.8E+02	3.9E+02	1.6E+03	2.9E+02	6.2E+01*
$R_3(\text{pgnm}^{-2})$	1.2E+02	3.8E+03*	2.7E+02	6.6E+02	3.0E+03	1.3E+02	1.3E+02*
$R_4(\text{pgnm}^{-2})$	0.0E+00	5.3E+04*	2.0E+02	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$R_5(\text{pgnm}^{-2})$	0.0E+00	5.4E+03*	2.5E+02*	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$R_6(\text{pgnm}^{-2})$	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	5.6E-05	2.4E-05	3.9E-04	6.4E-04	6.0E-04	3.5E-04	8.3E-04
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.6E-05	1.8E-05	2.7E+00*	8.5E-04*	6.6E-03*	3.6E-04	1.8E+02*
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	6.9E-05	2.4E-05	3.6E+02*	1.2E+03*	4.7E-01*	1.5E-04	6.0E+02*
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+00	5.9E-02*	7.1E+04*	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+00	3.1E-02*	8.9E+07*	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$T_{1b}$	3.1E-02*	9.7E-01*	5.7E-01*	1.3E+00	2.8E-02*	9.9E-01*	2.6E+02*
$T_{2b}$	9.3E-01	6.9E+00*	7.9E-04*	9.4E-04*	4.0E-02*	3.5E-01*	1.8E+03*
$T_{3b}$	2.6E-01*	1.4E-03*	6.6E-01	1.1E+00	7.7E-01	1.7E-02*	9.8E+00*
$T_{4b}$	0.0E+00	2.1E+01*	1.6E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$T_{5b}$	0.0E+00	9.4E-02*	7.8E-03*	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$T_{6b}$	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00
$D$	1.1E+00	4.0E-02*	2.7E-02	1.0E-01*	9.3E-01	2.0E-01*	1.5E-02*

Table 8.7: Parameter estimates from the GPN 2X2 model ran on the sensitized patient sensorgrams (continued).

	LT79 vs A2	LT43 vs B7	LT36 vs A2	LT36 vs A2 (repeat)	LT21 vs A2	LT16 vs A1	LT40 vs B40:02
$k_{a1}(M^{-1}s^{-1})$	2.4E-04	2.0E-04	2.4E-04	2.0E-04	2.5E-04	2.1E-04	2.1E-04
$k_{a2}(M^{-1}s^{-1})$	1.9E-03	2.5E-03	2.1E-03	2.5E-03	2.3E-03	2.5E-03	2.5E-03
$k_{d1}(s^{-1})$	1.1E+1	1.0E+1	1.0E+1	4.1E+1	2.6E+3	6.3E+2	6.6E+4
$k_{d1}(s^{-1})$	1.1E+4	2.7E+2	2.2E+3	1.5E+3	4.3E+9	6.4E+4	6.7E+3
$\beta_1 \text{ Mnm}^{-2} \text{ pg}^{-1}$	6.7E+4*	3.5E+2*	1.6E+0*	3.7E+2*	3.1E+8*	3.4E+4*	1.3E+6
$\beta_2 \text{ Mnm}^{-2} \text{ pg}^{-1}$	4.6E+2*	3.5E+3*	5.5E+1*	2.3E+2*	1.3E+1*	1.0E+2*	7.5E+1
$\gamma \text{ Mpg}^{-1} \text{ nm}^2$	1.1E+4*	2.7E+2*	2.2E+3*	1.5E+3*	4.3E+9*	6.4E+4*	6.7E+3
$R_1(\text{pgnm}^{-2})$	7.0E+4	3.7E+2	1.9E+3	8.4E+2	6.6E+1	3.0E+2	3.7E+2
$R_2(\text{pgnm}^{-2})$	2.1E+2	4.8E+2	1.5E+3	5.0E+2	5.1E+1	1.8E+2	1.8E+2
$R_3(\text{pgnm}^{-2})$	1.3E+2	4.1E+2	1.6E+3	5.8E+2	4.2E+1	3.9E+2	2.0E+2
$R_4(\text{pgnm}^{-2})$	9.3E+1*	4.2E+2	2.0E+3	9.1E+2	2.4E+4*	1.2E+2	2.3E+2
$R_5(\text{pgnm}^{-2})$	3.7E+0*	4.1E+2	2.1E+3	1.5E+3	5.9E+0*	5.6E+2*	1.3E+2
$R_6(\text{pgnm}^{-2})$	0.0E+0	4.5E+2	2.3E+3	9.6E+2	5.1E+1*	6.5E+2	0.0E+0
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.2E-03	3.4E-03*	5.3E-03	2.9E-03	7.5E+1*	3.0E-04	1.7E-03
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.4E-05	1.5E-03	1.2E+2*	1.8E-02*	7.2E+1*	4.7E-04	7.7E-02
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.9E-05	3.1E-03	7.6E+2*	1.8E-02*	1.2E+2*	2.4E-04	7.5E-03
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	3.5E-03*	4.4E-03*	5.7E+6*	5.0E-01*	2.3E+1*	6.2E-03*	4.0E-03
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.4E-05*	7.2E-02*	4.0E+7*	7.0E-02*	2.2E+2*	2.0E-02*	1.3E+2
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	1.3E-02*	9.3E+11*	7.9E-02*	3.6E-03*	9.3E-02*	0.0E+0
$T_{1b}$	2.1E+2*	1.5E+2*	5.7E-07*	4.4E-01*	6.0E-01	7.6E-01	2.9E+0
$T_{2b}$	2.0E-01*	8.1E-01	5.4E-03	6.5E-03	1.2E+0	5.0E-01	1.4E+0
$T_{3b}$	3.3E-01*	8.6E-01	1.8E-01	2.6E-01	5.3E+1*	1.1E+0*	1.3E+0
$T_{4b}$	3.2E+0*	1.9E+0*	4.4E-01	1.0E+0	2.7E+5*	4.4E-01	1.2E+0
$T_{5b}$	2.1E+0*	5.8E-01*	4.9E-01	2.3E+0	4.5E-02*	5.8E+0*	2.8E-03
$T_{6b}$	0.0E+0	1.7E-02*	4.7E-01	8.1E-01	4.3E+0*	8.4E+0	0.0E+0
$D$	7.9E-01	1.7E-01	3.6E-01	4.2E-01	7.1E-02	3.3E-01	1.9E-01

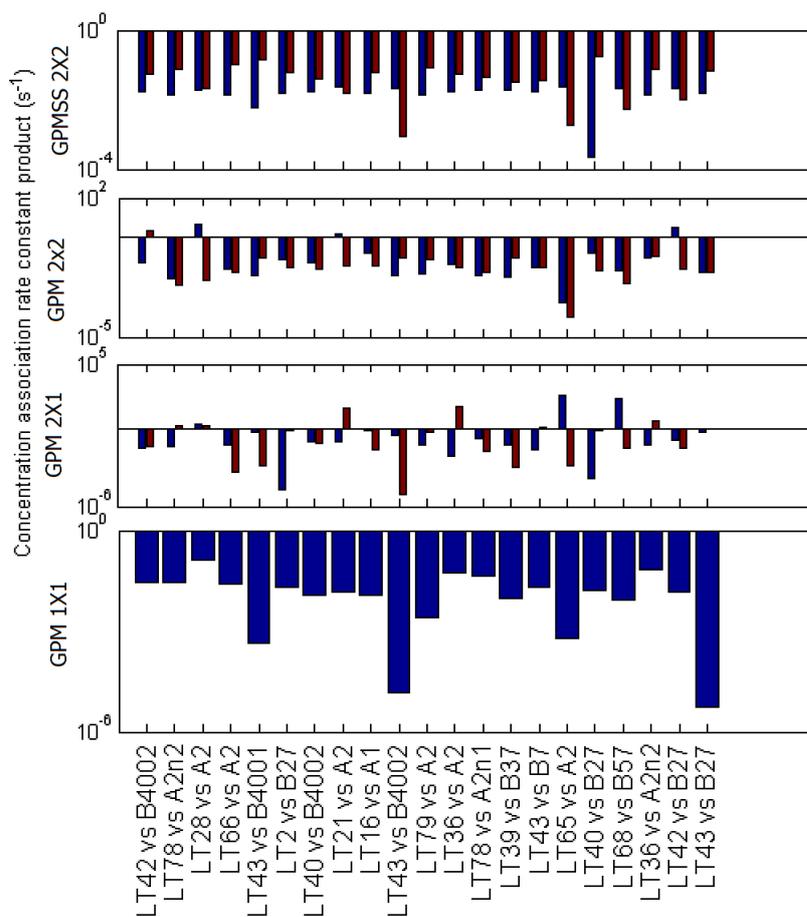
Table 8.8: Parameter estimates from the GPNSS 2X2 model ran on the sensitized patient sensorgrams.

	LT40 vs B27	LT43 vs B40:01	LT43 vs B40:02	LT43 vs B27	LT42 vs B27	LT42 vs B40:02	LT2 vs B27
$k_{a1}(M^{-1}s^{-1})$	2.2E-04	2.5E-04	2.1E-04	2.2E-04	2.3E-04	2.2E-04	2.3E-04
$k_{a2}(M^{-1}s^{-1})$	1.8E-03	1.9E-03	2.5E-03	2.0E-03	2.3E-03	2.3E-03	1.8E-03
$k_{d1}(s^{-1})$	1.2E+5	9.3E+3	6.6E+4	9.3E+5	8.5E+5	1.0E+1	5.8E+1
$k_{d1}(s^{-1})$	1.2E+3	2.6E+5	3.0E+4	2.5E+2	1.0E+3	9.5E+5	4.5E+3
$\beta_1 \text{ Mnm}^{-2} \text{ pg}^{-1}$	4.9E+5*	4.5E+4*	4.5E+4*	2.9E+6*	5.5E+8*	2.4E+1*	9.3E+0
$\beta_2 \text{ Mnm}^{-2} \text{ pg}^{-1}$	4.3E+2*	1.6E+0*	1.3E+1*	9.5E+2*	1.1E+1*	4.7E+2*	1.1E+2
$\gamma \text{ Mpg}^{-1} \text{ nm}^2$	1.2E+3*	2.6E+5*	3.0E+4*	2.5E+2*	1.0E+3*	9.5E+5*	4.5E+3
$R_1(\text{pgnm}^{-2})$	5.3E+1	6.9E+3*	1.9E+6*	5.4E+2*	1.3E+3	1.3E+4	7.6E+2
$R_2(\text{pgnm}^{-2})$	2.0E+4	1.4E+2*	1.5E+3*	4.6E+1	1.2E+3	2.6E+3	5.0E+2
$R_3(\text{pgnm}^{-2})$	6.0E+1	3.5E+1*	5.3E+2*	1.5E+3*	1.1E+3	2.9E+3	5.5E+2
$R_4(\text{pgnm}^{-2})$	6.2E+1	0.0E+0	0.0E+0	0.0E+0	1.2E+3	2.5E+3	8.2E+2
$R_5(\text{pgnm}^{-2})$	1.9E+2*	0.0E+0	0.0E+0	0.0E+0	9.6E+2	2.6E+3	5.2E+2
$R_6(\text{pgnm}^{-2})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	1.1E+3	2.1E+3	0.0E+0
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	7.2E-05	1.5E-05	9.6E-06	1.3E-05	2.1E+0*	2.4E-02	2.6E-03
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	2.8E-03	2.0E-05	3.1E-03*	8.1E-01*	3.5E+3*	5.5E-03	9.7E-04
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	4.3E-05	1.9E+0*	1.7E-02*	2.6E-01*	1.0E+4*	8.4E-03	2.7E-03
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.3E-04	0.0E+0	0.0E+0	0.0E+0	2.8E+6*	1.6E-02	1.5E-02
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	5.6E-02*	0.0E+0	0.0E+0	0.0E+0	7.0E+9*	1.0E-02	2.1E+0
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	4.3E+11*	2.3E+2*	0.0E+0
$T_{1b}$	1.2E-01	2.0E-01*	1.2E-01*	1.7E+0*	1.4E-01	3.8E+0*	4.0E-02
$T_{2b}$	4.0E+2*	3.4E-05*	2.7E-01*	2.6E+2*	8.8E-05*	5.1E-02*	2.0E-01
$T_{3b}$	8.9E-01	1.7E-01*	1.1E+0*	1.0E+0*	6.5E-02	1.6E-01	1.1E-01
$T_{4b}$	1.4E+0	0.0E+0	0.0E+0	0.0E+0	8.7E-02	5.9E-02*	1.3E+0
$T_{5b}$	6.9E+0*	0.0E+0	0.0E+0	0.0E+0	2.5E-01	1.2E-03*	5.3E-01
$T_{6b}$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	4.5E-01	3.1E-02*	0.0E+0
$D$	9.5E+1	3.2E+0*	3.7E-03*	5.0E-01*	4.4E-02	3.1E-01	5.3E-01

	LT39 vs B37	LT65 vs A2	LT68 vs B57	LT78 vs A2	LT78 vs A2	LT66 vs A2 (repeat)	LT28 vs A2 vs A2
$k_{a1}(M^{-1}s^{-1})$	2.1E-04	2.2E-04	2.0E-04	2.2E-04	2.1E-04	2.5E-04	2.1E-04
$k_{a2}(M^{-1}s^{-1})$	2.0E-03	2.5E-03	2.4E-03	2.5E-03	2.3E-03	1.9E-03	2.1E-03
$k_{d1}(s^{-1})$	2.5E+5	3.1E+1	2.8E+3	3.1E+1	8.3E+1	9.3E+3	1.5E+4
$k_{d1}(s^{-1})$	2.3E+3	2.0E+2	4.4E+1	2.6E+1	4.6E+3	2.4E+4	1.7E+4
$\beta_1 \text{ Mnm}^{-2} \text{ pg}^{-1}$	1.7E+6*	5.2E+1*	3.8E+5*	6.7E+1*	4.4E+2*	1.0E+4*	6.4E+5
$\beta_2 \text{ Mnm}^{-2} \text{ pg}^{-1}$	6.9E+0*	5.4E-02*	2.1E+2*	7.4E+2*	2.7E+1*	8.8E+1*	1.4E+2
$\gamma \text{ Mpg}^{-1} \text{ nm}^2$	2.3E+3*	2.0E+2*	4.4E+1*	2.6E+1*	4.6E+3*	2.4E+4*	1.7E+4
$R_1(\text{pgnm}^{-2})$	4.4E+2	7.1E+4*	2.8E+2	5.1E+2	1.1E+3	1.1E+3	3.2E+2
$R_2(\text{pgnm}^{-2})$	8.0E+1	4.8E+4*	2.1E+2	5.7E+2	9.3E+2	6.4E+2	5.0E+1
$R_3(\text{pgnm}^{-2})$	8.8E+1	1.9E+4*	3.3E+2	5.7E+2	5.7E+2	8.3E+1	3.9E+1
$R_4(\text{pgnm}^{-2})$	0.0E+0	8.1E+2*	2.6E+2	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$R_5(\text{pgnm}^{-2})$	0.0E+0	6.5E+2*	4.7E+4*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$R_6(\text{pgnm}^{-2})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$k_{m1}(\text{pgnm}^{-2}M^{-1}s^{-1})$	6.7E-05	1.1E-05	2.7E-04	1.1E-04	5.5E-04	2.8E-04	6.5E-06
$k_{m2}(\text{pgnm}^{-2}M^{-1}s^{-1})$	1.6E-02*	1.1E-05	8.8E-01*	1.5E-04	3.7E-04	5.0E-04	9.0E-05
$k_{m3}(\text{pgnm}^{-2}M^{-1}s^{-1})$	6.6E+3*	1.1E-05	2.0E-01*	1.8E-03	1.7E-03	4.4E-02*	3.5E-01
$k_{m4}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	1.2E-02*	1.1E+4*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$k_{m5}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	1.7E-06*	8.9E+4*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$k_{m6}(\text{pgnm}^{-2}M^{-1}s^{-1})$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$T_{1b}$	1.3E-02*	1.5E-03*	4.9E-04*	4.7E-04*	1.2E+0	4.1E-01	5.2E-02
$T_{2b}$	4.1E+1*	6.9E+0*	4.9E-04*	3.4E-02*	8.1E-01	9.6E-01	7.6E-01
$T_{3b}$	4.3E+0	2.2E-02*	6.5E-01	1.3E+0	7.6E-01	1.2E-02*	2.6E-01
$T_{4b}$	0.0E+0	9.9E-05*	1.7E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$T_{5b}$	0.0E+0	2.9E+0*	5.1E+1*	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$T_{6b}$	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0	0.0E+0
$D$	1.8E-01	7.4E-03*	2.3E-02	2.2E-01	4.7E-01	9.4E-01	1.1E-01

**Table 8.8:** Parameter estimates from the GPNSS 2X2 model ran on the sensitized patient sensorgrams (continued).

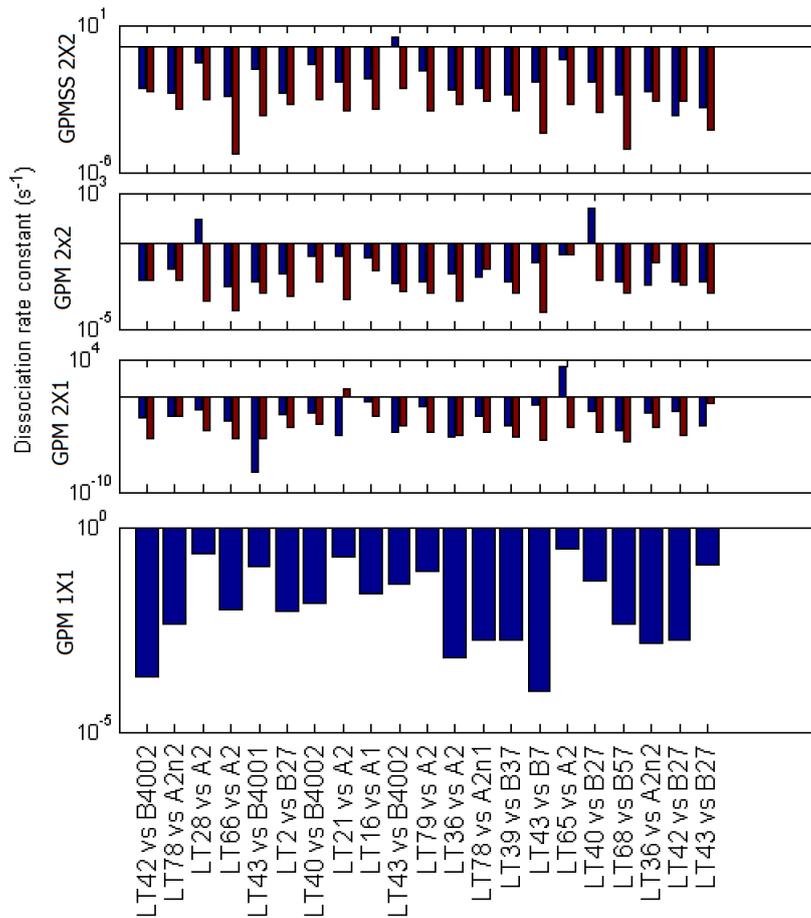


**Figure 8.18:** Bar charts of estimates of the dissociation rate constant for each model, for each of the patient-antigen pairings

It is harder to make comparisons based on the association rate constants, because they were only estimated as part of the association rate constants, concentration product, due to the unidentifiability of those parameters.

As a result even though these products a wide range of values in every case this could be explained by antibodies having association rate constants within an expected range, either in very high concentration, or low concentration respectively.

Interestingly these products are much lower for the large antibody in the size structured

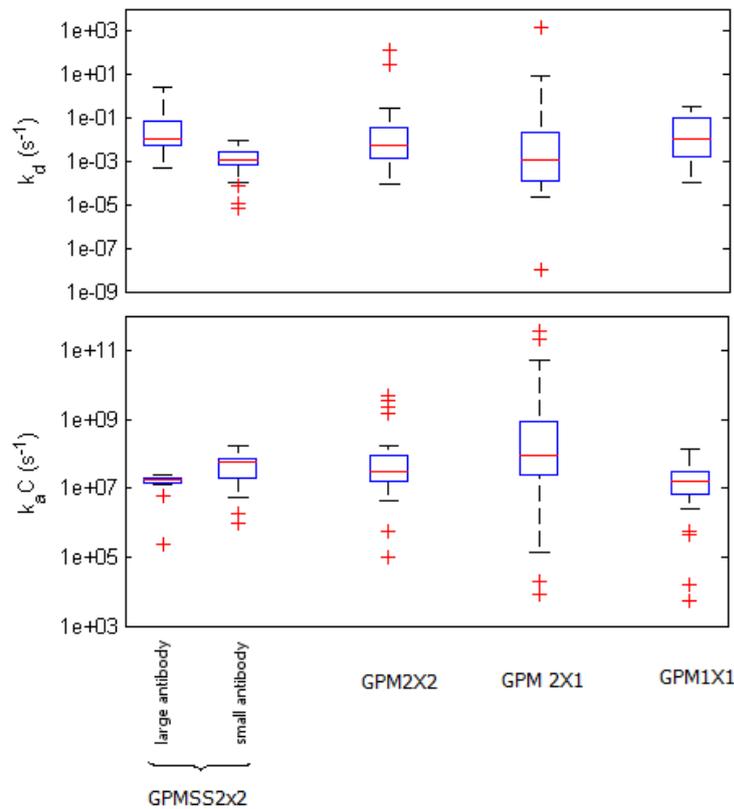


**Figure 8.19:** Bar charts of estimates of the association rate constant concentration product for each model, for each of the patient-antigen pairings

model. This may correspond to fact that IgM will only make up a much smaller portion of the antibody in the body than IgG - especially in the secondary immune response.

Surprisingly the highest estimates of the association rate constant concentration products do not correspond to the highest responses measured on the chips (compare Fig. 4.18). Illustrates a weakness of the fitting process.

The antigen on each chip was bound there by the same process, and although the chips surface is heterogeneous we should not expect unlimited variation in the parameters



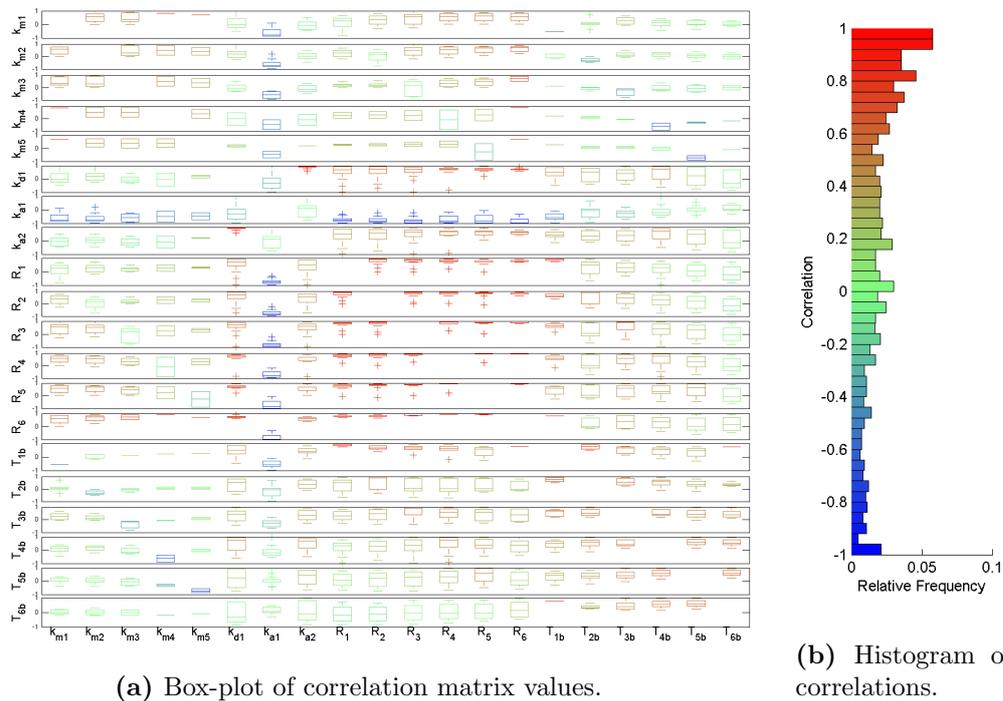
**Figure 8.20:** Box plots of parameter estimates for the patient data set with the polyclonal models

representing its concentration ( $R_m$  and  $R_b$ ).

### 8.5.1 Correlation of parameters

Whilst FACSIMILE fits the models it also analyses the correlation between the parameters used in fitting. At the end of the fitting process, it outputs a matrix showing the correlation coefficient for each pair of parameters.

Similar to Fig. 6.30 box-plots were made for the GPM1X1 2X1 and 2X2 as well as the GPMSS2X2. These show the distribution of the correlation coefficients for each pair



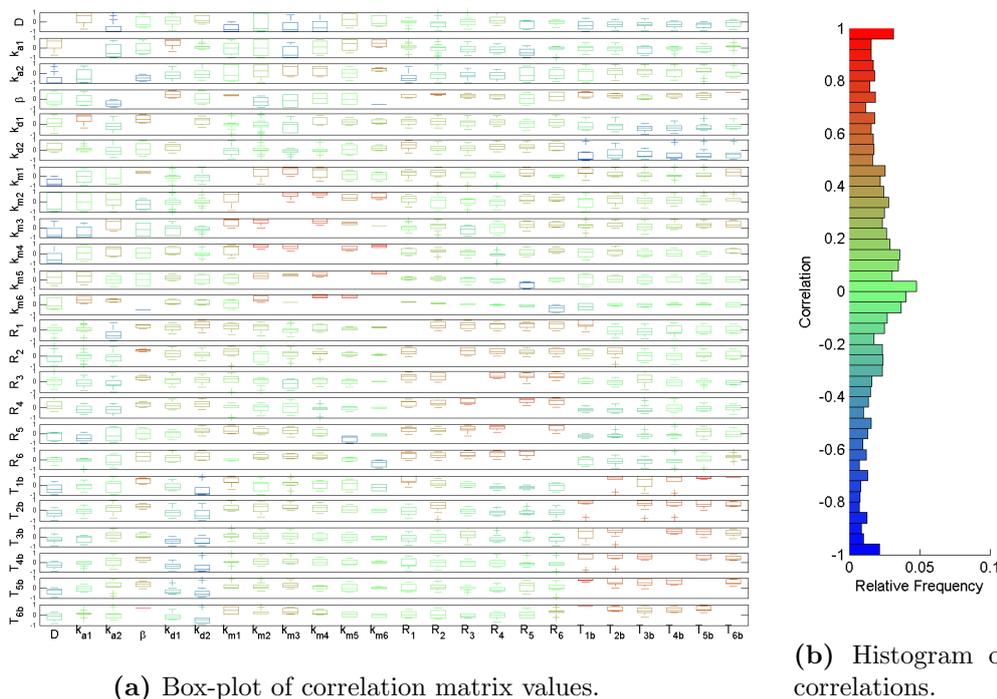
(a) Box-plot of correlation matrix values.

(b) Histogram of correlations.

**Figure 8.21:** Graphs of parameter correlations for the GPM1X1 fitted to the sensitised patient sensorgrams.

of parameter estimates from Tables 8.5-8.8. These box-plots are shown in Figs. 8.21-8.24, along with histograms showing the frequency of correlation coefficients across all pairs.

Figures 8.21(b)-8.24(b) illustrates the way the correlations vary between the models. The parameters of the GPM 1X1 model show much higher correlation with each other than for any of the other three models. Interestingly despite the fact that this model is both indistinguishable and based on the same assumptions as the BERCs model a much higher proportion of the parameter pair correlations take values above 0.9 (see Fig. 6.30). Together with the poorer fits that came from this model (Fig. 8.17), particularly that this model fits the data less well than the BERCs model did the monoclonal data (see Fig. 6.27), these seem indicative that the model is badly suited to the data.

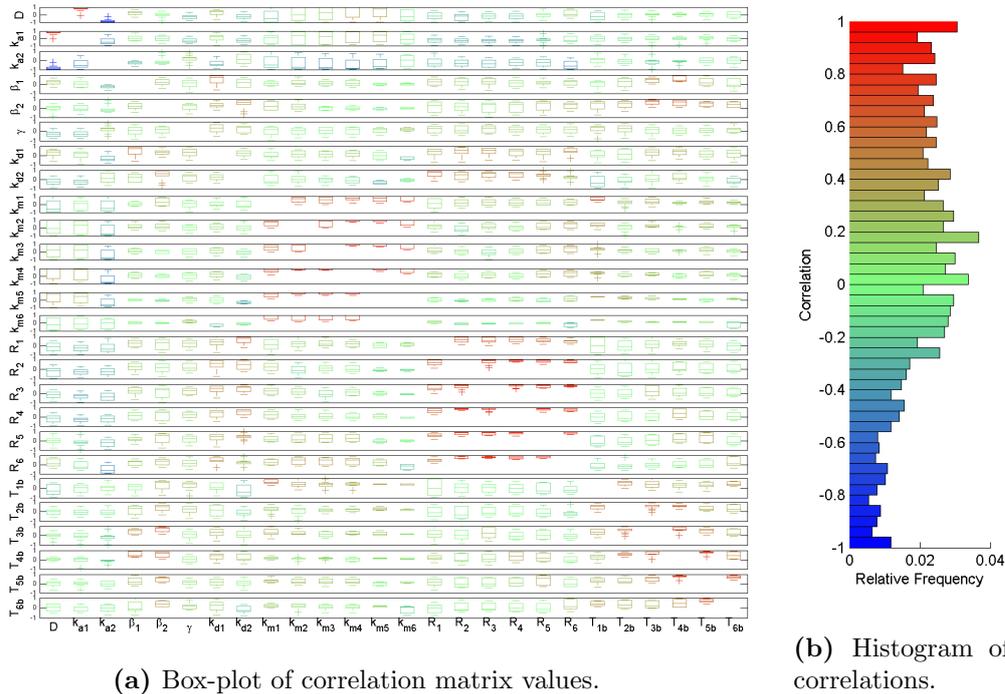


**Figure 8.22:** Graphs of parameter correlations for the GPM2X1 fitted to the sensitised patient sensorgrams.

The remaining three models have their pair correlations clustered to varying degrees around 0. that in the fitting of the sensorgrams a large number of the parameter pairs had high correlations. Comparing this with Fig. 6.30 (a) we see that there are consistently high correlations within the  $R_i$  parameters, and within the  $T_{iB}$  parameters. Interestingly the parameter  $k_{a1}$  has a consistently negative correlation with most of the other parameters.

This makes sense, parameters that represent the same quantity but for a different lane are correlated across lanes.

These box-plots suggest that there may be ways to reduce the complexity of the model. For example, the parameters  $R_i$  could be combined into a single parameter  $R$  representing the maximum density of analyte that can attach to the ligand on the sensor chip in  $\text{pg}^{-1}$



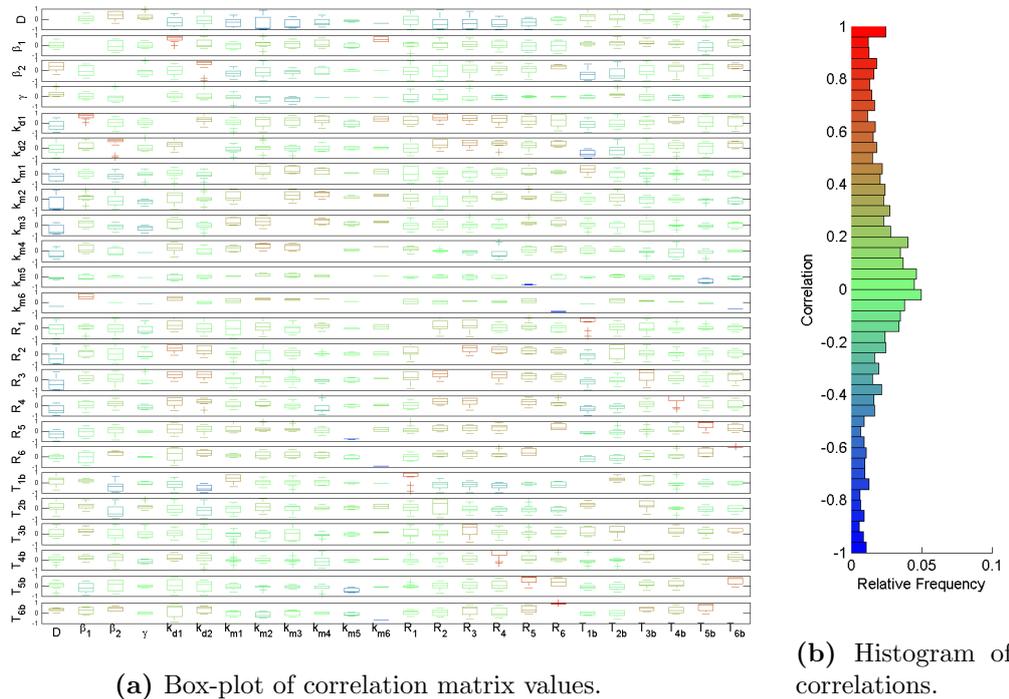
(a) Box-plot of correlation matrix values.

(b) Histogram of correlations.

**Figure 8.23:** Graphs of parameter correlations for the GPM2X2 fitted to the sensitised patient sensorgrams.

$\text{nm}^2$  at each of the interaction spots. The parameters  $T_{iB}$  could be combined to a single parameter  $T_B$ , allowing FACSIMILE to vary the proportion of antigen isolated or non-isolated at each interaction point simultaneously (See Eq. (6.6) ).

However such reductions were rejected. The parameters  $R_i$  and  $T_{iB}$  are only marked as NWD 1/57 and 2/57 times respectively. In part, this is because the BERCS model accurately represents the features of the data, but additionally because each sensorgram contains a large number of measurements, 6 measurements of the change of angle of SPR every 0.9 seconds for close to 2000 seconds. This means for every parameter there are over 600 individual SPR measurements.



**Figure 8.24:** Graphs of parameter correlations for the GPMSS2X2 fitted to the sensitised patient sensorgrams.

## 8.6 Discussion

This chapter opened with an initial study (Section 8.1) where the ERC2C model was found to be more effective than the ERC1 model, both in terms of parameter estimation, and minimising residuals. It also showed that from the selected patient samples both IgM and IgG binding was significant, although IgA binding was slight.

Experiments with plasma effluent and serum samples were fit with the two concentration Langmuir and ERC models (Section 8.2). These showed the same result as the monoclonal antibody studies (Subsection 6.4.1) the ERC assumption resulted in a significant reduction of residuals.

Whilst this fitting process continued models that included more complex features of antibody binding were developed for monoclonal antibody (such as bivalence and the effects of spatially isolated ligand, see Subsections 5.1.5 - 5.1.6). These resulted in more consistent estimates of kinetic rate constants and better fits (Subsections 6.4.2-6.6). This raised the question as to whether models featuring the same features of antibody binding would result in similar improvements for polyclonal antibody.

Experiments were conducted that showed very slow rates of change of signal in the dissociation phase, which lead to the Proteon manager software giving estimates of  $K_D$  from 8.7E-12 to 3.8E-24 (Subsection 4.7), vastly higher than what is thought possible for antibody binding. As a result, the new polyclonal models (GPM and its variations) were developed to include features that could explain this, such as multi-antibody multi-antigen binding leading to the formation of larger complexes, due to patient serum containing mixtures of antibody that bind to different antigen-epitopes.

To validate this hypothesis, further experiments were conducted with mixtures of monoclonal antibody (Subsection 4.8), and the concept of effective  $k_d$  was developed (Section 8.3). The sensorgrams from these experiments were processed in three ways. Firstly the effective  $k_d$  for each interaction spot was analysed (Section 8.3). It was shown that effective  $k_d$  for interaction spots exposed to both antibodies was consistently lower than either single antibody lanes. This demonstrated that the two antibodies were creating stronger bonds together- verifying the multi-antibody multi-antigen structure theory.

Secondly, the sensorgrams were fit with GPM 2X2 (Subsection 8.3.1), which allowed for two antibodies binding to two separate targets on the same epitope. The fitting process was used to estimate the concentrations of the antibody in the mixture. Whilst this produced accurate estimates of the concentrations involved, it produced poor fits to the data. This raised the question, as to whether these poorer fits were the result of the longer integration times, the more complex models required, or some kind of flaw in the

model.

To address this, the sensorgrams were processed in a third way (Subsection 8.3.2). Antibody concentration parameters were fixed at the correct values, and primary rate constants were fixed at the values they were estimated at in Section 6.6. This was done to reduce the number of parameters. The interaction spots chosen for each sensorgram were changed to reduce the number of variables. The resulting fits were improved and showed that the GPM 2X2 mostly captured the behaviour of the antibody mixture interacting with the antigen.

However, abnormalities in the parameters relating to the proportion of antigen available for bivalent binding lead to the investigation of the structure of the HLA-A2 protein and it was found that due to the proximity of the two epitopes SN607D8 and SN230G6 interfere with each others binding.

This again validates the underlying assumptions of the model and shows the value of the model for antibody mixtures.

Finally, the sensitised patient sensorgrams were processed in two ways. Firstly the effective  $k_d$  for each interaction patient antigen pairing was analysed, and it was shown that effective  $k_d$  for the polyclonal samples were generally lower than those for either the monoclonal antibody or the mixtures of monoclonal antibody, suggesting the possibility of multi-antigen multi-antibody complexes.

Patient sensorgrams were then fitted with variations of the GPM and the GPMSS model. The GPMSS 2X2 gave the best fits. However, the fits were not able to lead to accurate estimates of the affinity of the patient antibody involved with the reactions, as the concentration of antibody was unknown, as was the number of antibody, clones and number of antigen-epitopes they were binding to.

This highlights the need for size exclusion chromatography experiments as described in

Section 7.4. Individual size exclusion chromatography experiments for each patient-antigen pairing could help establish whether they have antibody against multiple antigen epitope targets, by testing directly whether it forms multi-antigen multi-antibody complexes when combined with the antigen. This would allow SPR experiments for each patient antigen pair to be analysed with an appropriate form of the GPM model.

The fact that IgM was found to cause significant binding in Section 8.1 and that the separate sizes version of the GPM model was the most successful suggests that IgM may commonly be involved with the reactions. This would make size exclusion chromatography experiments even more effective; not only could the ability of the patient's antibody to form large complexes with the antigen be analysed, but as complexes with IgM antibody will be heavier- patients with exclusively IgM or IgG interactions with the antigen could be identified, as could patients with both. Thus their sensorgrams could be appropriately modelled.

A key barrier to analysing affinity from patient antibody, however, is identifiability. So far in all the models analysed the association rate constant is identifiable if and only if the concentration of the antibody is known. The next chapter of this thesis will focus on new methods that could be applied to this problem using technology similar to SPR.

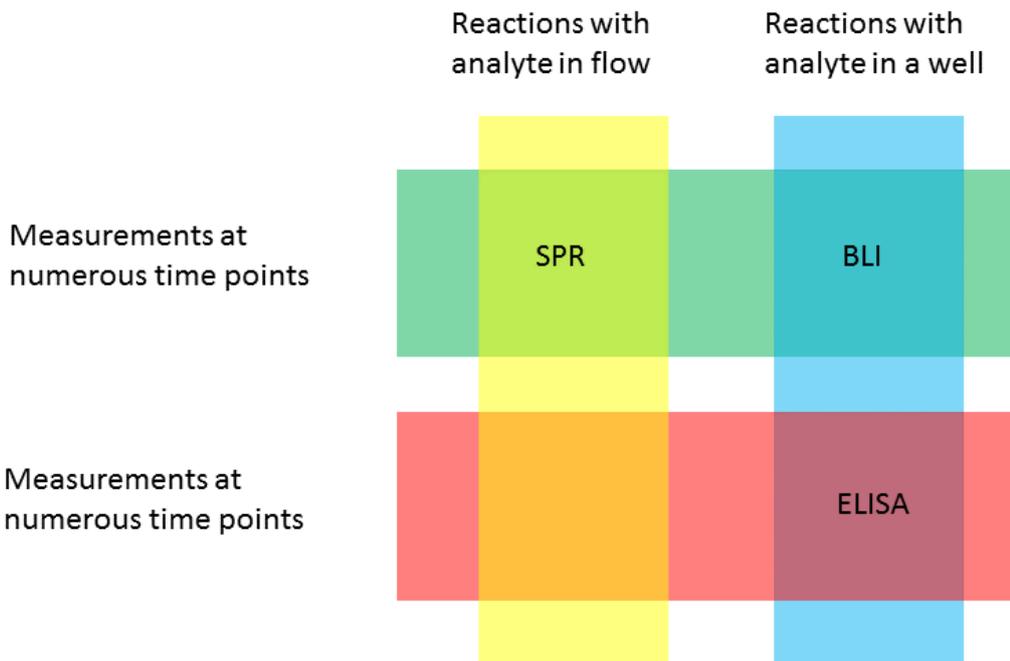
## Chapter 9

# Estimating the affinity of antibody with an unknown concentration

In the previous chapters, we have established that SPR is a useful tool for characterising the interactions between antibody and antigen. However, in cases where the concentration of antibody is unknown affinity cannot be estimated (see Section 5.3 and Section 7.3). It has also been established that this is because of problems with the underlying system as opposed to problems with the way measurements are made. As a result, other experimental systems where antibody-antigen interactions could be measured were modelled with the aim of finding a combination of methods that could be used to make association and dissociation rate constant estimates, even when antibody concentration is unknown.

Whilst SPR experiments have a continuous flow of antibody across a sensor coated with antigen, the other common techniques for measuring antibody-antigen affinity involve a well in which there is a fixed amount of antibody. The relationship between SPR and two other well known techniques is shown in Fig 9.1, The enzyme-linked immunosorbent assay (ELISA) and bio-layer interferometry (BLI). The key difference between these two is that ELISA gives a single time-point per reaction-well as its output, BLI gives measurements.

As ELISA did not require any new equipment to be used its use was investigated first,



**Figure 9.1:** Diagram of the relationship between three common techniques for affinity analysis.

however, due to the large number of reaction steps that ELISA involves, and the variability in the materials required for these reactions it was decided that the number of new parameters any model of it would have to require would be large enough to offset the extra information that could be gained through its use.

## 9.1 ELISA

The simplest ELISA technique, indirect ELISA, is illustrated in Fig. 9.2 (a). Antigen is added to each well of a plate, where it is given time to adhere to the plastic through charge interactions. The well is washed, leaving only antigen that is attached to the well. The antibody being measured is then added and binds to the antigen. The well is washed again leaving only antibody that is bound to the antigen. A secondary antibody conjugated with

an enzyme is added and binds to the antibody being studied. The well is washed again leaving only secondary antibody that is bound in a chain to the antibody being studied and the antigen bound to the well. A substrate is added to the well; it is catalysed by the enzyme, changing the colour of the substrate and giving a readout for the experiment. This readout shows the user whether antibody from the sample reacts with the antigen, but cannot be used to estimate affinity.

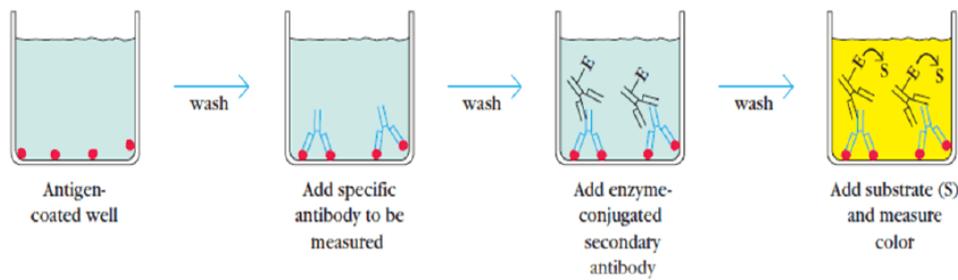
Variations of competition ELISA techniques are used to estimate affinity. The basic competition ELISA technique is illustrated in Fig. 9.2 (b) these add an extra step in which antibody is incubated in solution with the same antigen as is coated on the wells. Indirect ELISA can then be performed on this mixture. In this case, the change in colour resulting is a composite effect of the affinities involved and the amount of free antibody in the incubated solution when it is put through the indirect ELISA.

Larvor *et al.* (1994) uses a variation of competition ELISA in which the incubated antibody-antigen solution is diluted into a concentration far below the equilibrium dissociation constant. This causes the dissociation of the antibody-antigen complex and, at various times after the dilution, indirect ELISA is used to determine the concentration of free antibody. This produces a dissociation curve similar to those made in the dissociation phase of an SPR experiment.

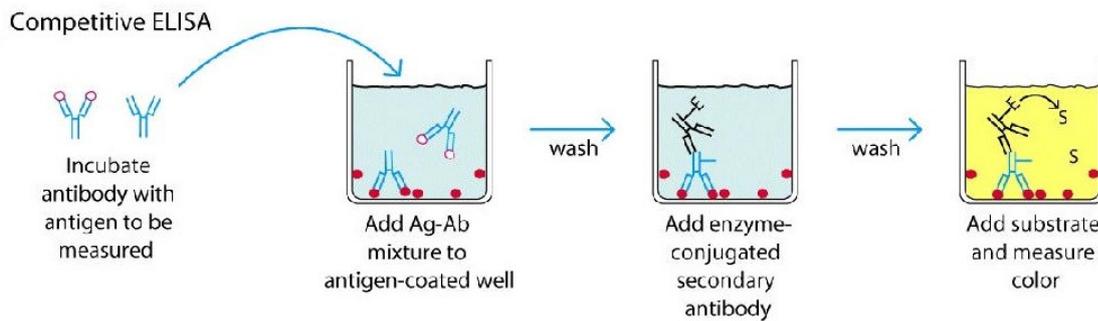
Zhuang *et al.* (2001) uses a variation of competition ELISA where the incubation period is varied, producing a curve similar to the association phase of an SPR experiment.

The barriers to the use of this technique were manifold. For patient antibody, much more complex reactions could be expected, resulting in the kinds of structures illustrated in Fig. 1.8, meaning output would reflect the avidity of antibody and antigen in whichever structures were formed in the incubation well.

It is also unclear what benefits that this dissociation curve would have beyond the SPR



(a) An illustration of the indirect ELISA method.



(b) An illustration of the competitive ELISA method.

**Figure 9.2:** Illustrations of ELISA methods taken from Pokhrel (2015)

dissociation curve, which would come with an association curve and tens of thousands of label-free measurements. In contrast, each measurement in an ELISA based experiment requires another well in which reactions can take place, meaning experiments are comparatively labour intensive. Additionally, the labelling process includes two additional reactions, the secondary antibody and the enzyme catalysing the substrate. Both of these introduce new potential for experimental error. Finally, the structures that should be expected in patient antibody experiments might not allow for each patient antibody to be bound to by a secondary antibody, preventing accurate measurement of the concentration of bound antibody.

However, some of the components of ELISA were modelled and although the technique was not used insights from this modelling process lead to the methods presented in Section 9.2- Section .

If we consider monoclonal antibody  $A$ , being added to a well coated with antigen  $L$ , and assume that only monovalent binding takes place, we have the reaction:



If we define  $R$  as the surface concentration of  $L$  in  $\text{mol nm}^{-2}$ , and  $x(t)$  as the surface concentration of  $AbL$  in  $\text{mol nm}^{-2}$ ,  $A$  as the area on which binding can take place in  $\text{nm}^2$ ,  $V$  as the volume of the solution containing  $Ab$  in  $\text{nm}^3$  then the rate of change of dimer concentration is:

$$\frac{dx}{dt} = k_a \left( C - \frac{A}{V}x \right) (R - x) - k_d x. \quad (9.2)$$

Notably, this is quadratic in  $x$ , whereas for the Langmuir model, Eq. 2.5, the state-space equation is linear. Interestingly the coefficient of  $x^2$  is  $k_a$  the association rate constant which was in the previously mentioned models simply globally unidentifiable. Although this equation misses out many key details of the interactions involved like bivalence and the multiple epitope targets different antibody may have, it suggested that antibody- antigen reactions in enclosed chambers may allow for the identification of  $k_a$ .

The technologies that allow for time series measurements of bound antibody do so directly, by similar effects to SPR (see Section 1.6) so do not introduce the same extra variability as techniques like ELISA. This means that the models developed for SPR can be readily repurposed for these experiments.

One such technology is bio-layer interferometry (BLI).

## 9.2 Comparing BLI and SPR models

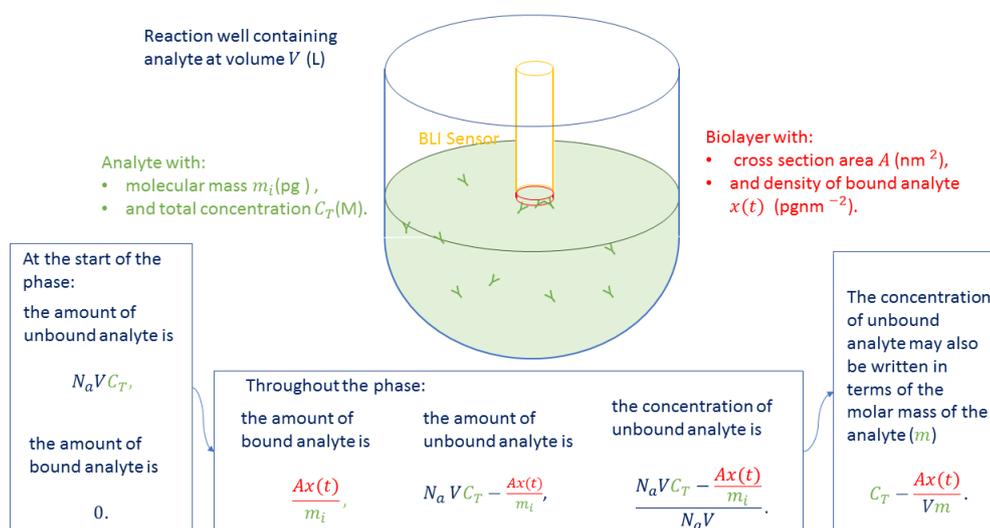
The models used for BLI are the same that are used for SPR (Abdiche *et al.*, 2008), so share structural and parameter identifiability. This is because it is assumed that the difference in experiment design, analyte in a well rather than in flow, does not affect the sensorgram produced.

It is simple to construct experiments on either apparatus where this assumption is valid. However, we will be interested in experimental protocol where the change in concentration due to antibody binding in the well is significant enough to create a measurable change in its rate.

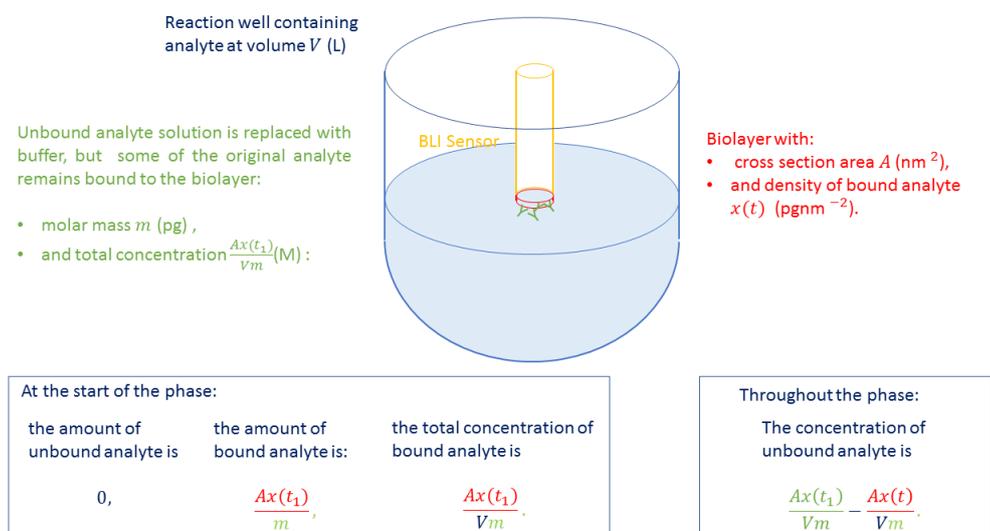
Notably, because SPR experiments rely on an analyte, stored in a well being passed over the surface of the interaction chip, the length of an experiment is limited by the size of the well. Experimental systems like BLI have no such limit. As a result, experiments can last dramatically longer, and as a result observe behaviours that cannot generally be observed with SPR technology can be observed, particularly behaviour much closer to the steady states.

In this thesis, we have used a consistent notation for the concentration of an unbound analyte in SPR experiments at the machines inlet,  $I(t)$ , the input variable. Figure 9.3 shows the derivation of a formula for the concentration of unbound analyte for either phase.

As a result, a simple substitution may be made for the monoclonal models to make them



(a) The association phase.



(b) The dissociation phase.

**Figure 9.3:** Diagram of BLI reaction well and derivation of equations for concentration in either phase Eq. (9.3), in terms of Avogadro's number  $N_a$ .

appropriate for BLI experiments:

$$\begin{aligned}
 I(t) &= W(t) - \frac{A}{V m} x(t), \\
 W(t) &= \begin{cases} C_T & : t \in T_a \\ \frac{A}{V m} x(t_1) & : t \in T_d, \end{cases} \quad (9.3)
 \end{aligned}$$

where  $V$  is the volume of the analyte in  $\text{nm}^3$ ,  $A$  is the area of the reaction spot in  $\text{nm}^2$ ,  $W(t)$  is the new control variable, initial concentration of analyte in the well in the current phase in M,  $m$  is the molar mass of the analyte in  $\text{pg mol}^{-1}$ , and  $x(t)$  is the total density of bound analyte at the time  $t$  in  $\text{pg nm}^{-2}$ , and  $T_a$  and  $T_d$  are intervals on the real line representing the association and dissociation phase, defined as  $T_a = [t_0, t_1]$ ,  $T_d = [t_1, t_2]$  respectively.

In this substitution  $W(t)$  becomes the new input variable, so will generally be relabeled  $I(t)$ . This substitution adds three new parameters to each model:  $m$ ,  $A$  and  $V$ . For antibodies  $m$  is known – 150kDa for IgG and 450kDa for IgM (2.5E-7pg and 1.6E-6pg respectively),  $A$  is always the same 2.8E+11 $\text{nm}^2$ , and  $V$  is determined by the experimenter but takes a minimum value of 4E-5L.

To analyse the impact of this substitution on a model, and these new parameters on the output we analyse the Langmuir model.

### 9.2.1 Langmuir Model

The L1 model was introduced in Example 2.1 in Section 2.2 and expanded on in Subsection 5.1.1 and has been discussed in a number of forms. Although it was not established as a model that was useful for antigen-antibody experiments, it may still be useful for experiments where the monomeric antigen-binding fragments (Fabs) of antibody are reacted with antigen and because of its simplicity it is useful as an illustration. The *fixed concentration*

Parameter	$k_a$	$k_d$	$R$	$t_1$	$t_2$	$m$	$A$	$V$
Value	$6\text{E-}4\text{M}^{-1}\text{s}^{-1}$	$7\text{E-}3\text{s}^{-1}$	$1000\text{pg nm}^{-2}$	1000s	2000s	$2.5\text{E-}7\text{pg}$	$2.8\text{E+}11\text{nm}^2$	$1\text{E-}4\text{L}$

**Table 9.1:** Parameters used in Fig. 9.4

*Langmuir model* (L1FC) is derived by substituting Eq. (9.3) into Eq. (2.5) and (2.6), this gives:

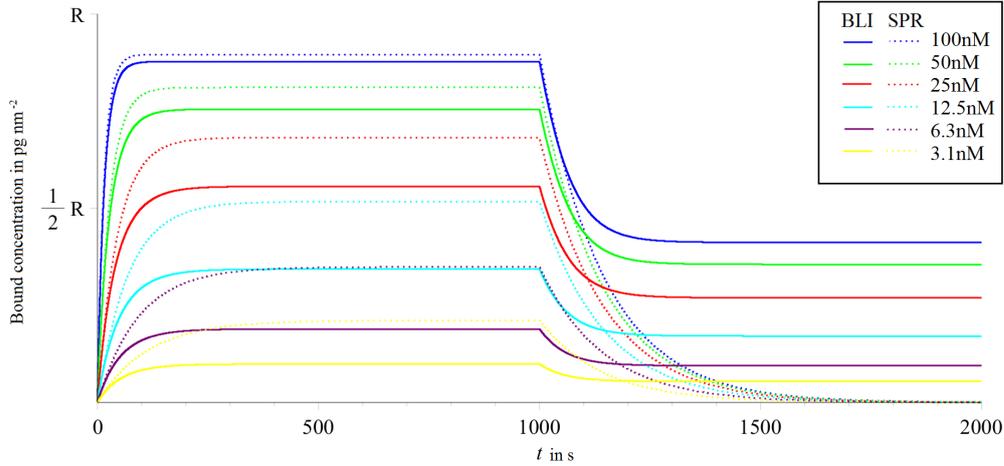
$$\Sigma = \begin{cases} \mathbf{p} & = (k_a, k_d, C_T, R) \\ I(t) & = \begin{cases} C_T & : t \in T_a \\ \frac{A}{V_m}x(t_1) & : t \in T_d, \end{cases} \\ \frac{dx}{dt} & = k_{a1} \left( I - \frac{A}{V_m}x \right) (R - x) - k_{d1}x \\ x(0; \mathbf{p}, I) & = 0; \\ y(t; \mathbf{p}, I) & = \alpha x(t; \mathbf{p}, I). \end{cases} \quad (9.4)$$

Whilst some new parameters are added to the model, these are all known so the models unknown parameters remain the same;  $k_a$ , the association rate constant, measured in  $\text{M}^{-1}\text{s}^{-1}$ , the dissociation rate constant measured in  $\text{s}^{-1}$ ,  $C_T$  the initial concentration of analyte in nM,  $R$  the antigen on the bio-layers capacity for antibody to bind  $\text{pg nm}^{-2}$ .

The output for this model is shown in 9.4, along with the output for the L1 model with the same parameters at 6 concentrations. Notably in this figure, in the dissociation phase we see in each case the amount of bound antibody  $x$  asymptotically tending towards a steady state.

This model differs from the single concentration Langmuir model for SPR in both its state-space and input equations, so we should expect it to be able to identify different parameters than the SPR model.

Due to the comparative simplicity of this model the output equation method may easily be



**Figure 9.4:** Graph of bound analyte  $x_1(t)$ , against time,  $t$  for the L1 and L1FC models, with the parameters shown in Table 9.1

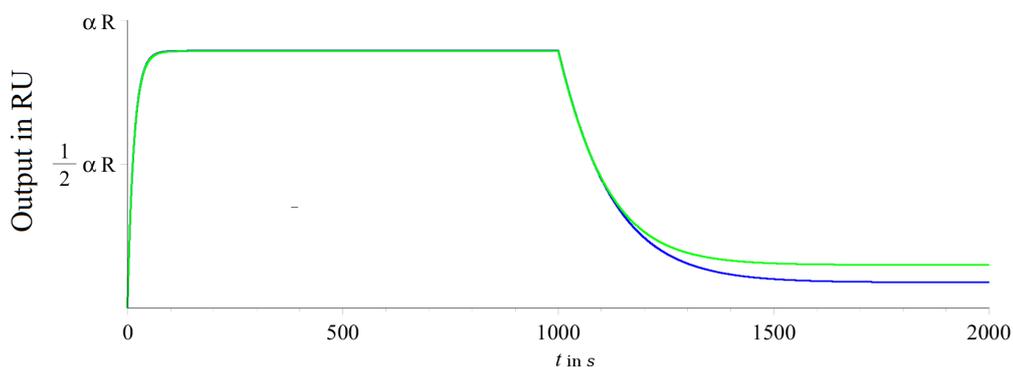
applied (Section 3.1.3), and a monovariant differential polynomial for its output in either phase can be found by substituting the output equation into the state-space equation, to eliminate  $x(t)$ :

$$\frac{d}{dt}y = \frac{1}{\alpha} \frac{k_{a1}A}{Vm} y^2 + \left( -k_{a1}C_T - \frac{k_{a1}AR}{Vm} - k_{d1} \right) y + \alpha k_{a1}C_T R \quad (9.5)$$

$$\frac{d}{dt}y = \frac{1}{\alpha} \frac{k_{a1}A}{Vm} y^2 + \left( -\frac{k_{a1}Ay(t_1)}{\alpha Vm} - \frac{k_{a1}AR}{Vm} - k_{d1} \right) y + \alpha \frac{k_{a1}Ax(t_1)R}{\alpha Vm} \quad (9.6)$$

The fact that  $k_{a,1}$  appears as the only unknown parameter in the coefficient of  $y^2$  means that it is uniquely identifiable. This is significant because this parameter has previously been simply structurally unidentifiable or structurally unidentifiable in every previous model. By the comparison of the remaining coefficients it may be quickly verified that it even when  $C_T$  is unknown all parameters are globally identifiable, i.e. the model is SGI.

This illustrates the value of BLI for use in a clinical setting where samples will be of unknown concentration. Notably in this case the problem of parameter identifiability is solved by changing the underlying system rather than the way in which the observations



**Figure 9.5:** Simulated experiments with analytes that would give identical results on the SPR platform

are made, the output function. Not only does this illustrate the value of BLI it also illustrates the value of making a distinction between the structural identifiability of a model and the simple structural identifiability of the underlying system.

In every model except the Langmuir with transport model, the three parameters  $k_{a1}$ ,  $C_T$  and  $k_m$  have appeared exclusively in products and quotients with each other in the system equations. As a result the association rate constant and the affinity ( $k_{d1}/k_{a1}$ ) of an antibody in a clinical sample has not been identifiable.

The substitution Eq. (9.3) that transforms models to make them appropriate for BLI, breaks up these products and quotients, meaning that in the parameters  $k_{a1}$ ,  $C$  and  $k_m$  can be expected to be simply structurally identifiable in the BLI versions of the models.

This result is illustrated in Fig. 9.2.1. The output of the L1FC model is shown for two parameter vectors 9.2 that would give identical outputs for the L1 model. The difference between these curves is large enough to be seen visually- suggesting that it will also be large enough to allow for parameter identification through model fitting.

The significance of this is that this tells us that BLI could be used to identify the affinity and concentration of an unknown sample of identical Fabs.

CHAPTER 9. ESTIMATING THE AFFINITY OF ANTIBODY WITH AN UNKNOWN CONCENTRATION

Parameter	$k_a$	$k_d$	$R$	$t_1$	$t_2$	$m$	$A$	$V$	$C_T$
Value	$6E+5M^{-1}s^{-1}$ $1.2E+4M^{-1}s^{-1}$	$7E-3s^{-1}$ $7E-3s^{-1}$	$1000pg\ nm^{-2}$ $1000pg\ nm^{-2}$	1000s 1000s	2000s 2000s	$2.5E-7pg$ $2.5E-7pg$	$2.8E+11nm^2$ $2.8E+11nm^2$	1E-4L 1E-4L	100nM 50nM

**Table 9.2:** Paramters used in Fig. 9.2.1

As the Langmuir model does not allow for antibody to bivalently bind the two simulations don't directly show how antibody would behave in the BLI reaction chamber. The simulation shown with the blue dotted line uses the parameters estimated in Section 6.6 for the interaction of the antibody SN230G6 against HLA- A2. In physical terms this may be thought of as how this binding reaction may take place if each IgG molecule had a single fab domain removed. As the analyte cannot bind bivalently, even though it has the same reaction constants it behaves quite differently- rapidly reaching equilibrium in either phase.

When the dissociation phase is began, the solution the reaction chip is exposed to is changed. Although this no new solution has no antibody, a very small amount is bound to the chip - this means the concentration of antibody in the system drops dramatically. In this particular example the concentration of antibody in the system drops by a factor of approximately 2000, as a result the equilibrium reached is very close to 0.

Figure shows a green and blue curve. These represent outputs from analyte that in an SPR experiment would be unidentifiable, as one has an association rate constant twice the other and is at half the concentration of the other sample. In the association phase the curves are nearly identical. They only differ because the blue curve reaches a microscopically lower equilibrium concentration, in the dissociation phase both curves differ widely. This is because previously the higher concentration of the first analyte produced a very similar effect to the higher association rate constant of the second analyte, but in the dissociation phase their concentrations are approximately the same. However, the difference between the two analytes is only apparent because of their high affinities. This raises the question of whether antibody-antigen binding affinities will be high enough for this model to predict

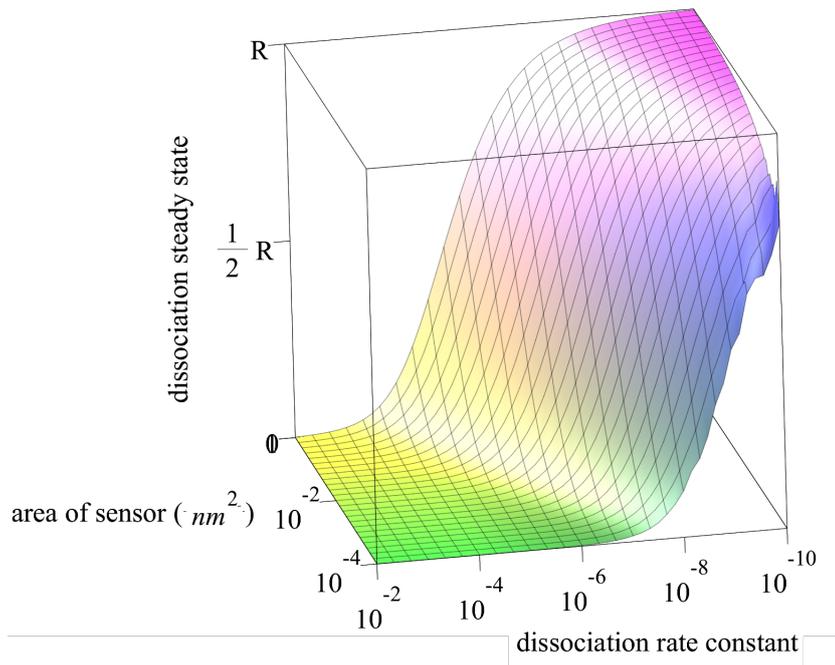
sensorgrams that would be discernibly different in concentrations for which the L2 model would predict identical sensorgrams.

A graph addressing this is shown in Fig. ???. This graph shows how the value of the steady state in the dissociation phase changes as the area of the chip antigen is bound to varies and the dissociation rate constant varies. To illustrate this the other parameters have been chosen to represent a  $400\mu\text{g}$  sample of SN230 antibody at a concentration of 100nM, interacting with A2 antigen.

The value of the steady state that can be reached in the dissociation phase has a sigmoidal relationship with either variable. That is an arbitrarily large choice of either variable puts the value of the dissociation phase steady state arbitrarily close either to 0 or  $R$ . In previous chapters the concept of an effective dissociation rate constant for analyte that undergoes a more complex binding to the ligand has been introduced (see Subsections 5.1.4 and 7.2.4).

As the effective dissociation rate constants of antibody are much lower for these models, this graph suggests that even for very small areas with antigen bound, the steady states in the dissociation phase will be a significant fraction of the total amount of antibody that could bind to the antigen ( $R$ ). This in turn suggests that concentration and as a result association rate constants will be possible to estimate through model fitting.

Figure 9.6 shows the output for two



**Figure 9.6:** An antibody in the proximity of antigen attached by linker to the chip's surface.

### 9.3 Estimating the affinity of single unknown monoclonal antibody

The most successful model from the previous chapters can be redeveloped for the BLI experimental apparatus- the bivalent ERC spacial model. This model was introduced in subsection 5.1.6.

$$\frac{dx_m}{dt} = \frac{k_{a,1}(k_{d,1}(x_1 + x_m) + k_m(I - \frac{A}{V}(x_1 + x_2 + x_m)))}{2k_m - k_{a,1}(x_1 + 2x_2 + x_m)} (R_m - x_m) - k_{d,1}B_m \quad (9.7)$$

$$\begin{aligned} \frac{dx_1}{dt} = & \frac{k_{a,1}(k_{d,1}(x_1 + x_m) + k_m(I - \frac{A}{V}(x_1 + x_2 + x_m)))}{2k_m - k_{a,1}(x_1 + 2x_2 + x_m)} (R_b - x_1 - x_2) - k_{d,1}x_1 \\ & - k_{a,2}x_1(R - x_1 - 2x_2) + 2k_{d,1}x_2 \end{aligned} \quad (9.8)$$

$$\frac{dx_2}{dt} = k_{a,2}x_1(R - x_1 - 2x_2) - 2k_{d,1}x_2 \quad (9.9)$$

$$I(t) = \begin{cases} C & : t \in (t_1, t_2) \\ x_1(t_2) + x_2(t_2) + x_m(t_2) & : t \in [t_2, t_3) \end{cases} \quad (9.10)$$

$$y = \alpha(x_1 + x_2 + x_m) \quad (9.11)$$

Unlike the SPR version of this model, in the BLI version the parameters  $C$  and  $k_{d1}$  are globally identifiable. As a result the model is structurally globally identifiable. This analysis was conducted with the iterative Gröbner method for finding the MVDP.

The effect that varying these parameters has is illustrated in Fig. 9.7 which shows two parameterisations that would be indistinguishable on the SPR equivalent of the model. This graph illustrates that on the BLI experiment the the difference between these curves is large enough to be seen visually- suggesting that it will also be large enough to allow for parameter identification through model fitting.

As with Fig. ??, the simulation shown with the blue dotted line uses the parameters estimated in Section 6.6 for the interaction of the antibody SN230G6 against HLA- A2. Although this model allows for bivalent binding- meaning the antibody in this simulation behave as they have been observed behaving in the experimental work.

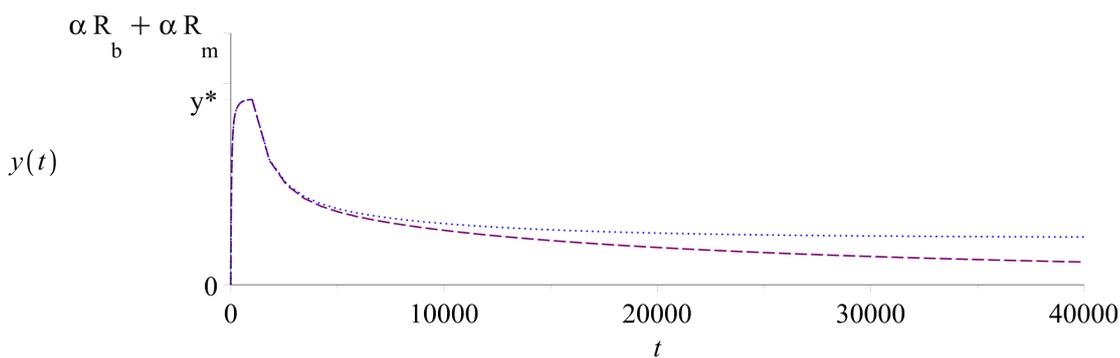
When the dissociation phase is began, the solution the reaction chip is exposed to changes. Although this no new solution has no antibody, a very small amount is bound to the

chip - this means the concentration of antibody in the system drops dramatically. In this particular example the concentration of antibody in the system drops by a factor of approximately 2000, as a result the equilibrium reached is very close to 0.

As the model allows for bivalent binding this dissociation is very slow.

In this diagram a purple dashed curve is also shown. This represents an analyte that in an SPR experiment would be unidentifiable from the one depicted with the blue dotted line, having an association rate constant 10 times higher and being at  $1/10^{th}$  the concentration of the other sample. In the association phase this curve is almost identical to the dotted blue curve. They only differ because the purple dashed curve reaches a microscopically lower equilibrium concentration, in the dissociation phase both curves are quite similar initially - whilst the change in signal is dominated by antibody unbinding from isolated ligand.

As either system reach their quasi steady states their outputs diverge. Like the Langmuir model this is because previously the higher concentration of the first analyte produced a very similar effect to the higher association rate constant of the second analyte, but in the dissociation phase their concentrations are approximately the same.



**Figure 9.7:** Simulations of the BERCs model and BLI version of the BERCs model with parameters averaged from Table 6.6.1

## 9.4 Estimating the affinity of multiple monoclonal antibodies or polyclonal serum

As there are a large number of different ways that polyclonal antibody can interact a general model for antibody and antigen interactions was developed, which allowed an arbitrary number of antibody species to interact competitively or non competitively with an arbitrary number of epitopes on a single species of antigen. Two specific version of this model were used to illustrate different ways the model could function. The simpler of these had its identifiability analysed, and it was found that it was simply structurally unidentifiable. It was also noted that in general the model was simply structurally unidentifiable because it has parameters that only appear in products and quotients in the statespace equations, with other parameters that appear in those quotients.

A BLI version of the general model may be developed by substituting 9.3 into the equations of subsection 7.2. It is anticipated that the general form of the model is simply structurally globally identifiable. This is because there are no longer parameters that only appear in products and quotients in the statespace equations, with other parameters that appear in those quotients, i.e.  $k_{aj}$ ,  $C_j$  and  $k_{mj}$  only appear together . As a result regardless of the output these parameters will be structurally unidentifiable.

The simpler specific version of the model that were previously developed, will now be analysed.

The structural identifiability of the BLI Langmuir form of the model for two antibodies competing over a single epitope can be analysed with the BVDP and Taylor series with limits around the discontinuity. The model has the parameter vector  $\mathbf{p} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_1, C_2, \beta, R, A, V)^T$ . For a second parameter vector  $\bar{\mathbf{p}}$ , the relationship  $\bar{\mathbf{p}} \sim \mathbf{p}$  holds if and only if:

$$\bar{\mathbf{p}} = (k_{a1}, k_{d1}, k_{a2}, k_{d2}, C_1, C_2, \beta, R)^T \text{ or } (k_{a2}, k_{d2}, k_{a1}, k_{d1}, C_1, C_2, \beta, R)^T .$$

That is the model is structurally locally identifiable. As with previous models that were not globally identifiable because their indexes could be swapped we may define  $k_{d1} > k_{d2}$  or a similar inequality to make the model structurally globally identifiable.

Similar to the monoclonal models, parameterisations that would produce identical SPR sensorgrams will produce divergent BLI sensorgrams. Visualisations of these will be similar to Fig. 9.7 in the later part of the dissociation phase. This is because both models allow bivalent binding, but no more complicated binding processes.

Models that include non-competitive antibody will see the higher steady states for realistic parameters, than any other model, as they allow for the development of more complicated complexes which need to go through more stages before antibody is released from the ligand. This results in higher effective dissociation rate constants (compare Fig. 5.11). As a result visual differences between outputs predicted with previously indistinguishable parameterisations will allow accurate parameter estimation.

# Chapter 10

## Conclusions

The overarching goal of this thesis is to develop models and fitting procedures that could be used to accurately estimate antibody-antigen affinity. The success of this goal will be analysed through the objectives set in Section 1.7.

The first objective was to develop models based on the chemical interaction schemes of both monoclonal and polyclonal antibody against antigen, that include the effects of antibody not being well mixed. This was accomplished with the development of the BERC and BERCS model for monoclonal antibody in Chapter 5, and the generalised polyclonal model (GPM) in Chapter 7.

The second and third objectives were to analyse the identifiability and indistinguishability of these models, and assess whether they would be useful for a pharmaceutical or clinical setting; and, if current methods do not work, to develop methods sufficient for this analysis. For most models adapted forms of the methods from the literature worked; although in some cases these adaptations were extensive, particularly the ORC method was adapted so it could be iteratively used with augmented forms of the model that treated observable variables as additional outputs and parameters as constant variables (Section 3.1.2). For the analysis of the most complex models new methods were developed – specifically the BVDP method (Section 3.2.3), and simple identifiability – which allowed for the analysis of the structural identifiability of models without analysing individual parameters (Section 3.2.1).

These methodologies allowed for a structural identifiability analysis of all models. For monoclonal models and parameter identifiability analysis for all of the monoclonal models.

In all cases the sample concentration and association rate constant were unidentifiable; and in cases where a full analysis could be conducted their product was identifiable, and in cases where only simple identifiability could be analysed their product was simply identifiable. This limitation might prevent the use of existing SPR machines with clinical samples where concentration is unknown.

The fourth objective was to participate in designing and conducting monoclonal and polyclonal antibody-antigen SPR experiments that will allow for these models to be tested against each other and those in the literature. This was accomplished and the results are shown in Chapter 4.

The fifth objective was to fit the models to the data, analyse whether they improve goodness of fit, and the consistency of affinity estimates. In the case of monoclonal antibody this was a significant success. BERCs Model fits were achieved dramatically better than those in the literature, and estimates were produced dramatically more consistent than those in the literature (this was discussed in Section 6.7).

In the case of polyclonal antibody the fits of the models to the data were also a significant success. In some SPR sensorgrams made with antibody samples taken from sensitized patients dissociation was vastly slower than it had been for any monoclonal experiment (Subsection 4.7, see also Fig. 8.14). The generalised polyclonal model was developed to explain this slower dissociation via multi-antigen multi-antibody complexes.

This hypothesis was validated through experiments with mixtures of monoclonal antibody with previously measured affinities. In these experiments dissociation was slower than the equivalent monoclonal experiments (Section 8.3), demonstrating that the two antibodies

were creating stronger bonds together- verifying the multi-antibody multi-antigen structure theory.

The appropriate form of the GPM was then fit to this mixed monoclonal data (Subsection 8.3.1 and Subsection 8.3.2). These fits showed that the model could accurately reproduce the sensorgrams. However the estimates of the parameters related to available antigen took unusual values – suggesting there was a mechanism in the binding process that was not in that form of the model this led to communication with a research group investigating the structure of the HLA-A2, who reported that due to the proximity of the two epitopes SN607D8 and SN230G6 interfere with each others binding (Jones *et al.*, 2016, personal communication) (see 8.13).

The fact that parameter estimates were so consistent, that their abnormalities could be used to detect unpublished features of the antigen, provides additional validation to the model and shows the value of the model for antibody mixtures.

Patient sensorgrams were processed with the GPM, which produced extremely accurate fits 8.16, although due to identifiability limitations (established by a previous simple identifiability analysis) these fits could not be used to estimate association rate constants, and thus affinity.

The final objective was that in a case where antibody samples being in unknown concentration limited the ability of SPR to identify antibody-antigen affinity, to investigate alternative experimental methods. This was accomplished in Chapter 9, where models from preceding chapters were adapted for the BLI machine and demonstrated to have association rate constant and concentration as globally identifiable parameters or at least as simply globally identifiable parameters.

## 10.1 Further Work

Due to the interdisciplinary nature of this work, these results create opportunities for further work in a variety of fields.

### 10.1.1 Clinical Research

One of the key results from this thesis is that many sensitised patient/antigen pairings show dynamics in SPR experiments that cannot be explained by single antibody-epitope interactions (Subsection 4.7), and that multiple antibody-epitope crosslinked binding can explain these dynamics (Section 8.3). This result is problematic because it suggests that patient antibody donor antigen interaction is dramatically more complex than previously assumed. Instead of characterising this reaction by affinity, a vector of affinities for each epitope will need to be used.

The next step in measuring patient antibody-antigen affinity is to conduct a study that combines size exclusion chromatography experiments with biosensor experiments using a BLI or SPR platform. The size exclusion chromatography experiments would use the methodology proposed in Section 7.4 to estimate the number of epitope targets involved in the antibody-antigen interaction as well as the class of the antibody involved. The Biosensor experiments would take a form similar to those in Section 4.9. If these were performed using BLI rather than SPR this would also remove the simple identifiability barrier to affinity estimation, and if these were analysed after the results of the size exclusion chromatography experiments this would allow each patient antigen pair to be analysed with an appropriate form of the GPM model. However if a BLI platform was not available, SPR could still be useful for an initial study.

These experiments could be repeated to Establish reproducibility of estimates for different

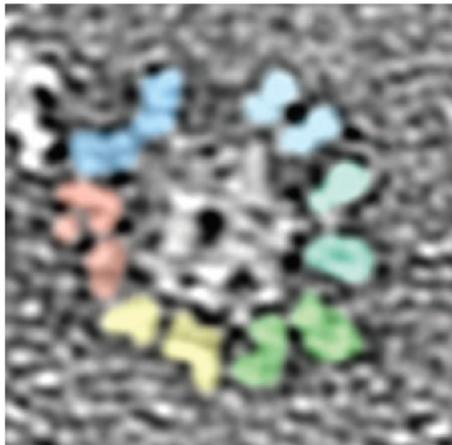
samples from the same patient. In Section 8.3 the hypothesis behind the GPM was tested with mixed samples of monoclonal antibody – additional tests like this could be performed to establish the error of association and dissociation rate constant estimates made with the GPM.

After the methodology of size exclusion chromatography and biosensor experiments is established it could be used to study antibody-antigen affinity for medical and more general purposes. There are three key medical questions could be investigated.

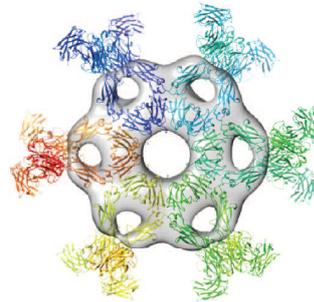
Diebold *et al.* (2014) establish a related result. In order to activate the complement system antibody have to form into hexameric ring-shaped multi-antibody multi-antigen complexes on the cell surface in which the Fc-region of the antibody bind to each other (see Section 1.2). When put together with the discussion of the stability of multi-antigen multi-antibody complexes in this thesis being enhanced if the antibody target multiple antigen-epitopes, this offers a potential explanation as to why donor antibody against the transplant does not guarantee rejection (Higgins *et al.*, 2011). In particular, depending on donor antigen positioning, and whether patient antibody targets multiple or single epitopes of donor antigen, different patients may develop very different numbers of multi-antibody multi-antigen structures on their cell surfaces, and may as a result experience very different complement activation.

### 10.1.2 SPR modelling

In Subsection 1.4.1 the high level of variation between affinity estimates for monoclonal antibody-antigen interactions was discussed, and it was noted that variation between estimates of the affinity of a single interaction in benchmark studies like Rich *et al.* (2009) and Yang *et al.* (2016) is vastly in excess of the variation for non-antibody antigen affinity estimates, such as ?. It was noted that in Rich *et al.* (2009) and Yang *et al.* (2016) the majority of modelling in these studies is done with the Langmuir 1 configuration model



(a) IgG Hexamer imaged with multi-angle light scattering with coloured Fab pairs.



(b) Surface rendering of a symmetrized Fc ring.

**Figure 10.1:** Visualisations of IgG hexamers from Diebolder *et al.* (2014).

(Eq. 2.5), which lacks many of the key features of antibody-antigen binding.

The key result of Chapter 6 is that as models of monoclonal antibody antigen interaction for SPR became more realistic through the inclusion of bivalence, the ERC assumption and heterogenous ligand distribution, RSS of fits decreased (see Fig. 6.8, Fig. 6.12, Fig. 6.20(a), Fig. 6.23(a) ), as did their AIC (see Fig. 6.32) and their affinity estimates became more consistent (see Fig. 6.6, Fig. 6.10, Fig. 6.14, Fig. 6.20(b), Fig.6.23(b)).

This result raises the question of whether the sensorgrams produced in studies like Rich *et al.* (2009) and Yang *et al.* (2016) could be reprocessed with the BERCs model, to produce more consistent results. A barrier to this is that as both were major international studies, and the sensorgrams were fitted with the default software for the SPR machines in the same laboratories that they were taken in. As a result, it is unlikely that the sensorgram data could be located and refit. The authors of both studies were contacted for collaboration but did not respond.

A paper that paralleled some of the development of the BERCs model was Hadzhiieva *et al.* (2017b). The authors noted that the density and distribution pattern of epitopes at the

surface of pathogens have a profound impact on immune responses and investigated the effect of varying antigen density on an SPR chip on the effective affinity of the interaction. Their analysis was limited by the lack of a dynamic model that could deal with the complexities of antibody binding to the isolated and non-isolated antigen. As a result, they used in the vast majority of cases equilibrium analysis (see Section 5.2), which effectively only considers one time-point in each sensorgram curve. In response to this, a collaboration was offered in which the BERCs model would be applied to their data.

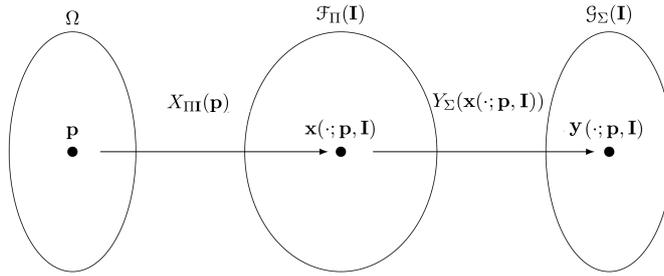
### 10.1.3 Identifiability and Indistinguishability

This thesis made a departure from the literature in how a model was conceptualised. This allowed for the concept of simple identifiability, and the visualisations of diagrams like those in Fig. 10.2. Notably because functions are either injective or not, this means that these four diagrams represent all possible models.

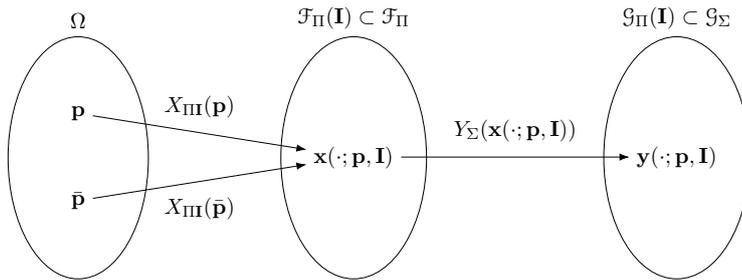
This allows for a different way of thinking about structural identifiability. Whether or not a model is structurally identifiable is a consequence of whether these mappings are injective or not (see Subsection 3.2.4).

The analysis of the mapping  $X_{\text{III}}(\cdot)$  is named simple identifiability, and is almost trivial, because the relationships between parameters that lead to the same state-space solutions may be obtained via inspection of the model.

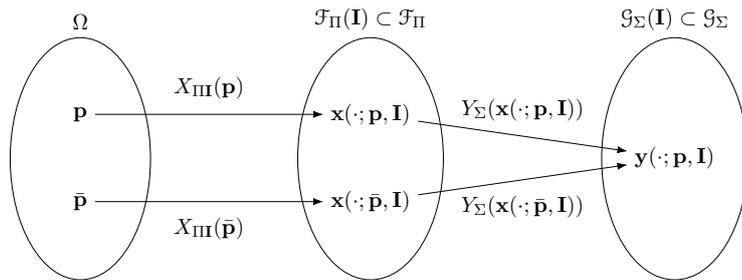
The ORC method was adapted for analysing the fibres of  $Y_{\Sigma}(\cdot)$ . In section 3.1.2 the concept of using the function  $\lambda(\mathbf{x}(t, \mathbf{p}))$  to examine the fibres of  $Y_{\Sigma}(\cdot)$ , by eliminating as many of the parameters from it as possible and using it as a pointwise mapping between indistinguishable state-space solutions. Significantly in this thesis all of the  $\lambda(\mathbf{x}(t, \mathbf{p}))$  mappings found have shared notable characteristics, each  $\lambda(\mathbf{x}(t, \mathbf{p}))$  mapping has been the multiplication of the state vector by a matrix and each matrix has had only one non-



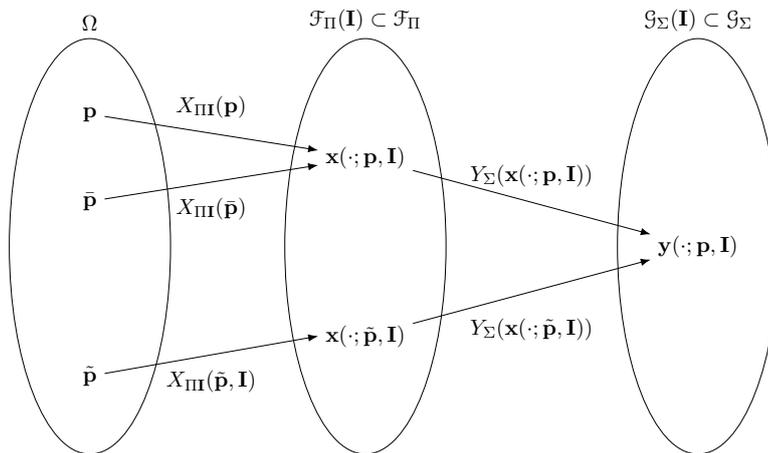
(a) A structurally identifiable model with injective  $X_{\text{III}}$  and  $Y_{\Sigma}$ .



(b) A model that is structurally unidentifiable because  $X_{\text{III}}$  is non-injective.



(c) A model that is structurally unidentifiable because  $Y_{\Sigma}$  is non-injective.



(d) A model that is structurally unidentifiable because neither  $X_{\text{III}}$  nor  $Y_{\Sigma}$  is injective.

**Figure 10.2:** Diagrams of  $X_{\text{III}}(\cdot)$  and  $Y_{\Sigma}(\cdot)$  as well as their domains and images for all models as defined in Definition 2.6.

zero number in every row, and one non-zero number in every column, and these non zero numbers have generally been one, viz. Eq. (3.29), Eq. (5.26), Eq. (5.27), Eq. (5.28), Eq. (5.29), Eq. (5.30) and these matrices have represented reflections and anisotropic scalings.

This motivates the following definition and conjecture.

**Definition 10.1.** *A scaled permutation matrix is a square matrix with a single non-zero element in each row and each column.*

**Conjecture 10.1.** For a model as defined in Definition 2.6, with an  $m$  dimensional state-space either

$$\lambda(\mathbf{x}(t, \mathbf{p})) = L(\mathbf{p})\mathbf{x}(t, \mathbf{p})$$

where  $L(\mathbf{p})$  is an  $m \times m$  scaled permutation matrix, or

$$\lambda(\mathbf{x}(t, \mathbf{p})) = L_1(\mathbf{p})\mathbf{x}(t, \mathbf{p}), \text{ or } L_2(\mathbf{p})\mathbf{x}(t, \mathbf{p}), \text{ or } \dots, \text{ or } L_l(\mathbf{p})\mathbf{x}(t, \mathbf{p})$$

where  $L_1(\mathbf{p}), L_2(\mathbf{p}), \dots, L_l(\mathbf{p})$  are  $m \times m$  scaled permutation matrices and  $l \in \mathbb{N}_{>0}$ .

The significance of this conjecture is that if proved it could allow for dramatically faster identifiability analyses, and analyses of models that were previously unidentifiable.

In addition to this approaches to identifiability that used the fact that SPR and BLI inputs are piecewise continuous, but not continuous were developed, which used the fact that the outputs were piecewise smooth and limits around the time point where the input is discontinuous to make symbolic structural identifiability analyses faster (see Fig. 3.9) and allow for the analyses of models that could not be otherwise analysed such as the BERCs model (see Subsection 5.3.6).

Whilst most inputs for models are not piecewise constant but not constant, this method

could easily be adapted for models with inputs that are piecewise smooth but not smooth. This is significant, because inputs are commonly piecewise smooth.

#### 10.1.4 Therapeutic applications

This thesis interacts with a large number of fields, but because of its focus on antibody-antigen interactions also has a strong links to current trends in medicine and pharmacy.

The monoclonal antibody drug market is projected to be worth \$125 billion by 2020 (Ecker *et al.*, 2015). Two of the key results of this thesis are that polyclonal antibodies can form much more permanent bonds with their antigen targets due to crosslinking and the development of larger multi-antibody multi- antigen complexes, and that mixtures of monoclonal antibodies against multiple epitopes of an antigen can create similar more permanent bonds. It is here assumed that these stronger cross linked bonds exist and serve immunological purposes within humans, and are not an artefact of SPR.

These two discoveries motivate the development of *oligoclonal* antibody therapies, that is antibody therapies in which a small number of monoclonal antibodies are mixed. This would have two key advantages over existing therapies. Firstly longer lasting bonds means that the antibodies in use could spend proportionally more time bound to their target, and less time in the blood stream of the patient. As a result smaller dosages could be used.

Secondarily and even more significantly these allow for better targeting of the antibody. Antibody bind to an epitope, and all antigen expressing that epitope (although this may be at different rates as established in Section 6.6 ). An oligoclonal therapy could be constructed to contain multiple antibodies that all bind to different epitopes of the same antigen, but so that there would not be other antigen that the antibodies could bind to.

This would mean that they could bind to the target antigen, and form the same much more permanent cross-linked structures with the antigen, but that they would only individually be able to bind to other patient antigen. As a result their bonds elsewhere would be much less permanent and have much smaller effects.

This would be useful for nearly all monoclonal antibody therapies, because unwanted binding is a problem for nearly all therapies and in many cases causes harmful side effects. Feyen *et al.* (2008) investigated adalimumab and infliximab, monoclonal antibodies that have been successfully used to treat rheumatic diseases and found that they also bind to proteins crucially involved in ribosomal protein synthesis which has been suggested as an explanation of the high rates of infection amongst patients undergoing the treatment (Kroesen *et al.*, 2003; Salliot *et al.*, 2007).

Simply put, to improve the efficacy of these treatments for rheumatic disease these monoclonal antibodies could be paired with other monoclonal antibodies that target other epitopes of the same antigen, and would then potentially benefit from multi-antibody multi-antigen binding, thus spend more time in contact with the target, and because they would not share these ribosome related targets, they would interact with them proportionately less. Additionally because multi-antibody multi-antigen bonds are longer lasting, lower dosages could be used further reducing side effects.

A monoclonal drug that may already be benefiting from a similar effect is rituximab. Rituximab binds specifically to two epitopes of its target antigen, CD20 (Binder *et al.*, 2006), and rituximab effects its target cells in ways that were not previously explained. Rituximab binding causes the distribution of CD20 in B cells to change, after exposure to rituximab CD20 forms a cap on one side of the cell as well as changes to the internal distribution of microtubules (Salliot *et al.*, 2007), this is shown in Fig. 10.3.

A potential explanation for this is the development of multi-antigen, multi-antibody structures on the cell. That is, if the CD20 antigen is free to move on the surface of

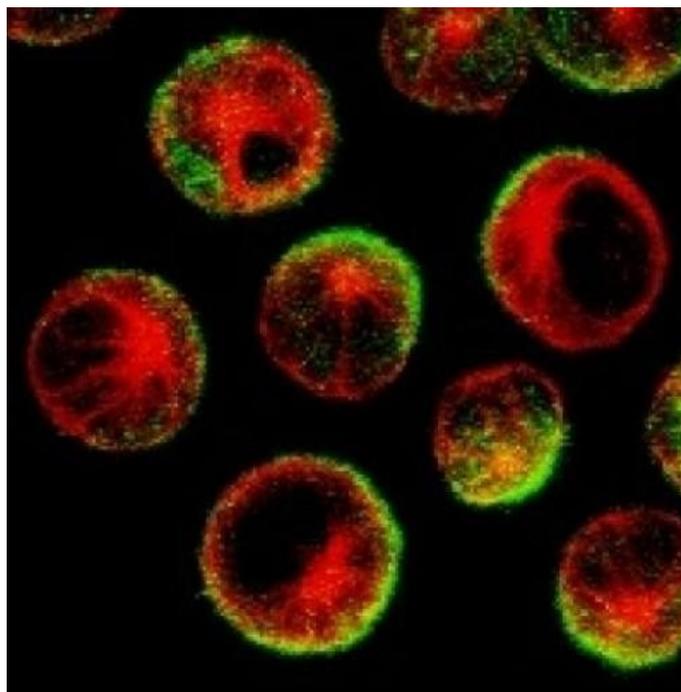
the cell, if one is bound to a rituximab antibody it may now bind to the second CD20; the new antigen- antibody-antigen has two free epitopes where rituximab can attach, meaning that if it becomes sufficiently close to other CD20 with an antibody bound to it, they may form an antigen-antibody-antigen-antibody-antigen structure, and so on, forming a chain of bound antibody and antigen on the surface of the cell.

Additionally these monoclonal antibody may also go through interactions like those discovered in Diebolder *et al.* (2014), i.e. the Fc-region of the bound antibodies bind to each other – allowing the chain to form relationships with itself, causing it to clump into the cap shown in Fig. 10.3. This process enhances the effectiveness of rituximab as measured by the rate at which T-cells kill the B-Cells. When the natural killer T-cells attach to the cell, 80% of the time they kill it if they latch on to the side where the protein has collected, compared to 40% of the time if they attached elsewhere or to a cell without a CD20 cap.

All of this points to the potential effectiveness of harnessing similar multi-antigen multi-antibody structures by using oligoclonal antibodies.

This explanation for the development of CD20 caps on the cells can be tested using enzymes that break apart antibody. The enzyme pepsin removes the Fc domain from an antibody leaving it as two linked antigen binding domains called a  $F(ab)_2$ . The enzyme papain breaks apart the antigen binding domains both from the heavy chain tail and each other leaving 2 Fab complexes (as shown in Fig. 10.4)

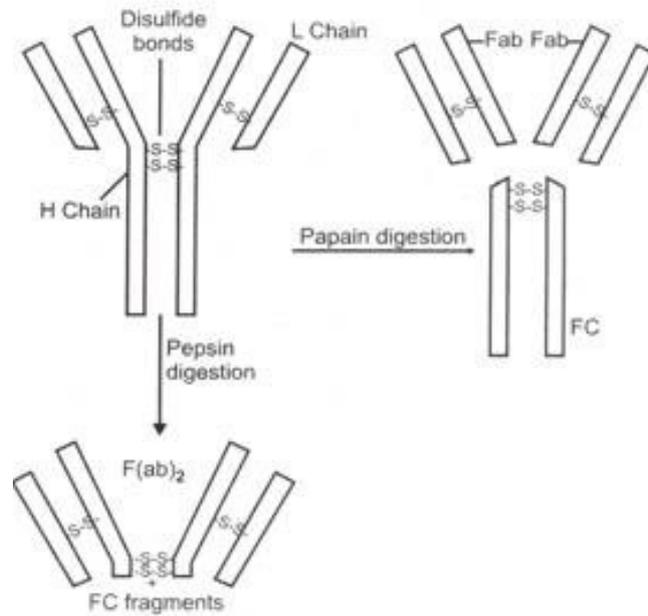
Three groups of B-cells could be exposed to intact rituximab antibodies, rituximab  $F(ab)_2$  complexes, and rituximab Fab complexes, and a fourth group of B-cells could be used as a control. These cells could be imaged and exposed to T-cells. If the formation of CD20 caps on the cells is due to multi-antibody multi-antigen complexes the cells exposed to the Fabs alone will not have these complexes. If binding of the Fc region of the antibody to other antibody is necessary for cap formation then the B-cells exposed to the rituximab



**Figure 10.3:** This image shows cancerous B cells that have been treated with rituximab. The protein CD20 (shown in green) has been drawn to the side of the cells from Kurzweil (2013).

$F(ab)_2$  complexes will not exhibit caps. Both of these groups can be compared to the cells exposed to intact rituximab and the cells that were not exposed to any form of the antibody, and a statistical analysis can be done. A second analysis could be based on the four groups of cells reaction to T-cells. This analysis would offer a simple way of confirming the role of multi-antibody multi-antigen binding in the formation of these caps, and potentially in cell death. If this experiment gave the result that cap formation is caused by of multi-antibody multi-antigen binding, this would make another potential case for oligoclonal antibody drugs, i.e. that they may result in T-cells more effectively causing cell death.

A second result that this experiment could produce would be whether T-cells interacting with cells with multi-antibody multi-antigen complexes where the antibody have their Fc domains removed are as effective in causing cell death.



**Figure 10.4:** Schematic of proteolytic fragmentation of immunoglobulin by brief treatment with enzymes papain and pepsin from ThermoFisher (2015).

Another therapy that could be developed, that uses multi-antibody multi-antigen binding would be *autologous modified polyclonal antibody therapy* (AMPAT). The goal of this therapy would be to use the properties of multi-antigen multi-antibody binding to block unwanted immune responses, for example as part of an autoimmune condition, or in antibody mediated transplant rejection.

Pre and post transplant antibody removal is already a well known and well established technique for dealing with immune incompatible transplants (Palmer *et al.*, 1989), typically either all patient antibody is removed, or only the antibody with affinity for the antigen on the transplant organ.

The first step of AMPAT is to remove patient antibody against the antigen, for which an immune reaction is to be discouraged. The antibody is then exposed to pepsin to create a polyclonal mixture of F(ab)<sub>2</sub> complexes. These are then injected into the patient, where they will bind to the target antigen.

Significantly as the patient has had all of his antibody against the antigen removed, these antibody fragments are free to bind to it and form the same kind of multi-antigen multi-antibody complexes as discussed earlier, except that these modified antibodies do not have Fc domains, so they will not interact with compliment (Diebolder *et al.*, 2014) resulting in one of the major parts of the patients immune system not interacting with the cells with the antigen.

Significantly if the rituximab B-cell experiments discussed earlier yield the result that the absence of Fc domains on the antibody in these complex leads to decreased rates of T-cell mediated cell death, then this treatment also reduces the T-cells role in the immune reaction as well as blocking compliment activation.

Whilst this happens the patients lymphocytes produce antibodies, but rather than being able to freely bind with the antigen, these have to compete with these multi-antibody multi-antigen structures, and in some cases be incorporated into them. Notably as initial amounts of whole antibody will be small compared to the amounts of antibody without their Cf domains, these intact antibody will generally not be able to form the hexamers that Diebolder *et al.* (2014) establishes were necessary for compliment activation- allowing this blockage to compliment and T-cell involvement in the immune reaction to persist.

The process of AMPAT could be repeated so that if the patient developed antibodies against other epitopes of the antigen these epitopes would also get covered in the antibody fragment without the Cf domain. The advantage of this proposed treatment is that it would be most useful in cases where the reaction was strongest, i.e for patients with large amounts of antibody this provides large amounts of the antibody fragment which can be used to block the antibody, and with antibodies against multiple epitopes on the antigen this provides a blocker able to form long lasting multi-antibody multi-antigen structures on cells with the target antigen.

Further in the future personalised oligoclonal antibody drugs, could be used in conjunction

with AMPAT, specifically antibody against the same antigen, without Cf domains possibly binding to different epitopes of the same antigen, and in high volume could also help block these responses.

Another big advantage of AMPAT is that it is not an immunosuppressant. Whilst it blocks a specific antibody interaction it leaves the whole immune system in tact, as a result it is likely not to create the same risks as general immunosuppressant drugs do, specifically cytopenia, infections with *Pneumocystis jirovecii*, nocardia, aspergillus, cryptococcus and reactivation of varicella zoster, herpes simplex, cytomegalovirus, hepatitis B and C as well as tuberculosis, cardiovascular disease and malignant cancer (Saag *et al.*, 2008). The testing and development of AMPAT should begin with an investigation of whether the removal of Cf domains from antibody decrease the effectiveness of T-cells attacking cells with multi-antibody multi-antigen complexes on them, such as those formed with rituximab as described above. However, as AMPAT is a therapy rather than a drug, i.e. a process rather than an object, it needs to be primarily studied through tests on living organisms.

Rats are an established model for transplantation (Wagner, 1980), and could be used to test renal transplants with AMPAT as a treatment. However, as this would rely on a surgical intervention for each test subject, this could be exceptionally costly.

A cheaper option would be to use one of the large number of established autoimmune disease models (Cohen & Miller, 1994) where an autoimmune disease is induced, such as adjuvant induced arthritis in rats. The benefit of these kinds of models is that AMPAT could be started as soon as the autoimmune disease begins – to simulate AMPAT beginning soon after immune-incompatible transplant, as well as later in the progress of the condition to simulate the treatment of a patient with an autoimmune disease as a pre-existing condition.

# Bibliography

“a person who cites his source brings deliverance to the world” (Avot 6:5)

- ABBAS, A.K., LICHTMAN, A.H., AND PILLAI, S. 2015. *Cellular and molecular immunology*. Elsevier Saunders.
- ABCAM. 2013a. *KD measurement: methods and materials*. <https://www.abcam.com/primary-antibodies/kd-measurement-methods-and-materials> [Accessed September 2018].
- ABCAM. 2013b. *KD value: a quantitative measurement of antibody affinity*. <http://www.abcam.com/primary-antibodies/kd-value-a-quantitative-measurement-of-antibody-affinity> [Accessed September 2018].
- ABDICHE, Y.N., MALASHOCK, D. S., PINKERTON, A, AND PONS, J. 2008. Determining Kinetics and Affinities of Protein Interactions Using a Parallel Real-time Label-free Biosensor, the Octet. *Analytical Biochemistry*, **377**(2), 209–217.
- AIKAWA, A., YAMASHITA, M., HADANO, T., OHARA, T., ARAI, K., KAWAMURA, T., AND HASEGAWA, A. 2003. ABO incompatible kidney transplantation - immunological aspect. *Exp. Clin. Transplant.*, **1**, 112–118.
- AKAIKE, H. 1974. A new look at the statistical model identification. *Transactions on Automatic Control*, **19**(6), 716–723.
- ANGUELOVA, M., KARLSSON, J., AND JIRSTRAND, M. 2012. Minimal output sets for identifiability. *Mathematical Biosciences*, **239**(1), 139 – 153.
- ARPALI E, KAUFMAN D, SOLLINGER H. 2016. Kidney Transplantation: Do Early Urinary Tract Complications Influence Long-Term Graft Survival? [abstract]. *Am J Transplant*, **16** (suppl 3).
- AWANGE, J.L. AND B.PALÁNCZ. 2016. *Geospatial algebraic computations: theory and applications*. Springer International Publishing.
- AYELA, C., ROQUET, F., VALERA, L., GRANIER, C., NICU, L., AND PUGNIERE, M. 2007. Antibody antigenic peptide interactions monitored by SPR and QCM-D: A model for SPR detection of IA-

## BIBLIOGRAPHY

---

- 2 autoantibodies in human serum. *Biosensors and Bioelectronics*, **22**(12), 3113 – 3119. Chem and Biosensing Transistors: from materials to systems.
- BELLU, G, SACCOMANI, M.P., S, AUDOLY, AND D'ANGIÒ, L. 2007. DAISY: a new software tool to test global identifiability of biological and physiological systems. *Computer methods and programs in biomedicine*, **88**(1), 52–61.
- BENDER, C.M. AND ORSZAG, S.A. 1978. *Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory*. New York: Springer.
- BENTALL, A.J. 2015. *Antibodies in kidney transplantation*. Ph.D. thesis, University of Birmingham.
- BEREK, C. & MILSTEIN, C. 1987. Mutation drift and repertoire shift in the maturation of the immune-response. *Immunological Reviews*, **96**, 23–41.
- BIACORE. 1997. *Kinetic and affinity analysis using BIA - Level 1*.
- BIACORE. 1998. *Kinetic and Affinity analysis using BIA - Level 2*.
- BINDER, M., OTTO, F., MERTELSMANN, R., VEELKEN, H., AND TREPPEL, M. 2006. The epitope recognized by rituximab. *Blood*, **108**(6), 1975–1978.
- BIO-RAD. 2010. *ProteOn XPR36 Protein Interaction Array System Protein Interaction Analysis User Manual*.
- BIO-RAD. 2011. *ProteOn Surface Plasmon Resonance Detection System*.
- BIO-RAD. 2010. *Affinity and Avidity of Antibodies*. <https://www.bio-rad-antibodies.com/antigen-antibody-interactions.html> [Accessed September 2018].
- BRITANNICA. 2016. *Encyclopaedia Britannica Online*. Encyclopaedia Britannica inc.
- CAI, J. & TERASAKI, P. I. 2005. Humoral theory of transplantation: Mechanism, prevention, and treatment. *Human Immunology*, **66**(4), 334–342.
- CHAIKEN, I., ROSE, S., AND KARLSSON, R. 1992. Analysis of macromolecular interactions using immobilized ligands. *Anal Biochem*, **201**(2), 197–210.
- CHEUNG, S. Y. AMY, YATES, JAMES W. T., AND AARONS, LEON. 2013. The design and analysis of parallel experiments to produce structurally identifiable models. *Journal of Pharmacokinetics and Pharmacodynamics*, **40**, 93–100.

- COHEN, IRUN R. AND MILLER, ARIEL (eds). 1994. *Autoimmune Disease Models*. San Diego: Academic Press.
- COLLINS, A. J., FOLEY, R. N., CHAVERS, B., GILBERTSON, D., HERZOG, C., ISHANI, A., JOHANSEN, K., KASISKE, B. L., KUTNER, N., LIU, J., *et al.* 2014. US Renal Data System 2013 Annual Data Report. *American journal of kidney diseases*.
- COMMONS, WIKIMEDIA. 2010. <https://upload.wikimedia.org/wikipedia/commons/9/9a/Monoclonals.png> [Accessed September 2018].
- COOPER, M. A. AND WILLIAMS, D. H. 1999. Kinetic analysis of antibody-antigen interactions at a supported lipid monolayer. *Analytical Biochemistry*, **276**, 36–47.
- COX, D, LITTLE, JOHN, AND O’SHEA, DONAL. 2007. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer.
- CUMMINGS, BENJAMIN. 2007. *Human Physiology: An Integrated Approach*. Pearson.
- DAGA, S, H, MOYSE, D, BRIGGS, D, LOWE, N2, EVANS, J4, JONES, R5, BUCHLI, C6, MCMURTREY, A7, MULDER, W6, HILDEBRAND, *et al.* 2017. Direct quantitative measurement of the kinetics of HLA-specific antibody interactions with isolated HLA proteins. *Hum Immunol.*, Oct., S0198–8859(17)30522–0.
- DAGA, SUNIL. 2015. *Biophysical characterisation and profile of HLA-specific antibodies in transplantation*. Ph.D. thesis, University of Warwick.
- DAGA, SUNIL, MOYSE, HARRY, BRIGGS, DAVID, LOWE, DAVID, EVANS, NEIL, JONES, JAMES, BUCHLI, RICO, MCMURTREY, CURTIS, MULDER, AREND, HILDEBRAND, WILLIAM, *et al.* 2018. Response to the comments on “Direct quantitative measurement of the kinetics of HLA-specific antibody interactions with isolated HLA proteins”. *Human Immunology*, **79**(2), 130 – 131.
- DENIS-VIDAL, L, JOLY-BLANCHARD, G, AND NOIRET, C. 2001. Some effective approaches to check the identifiability of uncontrolled nonlinear systems. *Mathematics in Computers and Simulation*, **57**, 35–44.
- DIEBOLDER, CHRISTOPH A., BEURSKENS, FRANK J., DE JONG, ROB N., KONING, ROMAN I., STRUMANE, KRISTIN, LINDORFER, MARGARET A., VOORHORST, MARLEEN, UGURLAR, DENIZ, ROSATI, SARA, HECK, ALBERT J. R., *et al.* 2014. Complement Is Activated by IgG Hexamers Assembled at the Cell Surface. *Science*, **343**(6176), 1260–1263.
- DRAPER, N.R.; SMITH, H. 1998. *Applied Regression Analysis (3rd ed.)*. John Wiley.

## BIBLIOGRAPHY

---

- DUQUESNOY, R.J.; ASKAR, M. 2007. HLAMatchmaker: a molecularly based algorithm for histocompatibility determination. V. Eplet matching for HLA-DR, HLA-DQ, and HLA-DP. *Hum Immunol.*, **68**, 12–25.
- DUQUESNOY, R.J. 2006. A structurally based approach to determine HLA compatibility at the humoral immune level. *Hum Immunol.*, **67**, 847–862.
- DUQUESNOY, R.J. 2014. Human leukocyte antigen epitope antigenicity and immunogenicity. *Curr Opin Organ Transplant.*, **19**, 428–435.
- ECKER, DAWN M., JONES, SUSAN DANA, AND LEVINE, HOWARD L. 2015 (Jan-Feb). *The therapeutic monoclonal antibody market.*
- EDWARDS, D., GOLDSTEIN, B., AND COHEN., D.S. 1999. Transport effects on surface volume biological reactions. *J. Math. Biol.*, **39**, 533–561.
- EDWARDS., D. A. 2001.. The effect of a receptor layer on the measurement of rate constants. *Bull Math Biol*, **63**(2), :301–327.
- EL NAHAS, A MEGUID ET AL. 2005. Chronic kidney disease: the global challenge. *The Lancet*, **Volume 365**(9456), 331 – 340.
- EVANS, N. D., CHAPMAN, M. J., CHAPPELL, M. J., AND GODFREY, K. R. 2002. Identifiability of uncontrolled nonlinear rational systems. *Automatica*, **38**(1799–1805).
- EVANS, N.D., CHAPPELL, M.J., CHAPMAN, M.J., AND GODFREY, K.R. 2004. Structural indistinguishability between uncontrolled (autonomous) nonlinear analytic systems. *Automatica*, **40**(11), 1947 – 1953.
- EVANS, N.D., MOYSE, H.A.J., LOWE, D., BRIGGS, D., HIGGINS, R., MITCHELL, D., ZEHNDER, D., AND CHAPPELL, M.J. 2013. Structural identifiability of surface binding reactions involving heterogeneous analyte: Application to surface plasmon resonance experiments. *Automatica*, **49**(1), 48 – 57.
- EVANS, NEIL, LOWE, DAVID, BRIGGS, DAVID, HIGGINS, ROBERT, BENTALL, ANDREW, BALL, SIMON, MITCHELL, DANIEL, ZEHNDER, DANIEL, AND J. CHAPPELL, MICHAEL. 2010. Estimation of kinetic rate constants from surface plasmon resonance experiments. *IET Seminar Digest*, **2010**(Jan.), 1–6.
- EVANS, NEIL D., WHITE, LISA J., CHAPMAND, MICHAEL J., GODFREYA, KEITH R., AND CHAPPELL, MICHAEL J. 2005. The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical Biosciences*, **194**(april), 175–197.

- FANG, Y. 2007. Label-free Optical Biosensors in Drug Discovery. *Trends in Biopharmaceutical Industry*, **3**, 34–38.
- F.E., JOHANSEN, R., BRAATHEN, AND P., BRANDTZAEG. 2000. Role of J Chain in Secretory Immunoglobulin Formation. *Scandinavian Journal of Immunology.*, **52**(3), 240–248.
- FEYEN, OLIVER, LUEKING, ANGELIKA, KOWALD, AXEL, STEPHAN, CHRISTIAN, MEYER, HELMUT E., GÖBEL, ULRICH, AND NIEHUES, TIM. 2008. Off-target activity of TNF- $\alpha$  inhibitors characterized by protein biochips. *Analytical and Bioanalytical Chemistry*, **391**(5), 1713–1720.
- FORSMAN, K. (1991). 1991. *Constructive commutative algebra in nonlinear control theory*. Ph.D. thesis, Linköping Institute of Technology.
- FORTEBIO. 2007. *BLI Technology*.
- GHERARDI, E. 2008. *Foundations of immunology*. lecture notes, University of Pavia.
- GILKEY, PETER, PARK, JEONGHYEONG, AND VÁZQUEZ-LORENZO., RAMÓN. 2015. *Aspects of differential geometry II*. San Rafael, California (1537 Fourth Street, San Rafael, CA 94901 USA) : Morgan & Claypool.
- GLASER, R. W. 1993. Antigen-antibody binding and mass transport by convection and diffusion to a surface: a two-dimensional computer model of binding and dissociation kinetics. *Anal. Biochem.*, **213**, 152–161.
- HADZHIEVA, M, A D. PASHOV, KAVERI, S V., LACROIX-DESMAZES, S, MOUQUET, H, AND DIMITROV, J. 2017a. Impact of Antigen Density on the Binding Mechanism of IgG Antibodies. *In: Scientific Reports*.
- HADZHIEVA, MAYA, PASHOV, ANASTAS D., KAVERI, SRINIVAS, LACROIX-DESMAZES, SÃ©BASTIEN, MOUQUET, HUGO, AND DIMITROV, JORDAN D. 2017b. Impact of Antigen Density on the Binding Mechanism of IgG Antibodies. *Scientific Reports*, **7**(1), 3767.
- H.A.J.MOYSE, D.LOWE, D.BRIGGS, R.HIGGINS, A.BENTALL, S.BALL, D.MITCHELL, M.J.CHAPPELL, ZEHNDER, D., AND N.EVANS. 2012. Estimation of antibody binding affinities with surface plasmon resonance experiments in incompatible blood type renal transplants. *In: 8th IFAC symposium on biological and medical systems*.
- HALBEISEN, LORENZ J. 2012. *Combinatorial set theory : with a gentle introduction to forcing*. London ; New York : Springer-Verlag London Ltd.

## BIBLIOGRAPHY

---

- HEARTY, S., LEONARD, P., AND O'KENNEDY, R. 2012. Measuring Antibody-Antigen Binding Kinetics Using Surface Plasmon Resonance. *Antibody Engineering. Methods in Molecular Biology (Methods and Protocols)*, **907**.
- HERMANN, R AND KRENER, A. J. 1977. Nonlinear controllability and observability. *IEEE Transactions on Automatic Control*, **AC-22**, 728–740.
- HIGGINS, R., HATHAWAY, M., LOWE, D., LAM, F., KASHI, H., TAN, L. C., IMRAY, C., FLETCHER, S., ZEHNDER, D., CHEN, K., *et al.* 2007. Blood levels of donor-specific human leukocyte antigen antibodies after renal transplantation: resolution of rejection in the presence of circulating donor-specific antibody. *Transplantation*, **84**(7), 876–884.
- HIGGINS, R., LOWE, D., HATHAWAY, M., WILLIAMS, C., LAM, F. T., KASHI, H., TAN, L. C., IMRAY, C., FLETCHER, S., CHEN, K., *et al.* 2011. Human leukocyte antigen antibody-incompatible renal transplantation: excellent medium-term outcomes with negative cytotoxic crossmatch.. *Transplantation*, **92**(8), 900–906.
- HOWELL, D. 2002. *Statistical Methods for Psychology*. Duxbury.
- HUANG, B., DAVERN, S., AND KENNEL, S.J. 2006. Mono and bivalent binding of a scFv and covalent diabody to murine laminin-1 using radioiodinated proteins and SPR measurements: Effects on tissue retention in vivo. *Journal of Immunological Methods*, **313**(1), 149 – 160.
- IN DRUG DISCOVERY, CURRENT BIOSENSOR TECHNOLOGIES. 2006. Cooper, M. A. *Drug Discovery World*, PP. 68–82.
- JACQUEZ., J.A. 1996. Compartmental analysis in biology and medicine. *Biomedware*.
- JAYARAMAN, B.N. MURTHY & S. SINHA & A. SUROLIA & S. S. INDI & N. 2008. SPR and ITC determination of the kinetics and the thermodynamics of bivalent versus monovalent sugar ligand lectin interactions. *Glycoconj J*, **25**(313-321).
- JONES, J, MOYSE, H.A.J., DAGA, S., AND LOWE, D. 2016 (Dec.). *A2 antigen*. E-mail conversation.
- KARLSSON, R., MICHAELSSON, A., AND MATTSSON, L. 1991. Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system. *Journal of Immunological Methods*, **145**(229-240;).

- KATSAMBA, P.S., I, NAVRATILOVA, M, CALDERON-CACIA, L, FAN, K, THORNTON, M, ZHU, TV, BOS, C, FORTE, D, FRIEND, I, LAIRD-OFFRINGA, *et al.* 2006. Kinetic analysis of a high-affinity antibody antigen interaction performed by multiple Biacore users. *Analytical Biochemistry*, PP. 208–221.
- KITOV, P.I. AND BUNDLE, D.R. 2003. On the Nature of the Multivalency Effect: A Thermodynamic Model. *Journal American Chemical Society*, **125**(52), 16271–16284.
- KLEIN, JOSHUA S. AND BJORKMAN, PAMELA J. 2010. Few and Far Between: How HIV May Be Evading Antibody Avidity. *PLOS Pathogens*, **6**(5), 1–6.
- KROESEN, S., WIDMER, A. F., TYNDALL, A., AND HASLER, P. 2003. Serious bacterial infections in patients with rheumatoid arthritis under anti-TNF- $\alpha$  therapy. *Rheumatology*, **42**(5), 617–621.
- KURZWEIL, RAY. 2013 (Apr.). *Scientists discover why a specific cancer drug is so effective.*
- KUZIEMKO, GEOFFREY M., STROH, MARK, AND STEVENS, RAYMOND C. 1996. Cholera Toxin Binding Affinity and Specificity for Gangliosides Determined by Surface Plasmon Resonance. *Biochemistry*, **35**(20), 6375–6384. PMID: 8639583.
- LANGMUIR., IRVING. 1916. The constitution and fundamental properties of solids and liquids. part i. solids. *J. Am. Chem. Soc.*, **38**(11), 2221–95,.
- LARVOR, M. P., DJAVADIOHANIAN, L., NALL, B., AND GOLDBERG, M. E. 1994. MEASUREMENT OF THE DISSOCIATION RATE-CONSTANT OF ANTIGEN/ANTIBODY COMPLEXES IN SOLUTION BY ENZYME-LINKED-IMMUNOSORBENT-ASSAY. *Journal of Immunological Methods*, **170**(2), 167–175.
- LEE, P. C., ZHU, L., TERASAKI, P. I., AND EVERLY, M. J. 2009. HLA-Specific Antibodies Developed in the First Year Posttransplant are Predictive of Chronic Rejection and Renal Graft Loss. *Transplantation*, **88**(4), 568–574.
- LEFSCHETZ, S. 2005. *Algebraic Geometry*, , Ch. I, 11.4. Dover Publications.
- LIU, J.K.H. 2014. The history of monoclonal antibody development: Progress, remaining challenges and future innovations. *Annals of Medicine and Surgery*, **3**(4), 113–116.
- LJUNG, L. AND GLAD., T. 1994. On global identifiability for arbitrary model parameterizations. *Automatica*, **30**, 265–276.
- LOWE, D. 2013. *Characterisation of HLA-specific antibodies.* Ph.D. thesis, Warwick Medical School.

## BIBLIOGRAPHY

---

- LUO, RUBEN. 2013. *Guide to SPR Data Processing on the ProteOn XPR36 Protein Interaction Array System*. Bio-Rad Laboratories, Inc., 2000 Alfred Nobel Drive, Hercules, CA 94547.
- MAO, Q., TERASAKI, P. I., CAI, J., BRILEY, K., CATROU, P., HAISCH, C., AND REBELLATO, L. 2007. Extremely high association between appearance of HLA antibodies and failure of kidney grafts in a five-year longitudinal study. *American Journal of Transplantation*, **7**(4), 864–871.
- MAPLESOFT, A DIVISION OF WATERLOO MAPLE INC. 2005-2015. *Maple User Manual*.
- MARGARIA, G. 2000. *Applications of differential algebra to the structural identifiability of non linear models*. Ph.D. thesis, Politecnico di Torino.
- MARGARIA, GABRIELLA, RICCOMAGNO, EVA, CHAPPELL, MICHAEL J., AND WYNN, HENRY P. 2001. Differential algebra methods for the study of the structural identifiability of rational function state-space models in the biosciences. *Mathematical Biosciences*, **174**(1), 1 – 26.
- MARQUART, J.A. 2013. *Surface Plasmon Resonance and Biomolecular interaction analysis*. Onbekend.
- MAZUMDAR, S. D.AND HARTMANN, M., KÄMPFER, P., AND M., KEUSGEN. 2007. Rapid method for detection of Salmonella in milk by surface plasmon resonance (SPR). *Biosensors and Bioelectronics*, **22**(9), 2040 – 2046. Selected Papers from the Ninth World Congress On Biosensors. Toronto, Canada 10 - 12 May 2006.
- MCMURTREY, C, LOWE, D, BUCHLI, R, DAGA, S, ROYER, D, AND A, HUMPHREY. 2014;. Profiling antibodies to class II HLA in transplant patient sera. *Human Immunology.*, **75**(3):, 261–70.
- MEIER-KRIESCHE, H.U., SCHOLD, J.D., AND KAPLAN, B. 2004. Long-term renal allograft survival: have we made significant progress or is it time to rethink our analytic and therapeutic strategies? *Am. J. Transplant.*
- MESHKAT, N., EISENBERG, M., AND DISTEFANO 3RD, J. J. 2009. An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner Bases,. *Mathematical Biosciences*, PP. 61–72.
- MOGHADAMNIA, A.A., ROSTAMI-HODJEGAN, A, ABDUL-MANAP, R, WRIGHT, CE, MORICE, AH, AND TUCKER, GT. 2003. Physiologically based modelling of inhibition of metabolism and assessment of the relative potency of drug and metabolite: dextromethorphan vs. dextrorphan using quinidine inhibition. *Br J Clin Pharmacol*, **56**, 57–67.

- MULDER, A, C, EIJSINK, MJ, KARDOL, VAN DIJK MEI, FRANKE, VAN DER BURG SH, AND M, KESTER. 2003;. Identification, isolation, and culture of HLA-A2-specific B lymphocytes using MHC class I tetramers. *Journal of Immunology*, **171**(12), 6599–6603.
- MURPHY, K. 2001. *Janeway's Immunobiology*. Garland science.
- MURRAY, J.D. 2002. *Mathematical Biology*. Springer-Verlag New York.
- MYSZKA, D. G. 1999. Survey of the 1998 optical biosensor literature. *J.Mol.Recognit*, **12**, 390–408.
- MYSZKA, D. G., JONSEN, M.D., AND GRAVES, B.J. 1998a. Equilibrium analysis of high affinity interactions using BIACORE. *Analytical Biochemistry*, **265**, 326–330.
- MYSZKA, D G, HE, X, DEMBO, M, MORTON, T A, AND GOLDSTEIN, B. 1998b. Extending the range of rate constants available from BIACORE: interpreting mass transport-influenced binding data. *Biophysical Journal*, **75**(2), 583–594.
- NABOK, A.V., TSARGORODSKAYA, A., HASSAN, A.K., AND STARODUB, N.F. 2005. Total internal reflection ellipsometry and SPR detection of low molecular weight environmental toxins. *Applied Surface Science*, **246**(4), 381 – 386. 9th European Conferences on Organised Films.
- NAKAJIMA, HIDEKI, KIYOKAWA, NOBUTAKA, KATAGIRI, YOHKO U., TAGUCHI, TOMOKO, SUZUKI, TOYO, SEKINO, TAKAOMI, MIMORI, KENICHI, EBATA, TOMOHIKO, SAITO, MASAHIRO, NAKAO, HIROSHI, *et al.* 2001. Kinetic Analysis of Binding between Shiga Toxin and Receptor Glycolipid Gb3Cer by Surface Plasmon Resonance. *Journal of Biological Chemistry*, **276**(46), 42915–42922.
- NANKIVELL, B. J. AND CHAPMAN, J. R. 2006. Chronic allograft nephropathy: current concepts and future directions. *Transplantation*,, **81** (5):, 643–654.
- NANKIVELL, B. J., BORROWS, R. J., FUNG, C. L. S., O'CONNELL, P. J., AND ALLEN, R. D. M. & CHAPMAN, J. R. 2003. The natural history of chronic allograft nephropathy. *New England Journal of Medicine*, **349** (24), 2326– 2333.
- NEDERPELT, R. 2004. *Logical Reasoning: A First Course*. 3 edn. Texts in Computing. King's College Publications.
- NELSON E, SCHALDEMOSE I. 1959. Urinary excretion kinetics for evaluation of drug absorption. *Journal of American Pharmacists Association*, PP. 489–495.
- NERI, D., MONTIGIANI, S., AND KIRKHAM, P. M. 1996. Biophysical methods for the determination of antibody-antigen affinities. *Trends in Biotechnology*,, **14**(12), 465–470.

## BIBLIOGRAPHY

---

- NHSBT. 2016. *NHS Blood and Transplant Annual report on kidney transplantation report for 2015/2016*.
- NIELSEN, ULRIC B., ADAMS, GREGORY P., WEINER, LOUIS M., AND MARKS, JAMES D. 2000. Targeting of Bivalent Anti-ErbB2 Diabody Antibody Fragments to Tumor Cells Is Independent of the Intrinsic Antibody Affinity. *Cancer Research*, **60**(22), 6434–6440.
- NITSCH, DOROTHEA, GRAMS, MORGAN, SANG, YINGYING, BLACK, CORRI, CIRILLO, MASSIMO, DJURDJEV, OGNJENKA, ISEKI, KUNITOSHI, JASSAL, SIMERJOT K, KIMM, HEEJIN, KRONENBERG, FLORIAN, *et al.* 2013. Associations of estimated glomerular filtration rate and albuminuria with mortality and renal failure by sex: a meta-analysis. *BMJ*, **346**.
- ONG, G.L. AND MATTES, M.J. 1993. Re-evaluation of the concept of functional affinity as applied to bivalent antibody binding to cell surface antigens. *Mol Immunol.*, **30**(16), 1455–62.
- OWEN, J., PUNT, J., AND STRANFORD, S. 2013. *Kuby Immunology*. W. H. Freeman.
- PALMER, A., WELSH, K., GJORSTRUP, P., TAUBE, D., BEWICK, M., AND THICK, M. 1989. Removal of anti-hla antibodies by extracorporeal immunoabsorption to enable renal transplantation. *The Lancet*, **333**(8628), 10 – 12.
- PAPALIA, GIUSEPPE A., LEAVITT, STEPHANIE, BYNUM, MAGGIE A., KATSAMBA, PHINIKOULA S., WILTON, ROSEMARIE, QIU, HUAWEI, STEUKERS, MIEKE, WANG, SIMING, BINDU, LAKSHMAN, PHOGAT, SANJAY, *et al.* 2006. Comparative analysis of 10 small molecules binding to carbonic anhydrase II by different investigators using Biacore technology. *Analytical Biochemistry*, **359**(1), 94 – 105.
- PATEL, R. & TERASAKI, P. I. 1969. Significance of the positive cross match test in kidney transplantation. *New England Journal of Medicine*, **280**(14), 735–739.
- POHJANPALO, H. 1978a. SystemIdentifiability based on the power series of the solution. *Mathematical Biosciences*, **41**, 21–33.
- POHJANPALO, H. 1978b. system identifiability for arbitrary model parameterizations. *Automatica*, **41**, 265–276.
- POKHREL, P. 2015. *ELISA-principle Types and Applications*. MicrobiologyNotes.com.
- RICARD-BLUM, SYLVIE, FÃ©RAUD, OLIVIER, LORTAT-JACOB, HUGUES, RENCUROSÌ, ANNA, FUKAI, NAOMI, DKHISSI, FATIMA, VITTET, DANIEL, IMBERTY, ANNE, OLSEN, BJORN R., AND VAN DER REST, MICHEL. 2004. Characterization of Endostatin Binding to Heparin and Heparan Sulfate by Surface

- Plasmon Resonance and Molecular Modeling: ROLE OF DIVALENT CATIONS. *Journal of Biological Chemistry*, **279**(4), 2927–2936.
- RICH, R. L.; MYSZKA, D. G. 2007. Higher-throughput, label-free, real-time molecular interaction analysis. *Analytical Biochemistry*, **361**, 1–6.
- RICH, REBECCA L., PAPALIA, GIUSEPPE A., FLYNN, PETER J., FURNEISEN, JAMIE, QUINN, JOHN, KLEIN, JOSHUA S., KATSAMBA, PHINI S., WADDELL, M. BRENT, SCOTT, MICHAEL, THOMPSON, JOSHUA, *et al.* 2009. A global benchmark study using affinity-based biosensors. *Analytical Biochemistry*, **386**(2), 194 – 216.
- RICHARD B. M. SCHASFOORT, ANNA J. TUDOS. 2008. *Handbook of Surface Plasmon Resonance*. Royal Society of Chemistry.
- ROBINSON, J., HALLIWELL, J.A., HAYHURST, J.H., FLICEK, P., PARHAM, P., AND MARSH, S.G.E. 2015. The IPD and IMGT/HLA database: allele variant databases. *Nucleic Acids Research*, **43**, D423–431.
- ROITT, I.M., BURTON, D.R., DELVES, P.J., AND MARTIN, S.J. 2017. *Essential immunology*. Chichester, West Sussex ; Hoboken, NJ : Wiley-Blackwell, 201.
- RUDNICK, STEPHEN I. AND ADAMS, GREGORY P. 2009. Affinity and Avidity in Antibody-Based Tumor Targeting. *CANCER BIOTHERAPY AND RADIOPHARMACEUTICALS*, **24**(2).
- SAAG, KENNETH G., TENG, GIM GEE, PATKAR, NIVEDITA M., ANUNTIYO, JEREMY, FINNEY, CATHERINE, CURTIS, JEFFREY R., PAULUS, HAROLD E., MUDANO, AMY, PISU, MARIA, ELKINS-MELTON, MARY, *et al.* 2008. American College of Rheumatology 2008 recommendations for the use of nonbiologic and biologic disease-modifying antirheumatic drugs in rheumatoid arthritis. *Arthritis Care & Research*, **59**(6), 762–784.
- SABBAN, SARI. 2011. *Development of an in vitro model system for studying the interaction of Equus caballus IgE with its high- affinity FceRI receptor*. phdthesis, The University of Sheffield.
- SABER, R., SARKAR, S., GILL, P., NAZARI, B., AND FARIDANI, F. 2011. High resolution imaging of IgG and IgM molecules by scanning tunneling microscopy in air condition. *Scientia Iranica*, **18**(6), 1643 – 1646.
- SALLIOT, C., GOSSEC, L., RUYSSSEN-WITRAND, A., LUC, M., DUCLOS, M., GUIGNARD, S., AND DOUGADOS, M. 2007. Infections during tumour necrosis factor- $\alpha$  blocker therapy for rheumatic diseases in daily practice: a systematic retrospective study of 709 patients. *Rheumatology*, **46**(2), 327–334.

## BIBLIOGRAPHY

---

- SARID, D. AND CHALLENGER, W. 2010. *Modern Introduction to Surface Plasmons: Theory, Mathematical Modeling, and Applications*. . Cambridge University Press.
- SCHAUER, M. AND HEINRICH, R. 1983. Quasi-steady-state approximation in the mathematical modeling of biochemical reaction networks. *Mathematical Biosciences*, **65**(2), 155 – 170.
- SCIENCE AND COUNCIL, TECHNOLOGY FACILITIES. 2015 (Apr.).
- SENNHAUSER, F. H., MACDONALD, R. A., ROBERTON, D. M., AND HOSKING, C. S. 1989. Comparison of concentration and avidity of specific antibodies to E. coli in breast milk and serum. *Immunology*, **66**, 394–397.
- SONTAG, E.D. 1998. *Mathematical Control Theory*. Springer-Verlag New York.
- SUN, Y., XIAO, S., CHEN, J., WANG, M., Z., ZHENG, S., SONG, AND L., ZHANG. 2015. Heat Shock Protein 90 Mediates the Apoptosis and Autophagy in Nicotinic-Mycoepoxydiene-Treated HeLa Cells. *Acta Biochim Biophys Sin*, **47**(6), 451–8.
- TERASAKI, P. I. AND CAI, J. C. 2008. Human leukocyte antigen antibodies and chronic rejection: From association to causation. *Transplantation*, **86**(3), 377–383.
- THERMOFISHER, SCIENTIFIC. 2015. *Antibody Fragmentation*.
- TOBIAS, R. AND KUMARASWAMY, S. 2013. *Biomolecular Binding Kinetics Assays on the Octet Platform*.
- VAJDA, S. 1984. Structural Identifiability of Linear, Bilinear, Polynomial and Rational Systems. *IFAC Proceedings Volumes*, **17**(2), 717 – 722. 9th IFAC World Congress: A Bridge Between Control Science and Technology, Budapest, Hungary, 2-6 July 1984.
- VAJDA, S, GODFREY, K, AND RABITZ, H. 1989. Similarity transformation approach to identifiability analysis of nonlinear compartmental models. *Mathematical Biosciences*, **93**(217-248).
- VILLAVERDE, ALEJANDRO F., BARREIRO, ANTONIO, AND PAPACHRISTODOULOU, ANTONIS. 2016. Structural Identifiability of Dynamic Systems Biology Models. *PLOS Computational Biology*, **12**(10), 1–22.
- WAGNER, E. 1980. The rat as experimental model for organ transplantation: technique of rat kidney transplantation. *Contrib Nephrol.*, **19**, 167–7.
- WALTER, E; LECOURTIER, Y. 1982. Global approaches to identifiability testing for linear and nonlinear state space models. *Mathematics and Computers in Simulation*, **24**.

- WEISSTEIN, ERIC W. 2010. *Fibre*. MathWorld—A Wolfram Web Resource.
- WILLIAMS, R. C., MALONE, C., BLOOD, B., AND SILVESTRIS, F. 1999. Anti-DNA and anti-nucleosome antibody affinity - A mirror image of lupus nephritis? *Journal of Rheumatology*, **26**(2), 331–346.
- WOLFE, R. A., ASHBY, V. B., MILFORD, E. L., OJO, A. O., ETTENGER, R. E., AGODOA, L. Y. C., HELD, P. J., AND PORT, F. K. 1999. Comparison of mortality in all patients on dialysis, patients on dialysis awaiting transplantation, and recipients of a first cadaveric transplant. *New England Journal of Medicine*, **341**(23), 1725–1730.
- XIA X, MOOG CH. 2003. Identifiability of nonlinear systems with applications to HIV and AIDs models. *Transactions on Automatic Control*, **48**, 330–336.
- YANG, D., SINGH, A., WU, H., AND KROE, R.R. 2016. Comparison of biosensor platforms in the evaluation of high affinity antibody antigen binding kinetics. *Analytical Biochemistry*.
- YARMUSH, M.L., PATANKAR, D.B., AND YARMUSH., D.M. 1996. An analysis of transport resistance in the operation of biacore; implications for kinetic studies of biospecific interactions. *Mol. Immunol*, **33**, 1203–1214.
- ZAHND, C, SPINELLI, S, LUGINBÄHL, B, AMSTUTZ, P, CAMBILLAU, C, AND PLÄCKHUN, A. 2004. Directed in vitro evolution and crystallographic analysis of a peptide-binding single chain antibody fragment (scFv) with low picomolar affinity. *J Biol Chem.*, **279**(18), 18870–7.
- ZHUANG, G. Q., KATAKURA, Y., OMASA, T., KISHIMOTO, M., AND SUGA, K. I. 2001. Measurement of association rate constant of antibody-antigen interaction in solution based on enzyme-linked immunosorbent assay. *Journal of Bioscience and Bioengineering*, **92**(4), 330–336.
- Z.SUN AND M.SARABIA. 2011. *Stability Theory of Switched Dynamical Systems*. 1 edn. Communications and Control Engineering. Springer-Verlag.

# Appendix A

## Samples of Maple and Facsimile code

Much of this thesis is focused on problems that are solved with a mathematical computer program, Maple, rather than by hand. As a result code is presented.

### A.1 Maple code from Chapter 2

#### A.1.1 Deriving the state-space solution for the Langmuir 2 concentration model

For the Graphs of Example 2.2 the following code was used.

```
ode := [diff(x[1](t), t) = k_a1*C*(R-x[1](t)-x[2](t))-k_d1*x[1](t),  
diff(x[2](t), t) = k_a2*C*(R-x[1](t)-x[2](t))-k_d2*x[2](t)];  
S := dsolve(ode);  
solve(''[ '=' ](0, ''[rhs](subs(t = 0, S))), _C1, _C2);  
solve(''[rhs](subs(timeless, ode)), x[1], x[2])  
S2 := dsolve(subs(C = 0, ode), x[1], x[2])
```

## A.2 Chapter 3

Neil Evans wrote two procedures that were used in the majority of examples of Chapter 2.

This code relies on two packages that have to be opened.

```
with(LinearAlgebra): with(Groebner):
```

The first procedure, `lieDer`, determines the Lie derivative of `H` along `F`.

```
lieDer := proc (H, F)
  local N, V, vars;
  N := nops(F);
  vars := [seq(x[t], t = 1 .. N)];
  V:=map( (a, b) -> diff(b, a), vars, H):
  DotProduct(Vector(F),Vector(V),conjugate=false)
end:
```

The second procedure, `listLieDer`, determines the list of the first `N` Lie derivatives of `H` along `F`.

```
listLieDer := proc (H, F, N)
  local L, i, tmp;
  L := [y[0]-H];
  tmp := H;
  for i to N do
    tmp := lieDer(tmp, F);
    L := [op(L), y[i]-tmp]
  od;
end:
```

### A.3 Structural Identifiability

Maple code to implement an adapted form of the *Taylor series approach* method for SPR experiments was written. This procedure takes the input  $H$ ,  $F$ ,  $NO$ ,  $Na$ ,  $Nd$ ,  $Pnum$ ,  $Xnum$  where  $H$  is the output mapping of the model,  $F$  the state-space function of a model and  $In$  is a list containing the two values in order the input equation will take,  $NO$ ,  $Na$ ,  $Nd$  are the number of iterative Lie - derivatives that should be taken at each of three points,  $Pnum$  and  $Xnum$  are the numbers of parameters and state variables respectively, and the parameters are ordered so that concentration is the last one.

```
GTS := proc (H, F, In, NO, Na, Nd, Pnum, Xnum)
local H0, Ha, Hd, pqsubs, H0q, Hdq, Haq, Sols;
H0 := listLieDer(H, subs(In=In[1], F), Na);
H0 := subs([x[1] = 0, x[2] = 0, y = y0], Ha);
Ha := listLieDer(H, subs(In=In[1], F), Na);
Ha := subs(y = ya, Ha);
Hd := listLieDer(H, subs(In=In[2], F), Nd);
pqsubs := p = q;
H0q := subs(pqsubs, H0);
Hdq := subs(pqsubs, Hd);
Haq := subs(pqsubs, Ha);
Sols := solve([op(op(H0q)), op(op(H0)), op(Hdq), op(Hd), op(Haq), op(Ha),
  alpha <> 0, seq(p[i] <> 0, i = 1 .. Pnum), seq(x[i] <> 0, i = 1 .. Xnum),
  y0[1] <> 0])
end proc;
```

#### A.3.1 Example 2.2

First the system and output equations are written.

```
F := [k[a1]*C[T]*(R-x[1]-x[2])-k[d1]*x[1],
      k[a2]*C[T]*(R-x[1]-x[2])-k[d2]*x[2]];
In:= [C[T], 0];
H := alpha*(x[1]+x[2]);
```

The system equations were converted to a form that could be input into for the `GTS` proc, and the `GTS` proc was run, producing a system of equations representing the relationships between parameters and state variables at the start of the dissociation phase that are necessary for two parameterisations to have the same output at all points.

```
psubs:= [p[1]=k[a1], p[2]=R, p[3]=k[d1], p[4]=k[a2], p[5]=k[d2], p[6]=C[T]];
F := subs(psubs, F);
In:= subs(psubs, In);
GTS(H, F, In, 4, 4, 4, 6, 2)
```

$$\begin{bmatrix} xq_1 = x_2, xq_2 = x_1, q_1 = \frac{q_4 p_4}{p_1}, q_2 = p_2, q_3 = p_5, q_4 = q_4, q_5 = p_3, q_6 = \frac{p_1 p_6}{q_4} \\ xq_1 = x_1, xq_2 = x_2, q_1 = \frac{q_4 p_1}{p_4}, q_2 = p_2, q_3 = p_3, q_4 = q_4, q_5 = p_5, q_6 = \frac{p_4 p_6}{q_4} \end{bmatrix}$$

The output can be rewritten in terms of the original parameters by hand, giving the results written in example 2.2.

### A.3.2 Measuring the time taken to find, and dimension of a solution by the TSA

First the packages were loaded and programs were created to find solutions and partial solutions.

```
with(LinearAlgebra);
with(Groebner);
```

```

lieDer := proc (H, F) local N, V, vars;
N := nops(F);
vars := [seq(x[t], t = 1 .. N)];
V := map(proc (a, b) options operator, arrow;
diff(b, a) end proc, vars, H);
DotProduct(Vector(F), Vector(V), conjugate = false) end proc;

listLieDer := proc (H, F, N) local L, i, tmp;
L := [y[0]-H];
tmp := H;
for i to N do tmp := lieDer(tmp, F);
L := [op(L), y[i]-tmp] end do end proc;

TypeTools:-AddType(hname, proc (n) local r;
r := n;
if not n::name then return false end if;
while not r::symbol do r := op(0, r) end do;
evalb(substring(r, -1 .. -1) = ':-h') end proc);

GTS := proc (H, F, Na, Nd) local Ha, Hd, hatsubs, Hdhat, Hahat, Sols;
Ha := listLieDer(H, F, Na);
Ha := subs([x[1] = 0, x[2] = 0, y = ya], Ha);
Hd := listLieDer(H, subs(C[T] = 0, F), Nd);
Hd := subs(y = yd, Hd);
hatsubs := [k = kh, C = Ch, R = Rh, x = xh];
Hdhat := subs(hatsubs, Hd);
Hahat := subs(hatsubs, Ha);
Sols := solve([op('~['-'')(Ha, Hahat)), op('~['-'')(Hd, Hdhat)),
x[1] <> 0, Ch[T] <> 0, C[T] <> 0]) end proc;
GTS2 := proc (H, F, Na, Nd) local Sols, Sols2, Sols3;
Sols := [GTS(H, F, Na, Nd)];
numelems(Sols);

```

```

Sols2 := remove(proc (s) options operator, arrow;
true in map(proc (e) options operator, arrow;
'assuming'([is(rhs(e) <= 0)], [positive]) end proc, s) end proc, Sols);
numelems(Sols2);
Sols3 := select(proc (s) options operator, arrow;
andmap(hastype, remove(evalb, s), hname) end proc, Sols2);
numelems(Sols3); Sols3 end proc;

```

A second group of programs were developed to measure the time taken by the first group of programs and the dimension of their solutions.

```

GTS2timer2 := proc (H, F, Na, Nd) local st;
st := time();
try timelimit(2000, GTS2(H, F, Na, Nd))
catch "time expired": printf("Time Expired - I'm baling out on iteration i=%d
n", Nd) finally time()-st end try end proc;
Dims := proc (S::(set('='))) options operator, arrow;
nops(select(evalb, S)) end proc;
DimsPars := proc (S::(set('='))) options operator, arrow;
select(evalb, S) end proc;
TypeTools:-AddType('Variable', proc (v) options operator, arrow;
v::'indexed' and (op(0, v))::('identical')('x', 'y'), 'Variable' end proc);
VarsParams := proc (e) options operator, arrow;
selectremove(type, indets(e, 'name'), 'Variable') end proc;
CountVP := proc (e) options operator, arrow;
' '[nops]([VarsParams(e)]) end proc;
AllVP := proc (e) options operator, arrow;
' '[op]([VarsParams(e)]) end proc;
CountAllVP := proc (e) options operator, arrow;
nops(AllVP(e)) end proc;

```

```

Dims := proc (S::(set('='))) options operator, arrow;
nops(select(evalb, S)) end proc;
DimsPars := proc (S::(set('='))) options operator, arrow;
select(evalb, S) end proc;

```

The timer was run, and dimensions were counted, producing the results graphed in Fig. 3.5.

```

Nd := 4;
Na := 1;
v1 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv1 := seq(max('~'[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
Na := 2;
v2 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv2 := seq(max('~'[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
Na := 3;
v3 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv3 := seq(max('~'[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
Na := 4;
v4 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv4 := seq(max('~'[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
Na := 5;
v5 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv5 := seq(max('~'[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
Na := 6;
v6 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv6 := seq(max('~'[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
Na := 7;
v7 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv7 := seq(max('~'[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
Na := 8;

```

```
v8 := [seq(GTS2timer2(H, F, Na, i), i = 1 .. Nd)];
Dimv8 := seq(max(‘‘[Dims](GTS(H, Vector(F), Na, i))), i = 1 .. Nd);
```

### A.3.3 Output equation method

Neil Evans wrote three procedures were made that will calculate with the output MVDP given the system equations F, and the output equation H.

```
outptEqn1 := proc (F, H) local N, L, vars; N := nops(F);
L := listLieDer(H, F, N);
L := map(expand, L);
vars := [seq(x[t], t = 1 .. N), y[N]];
UnivariatePolynomial(y[N], L, vars) end proc;
```

```
outptEqn2 := proc (F, H) local N, G, L, vars;
N := nops(F);
L := listLieDer(H, F, N);
vars := [seq(x[t], t = 1 .. N), ‘$(y[N-i], i = 0 .. N)];
G := Basis(L, plex(op(vars)));
G[1] end proc;
```

```
outptEqn3 := proc (F, H) local N, G, L, vars;
N := nops(F);
L := listLieDer(H, F, N);
vars := [seq(x[t], t = 1 .. N), y[N]];
G := Basis(L, plex(op(vars)));
G[1] end proc
```

A fourth procedure was written, `outptEqn4`, which allows for faster computation by iteratively finding the Gröbner basis and differentiating it - hence minimising the order of polynomials being differentiated, allowing for faster computation. This requires a slightly

different method for calculating the Lie - derivative as Lie - derivatives will now be taken of polynomials including  $y$ . An algorithm to do this was written `lieDer2`.

```

lieDer2 := proc (H, F)
local N, V, vars; N := nops(F);
vars := [seq(x[t], t = 1 .. N), seq(y[t], t = 0 .. N+1)];
V := map(proc (a, b) options operator, arrow; diff(b, a)
end proc,

outptEqn4 := proc (F, H)
local N, G, L, i, G2, dG, dG2, vars;
N := nops(F);
L := listLieDer(H, F, 1);
G := [seq(numer(L[i])*denom(L[i]), i = 1 .. numelems(L))];
vars := [seq(x[t], t = 1 .. N)];
for i to N-1 do
G2 := Groebner:-Basis([op(G)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
G := [op(G2), op(dG2)]
end do;
G2 := Groebner:-Basis([op(G)], plex(op(vars), y[N])); G2 end proc;

```

### A.3.4 Example 3.1.3

The system is declared, and the MVDP for either phase (Equations 3.35 and 3.36) is derived in the first block of code. In the second block the coefficients of these MVDPs are extracted, equations are formed from them and solved giving the conditions required for two parameterisations to give equal outputs.

```
F := [k[a1]*C[T]*(R-x[1]-x[2])-k[d1]*x[1],
```

```

    k[a2]*C[T]*(R-x[1]-x[2])-k[d2]*x[2]];
In=[C[T],0];
H := alpha*(x[1]+x[2]);
MVDPa:=outptEqn2(subs(in=in[1],F), H);
MVDPd:=outptEqn2(subs(in=in[2],F), H);

ParsList:=[k=kh,R=Rh,C=Ch]:
CoeffList:=(op(coeffs(MVDPa,[x[1],x[2]]),op(coeffs(MVDPd,[x[1],x[2]])))));
Eqs:=CoeffList=~subs(ParsList,CoeffList);
solve(Eqs);

```

### A.3.5 Output limits at discontinuity

To derive a set of output BVDPs for a model two procedures are used. The first, `lieDer2` allows Lie-derivatives to be taken of a multivariate polynomial  $H$  both in terms of the components of the state variables  $\mathbf{x}$  and the derivatives of the output  $y$ . The second BVDP takes in five arguments  $F$  the rates of change of the state variables,  $H$  the output of the model,  $In$  the input taken by the model, and  $Na$  and  $Nd$  the number of derivatives to be considered in either phase. This procedure outputs a Gröbner basis for polynomial ideal composed of the first  $Na$  Lie - derivatives of the output for the association phase, and the the first  $Nd$  Lie - derivatives of the output for the dissociation phase. Due to the choice of monomial ordering, the Gröbner basis will include a set of output BVDPs. This set will be of non-zero size if the sum of  $Na$  and  $Nd$  are greater than the number of state variables.

```

lieDer2 := proc (H, F)
local N, V, vars;
N := nops(F);
vars := [seq(x[t], t = 1 .. N), seq(y[t], t = 0 .. N+1)];
V := map(proc (a, b) options operator, arrow; diff(b, a) end proc, vars, H);

```

```

DotProduct(Vector([op(F), seq(y[t+1], t = 0 .. N+1)]), Vector(V), conjugate = false)
end proc;

BVDP := proc (F, H, In, Na, Nd)
local N, Ga, Gd, La, Ld, i, G2, dG, dG2, vars;
N := nops(F); La := listLieDer(H, subs(In[1], F), 1);
Ld := listLieDer(H, subs(In[2], F), 1);
Ga := [seq(numer(La[i])*denom(La[i]), i = 1 .. numelems(La))];
Gd := [seq(numer(Ld[i])*denom(Ld[i]), i = 1 .. numelems(Ld))];
vars := [seq(x[t], t = 1 .. N)];
for i to Na-1 do G2 := Basis([op(Ga)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Ga := [op(G2), op(dG2)]
end do;
for i to Nd-1 do G2 := Basis([op(Gd)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Gd := [op(G2), op(dG2)]
end do;
G2 := Basis([op(subs(y = ya, Ga)), op(subs(y = yd, Gd))], plex(op(vars), seq(ya[Na-i], i = 0 .. Na), seq(yd[Nd-i],
G2
end proc

```

### A.3.6 Measuring the time taken to find, and dimension of a solution by the BVDP and TSA at limits

The model is defined.

```

F := [k[a1]*C[T]*(R-x[1]-x[2])-k[d1]*x[1],
k[a2]*C[T]*(R-x[1]-x[2])-k[d2]*x[2]];
H := 1000*(x[1]+x[2])

```

Basic programs for counting dimensions, recognising parameters from that hatted parameter vector, here represented by adding an *h* to the end of the parameter name, and variables from its evaluation, again represented by adding an *h* to the end of the parameter name.

```
TypeTools:-AddType(hname, proc (n) local r;
  r := n;
  if not n::name then return false end if;
  while not r::symbol do r := op(0, r) end do;
  evalb(substring(r, -1 .. -1) = ':-h') end proc)
Dims := proc (S::(set('='))) options operator, arrow;
  nops(select(evalb, S)) end proc;
DimsPars := proc (S::(set('='))) options operator, arrow;
  select(evalb, S) end proc;
TypeTools:-AddType('Variable', proc (v) options operator, arrow;
  v::'indexed' and (op(0, v))::('identical')('x', 'y'), 'Variable' end proc);
VarsParams := proc (e) options operator, arrow;
  selectremove(type, indets(e, 'name'), 'Variable') end proc;
CountVP := proc (e) options operator, arrow;
  '--[nops]([VarsParams(e)]) end proc;
AllVP := proc (e) options operator, arrow;
  '--[op]([VarsParams(e)]) end proc;
CountAllVP := proc (e) options operator, arrow;
  nops(AllVP(e)) end proc
```

Create programs to derive the BVDP and evaluate the Wronskian of its monomials.

```
with(LinearAlgebra);
with(Groebner);
lieDer2 := proc (H, F) local N, V, vars, out;
  N := numelems(F);
  vars := [seq(x[t], t = 1 .. N), seq(y[t], t = 0 .. N+1)];
```

```

V := map(proc (a, b) options operator, arrow;
diff(b, a) end proc, vars, H);
out := DotProduct(Vector([seq(F(t), t = 1 .. N), seq(y[t+1], t = 0 .. N+1)]), Vector(V), conjugate = false) end p

listLieDer2 := proc (H, F, N) local L, i, tmp;
L := [y[0]-H];
tmp := H;
for i to N do tmp := lieDer2(tmp, F);
L := [op(L), y[i]-tmp] end do end proc;

BVDP := proc (F, H, In, Na, Nd) local N, Ga, Gd, La, Ld, i, G2, dG, dG2, vars, generators;
N := nops(F);
La := listLieDer2(H, subs(In[1], F), 1);
Ld := listLieDer2(H, subs(In[2], F), 1);
Ga := [seq(numer(La[i])*denom(La[i]), i = 1 .. numelems(La))];
Gd := [seq(numer(Ld[i])*denom(Ld[i]), i = 1 .. numelems(Ld))];
vars := [seq(x[t], t = 1 .. N)];
for i to Na-1 do G2 := Basis([op(Ga)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Ga := [op(G2), op(dG2)] end do;
for i to Nd-1 do G2 := Basis([op(Gd)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Gd := [op(G2), op(dG2)] end do;
generators := [op(subs(y = ya, Ga)), op(subs(y = yd, Gd))];
G2 := Basis(generators, plex(op(vars), seq(ya[i], i = 0 .. Na), seq(yd[i], i = 0 .. Nd)));
remove(has, G2, x) end proc;

NonZeroSols := proc (S) local S1, S2, m, n, Res;
S1 := '~'[convert](S, list);
S2 := '~'[rhs](Array(S1));
m, n := op(S2)[1 .. 2];
Res := seq(AllNonZero(S2[i]), i = m);

```

```
ListTools:-SearchAll(true, [Res]) end proc;
```

```
BVDPwronskian := proc (BVDP, MaxDiff) coeffs(BVDP, [seq(ya[i], i = 0 .. MaxDiff), seq(yd[i], i = 0 .. MaxDiff)])
```

Use these programs to calculate and test the BVDP.

```
bvdp := BVDP(Vector(F), H, [[C[T] = C[T]], [C[T] = 0]], 1, 1)
bvdpCoeffs := [coeffs(bvdp[1], [ya[1], yd[0], yd[1]])]
EqsBVDP := '‘[‘=‘](bvdpCoeffs, subs([k = kh, R = Rh, C = Ch], bvdpCoeffs));
S1 := [solve(EqsBVDP)];
BVDPsols := remove(proc (s) options operator, arrow;
true in map(proc (e) options operator, arrow;
‘assuming‘([is(rhs(e) <= 0)], [positive]) end proc, s) end proc, S1);
BVDPsols := op(BVDPsols)
```

Create programs that allow the TSA to be used with limits and assuming the parameter relationships derived in the BVDP.

```
GTS3 := proc (H, F, In, N0, Na, Nd, BVDPsols) local H0, Ha, Hd, hatsubs, H0hat, Hdhat, Hahat, Sols, Nvars;
Nvars := numelems(F);
H0 := listLieDer2(H, subs(In[1], F), N0);
H0 := subs([seq(x[i] = 0, i = 1 .. Nvars), op(In[1]), y[0] = 0], H0);
Ha := listLieDer2(H, subs(In[1], F), Na);
Ha := subs(y = ya, Ha);
Ha := subs(ya[0] = yd[0], Ha);
Hd := listLieDer2(H, subs(In[2], F), Nd);
Hd := subs(y = yd, Hd);
hatsubs := [k = kh, C = Ch, R = Rh, x = xh];
H0hat := subs(hatsubs, H0);
Hdhat := subs(hatsubs, Hd);
Hahat := subs(hatsubs, Ha);
Sols := [solve([op('‘[‘-‘](Ha, Hahat)), op('‘[‘-‘](Hd, Hdhat)), op('‘[‘-‘](H0, H0hat)), op(BVDPsols)]);
Sols := remove(proc (s) options operator, arrow;
```

```

true in map(proc (e) options operator, arrow;
'assuming'([is(rhs(e) <= 0)], [positive]) end proc, s) end proc, Sols);
Sols := select(proc (s) options operator, arrow; andmap(hastype, remove(evalb, s), hname) end proc, Sols) end proc;

GTS3timer2 := proc (H, F, Na, Nd, BVDPsols) local st;
st := time();
try timelimit(2000, GTS3(H, Vector(F), [[C[T] = C[T]], [C[T] = 0]], Na, Nd, Nd, BVDPsols)) catch "time expired": p
n", Nd) finally time()-st end try end proc

```

Measure the time taken to produce a solution.

```

Nd := 4;
Na := 1;
v1 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 2;
v2 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 3;
v3 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 4;
v4 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 5;
v5 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 6;
v6 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 7;
v7 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 8;
v8 := [seq(GTS3timer2(H, F, Na, i, BVDPsols), i = 1 .. Nd)];
Na := 1;

```

Measure the dimesnion of the solution.

```

Na := 1;
Dimv1 := seq(max(''[Dims](GTS3(H, Vector(F), Input, Na, i, i, BVDPsols))), i = 1 .. 4);
Na := 2;
Dimv2 := seq(max(''[Dims](GTS3(H, Vector(F), Input, Na, i, i, BVDPsols))), i = 1 .. 4);
Na := 3;
Dimv3 := seq(max(''[Dims](GTS3(H, Vector(F), Input, Na, i, i, BVDPsols))), i = 1 .. 4);
Na := 4;
Dimv4 := seq(max(''[Dims](GTS3(H, Vector(F), Input, Na, i, i, BVDPsols))), i = 1 .. 4);

```

It cannot decrease any further, so there is no point in doing this for larger values of `Na`. The results of `Dims` also include the dimension of the original parameter vector and statespace vector which add 8 to the result, and they need this to be taken away before they will give the output shown in Fig. 3.9(a).

## A.4 Derivations

In Subsections 5.2.1 and 5.2 formula for the steady states in the association phase, and rates of change at these points in the dissociation are referred to. Code to derive these formulae is given:

```

# define models
L1 := diff(x(t), t) = k[a]*In(t)*(R-x(t))-k[d]*x(t);
ERC1 := diff(x(t), t) = (k[a]*In(t)*(R-x(t))-k[d]*x(t))
/(1+k[a]*(R-x(t))/k[m]);
BIV := [diff(x[1](t), t) = k[a1]*In(t)*(R-x[1](t)-2*x[2](t))
-k[d1]*x[1](t)-k[a2]*x[1](t)*(R-x[1](t)-2*x[2](t))+2*k[d2]*x[2](t),
diff(x[2](t), t) = k[a2]*x[1](t)*(R-x[1](t)-2*x[2](t))-2*k[d2]*x[2](t)];
BERC := [diff(x[1](t), t) = (2*k[a1]*In(t)*d*(R-x[1](t)-2*x[2](t))-k[d1]*x[1](t))
/(1+2*k[a1]*(R-x[1](t)-2*x[2](t))/k[m])
-k[a2]*x[1](t)*(R-x[1](t)-2*x[2](t))+2*k[d2]*x[2](t),
diff(x[2](t), t) = k[a2]*x[1](t)*(R-x[1](t)-2*x[2](t))-2*k[d2]*x[2](t)];

```

```

BERCs := [diff(x[1](t), t) = k[a1]*(k[m]*In(t)-k[d1]*x[1](t)-k[d1]*x[m](t))
*(R[b]-x[1](t)-2*x[2](t))/((x[1](t)+2*x[2](t)+x[m](t)-R[b]-R[m])*k[a1]+k[m])
-k[d1]*x[1](t)-k[a2]*x[1](t)*(R[b]-x[1](t)-2*x[2](t))+2*k[d2]*x[2](t),
diff(x[2](t), t) = k[a2]*x[1](t)*(R[b]-x[1](t)-2*x[2](t))-2*k[d2]*x[2](t)
, diff(x[m](t), t) = k[a1]*(k[m]*In(t)-k[d1]*x[1](t)-k[d1]*x[m](t))
*(R[m]-x[m](t))/((x[1](t)+2*x[2](t)+x[m](t)-R[b]-R[m])*k[a1]+k[m])
-k[d1]*x[m](t)];
timeless := [x[m](t) = x[m], x[1](t) = x[1], x[2](t) = x[2], x(t) = x];
# Get stationary points in the association phase
S_xL1 := x = solve(subs([op(timeless), In(t) = C[T]], rhs(L1)), x);
S_xERC1 := x = solve(subs([op(timeless), In(t) = C[T]], rhs(ERC1)), x);
S_xBIV := solve(subs([op(timeless), In(t) = C[T]], map(rhs, BIV)),
[x[1], x[2]]);
S_xBERC := solve(subs([op(timeless), In(t) = C[T]], map(rhs, BERC)),
[x[1], x[2]]);
S_xBERCs := solve(subs([op(timeless), In(t) = C[T]], map(rhs, BERCs)),
[x[1], x[2], x[m]]);

# Get output at stationary point
#
S_yL1 := S_xL1*alpha; S_yERC1 := S_xERC1*alpha;
S_yBIV := alpha*(S_xBIV[1, 1]+S_xBIV[1, 2]);
S_yBERC := alpha*(S_xBERC[1, 1]+S_xBERC[1, 2]);
S_yBERCs := alpha*(S_xBERCs[1, 1]+S_xBERCs[1, 2]+S_xBERCs[1, 3]);
# Get initial ROC in the dissociation phase
#
S_dxL1 := rhs(subs(timeless, [S_xL1, In(t) = 0], L1));
S_dxERC1 := collect(simplify(rhs(subs(timeless, [S_xERC1, In(t) = 0], ERC1))), k[a]);
S_dxBIV := subs(timeless, [op(op(S_xBIV)), In(t) = 0], map(rhs, BIV));
S_dxBERC := subs(timeless, [op(op(S_xBERC)), In(t) = 0], map(rhs, BERC));
S_dxBERCs := subs(timeless, [op(op(S_xBERCs)), In(t) = 0], map(rhs, BERCs));

# Get initial ROC of output in the dissociation phase

```

```
#
S_dyL1 := alpha*S_dxL1; S_dyERC1 := alpha*S_dxERC1;
S_dyBIV := alpha*(S_dxBIV[1]+S_dxBIV[2]);
S_dyBERC := alpha*(S_dxBERC[1]+S_dxBERC[2]);
S_dyBERCs := alpha*(S_dxBERCs[1]+S_dxBERCs[2]+S_dxBERCs[3]);
```

## A.5 Identifiability analyses

Results from analyses are presented. Complete code for two key examples is presented here.

### A.5.1 LwT Identifiability

Input the model.

```
F1 := Vector([(-k[a]*x[2]*(R-x[1])+k[d]*x[1]+k[m]*(Inp-x[2]))/k[h],
k[a]*x[2]*(R-x[1])-k[d]*x[1]]);
H1 := alpha*x[1];
Input := [[Inp = C[T]], [Inp = 0]]
```

Define programs required to find the BVDP and test it.

```
with(LinearAlgebra);
with(Groebner);
lieDer2 := proc (H, F) local N, V, vars, out;
N := numelems(F);
vars := [seq(x[t], t = 1 .. N), seq(y[t], t = 0 .. N+1)];
V := map(proc (a, b) options operator, arrow; diff(b, a) end proc, vars, H);
out := DotProduct(Vector([seq(F(t), t = 1 .. N),
seq(y[t+1], t = 0 .. N+1)]), Vector(V), conjugate = false) end proc;
```

```

listLieDer2 := proc (H, F, N) local L, i, tmp;
L := [y[0]-H];
tmp := H;
for i to N do tmp := lieDer2(tmp, F);
L := [op(L), y[i]-tmp] end do end proc;

BVDP := proc (F, H, In, Na, Nd) local N, Ga, Gd, La, Ld, i, G2, dG, dG2, vars, generators;
N := numelems(F);
La := listLieDer2(H, subs(In[1], F), 1);
Ld := listLieDer2(H, subs(In[2], F), 1);
Ga := [seq(numer(La[i])*denom(La[i]), i = 1 .. numelems(La))];
Gd := [seq(numer(Ld[i])*denom(Ld[i]), i = 1 .. numelems(Ld))];
vars := [seq(x[t], t = 1 .. N)];
for i to Na-1 do G2 := Basis([op(Ga)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Ga := [op(G2), op(dG2)] end do;
for i to Nd-1 do G2 := Basis([op(Gd)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Gd := [op(G2), op(dG2)] end do;
generators := [op(subs(y = ya, Ga)), op(subs(y = yd, Gd))];
G2 := Basis(generators, plex(op(vars), seq(ya[i], i = 0 .. Na), seq(yd[i], i = 0 .. Nd)));
remove(has, G2, x) end proc;

NonZeroSols := proc (S) local S1, S2, m, n, Res;
S1 := '~'[convert](S, list);
S2 := '~'[rhs](Array(S1));
m, n := op(S2)[1 .. 2];
Res := seq(AllNonZero(S2[i]), i = m);
ListTools:-SearchAll(true, [Res]) end proc;

BVDPwronskian := proc (BVDP, MaxDiff) coeffs(BVDP, [seq(ya[i], i = 0 .. MaxDiff), seq(yd[i], i = 0 .. MaxDiff)], M

```

Find the BVDP and test it.

```

S1 := BVDP(F1, H1, Input, 1, 1);
unassign('S1Monos'); S1Coeffs := [coeffs(S1[1], [ya[1], yd[1]], S1Monos)];
  ''[print](S1Coeffs)[]; ''[print](S1Monos)[];
ReducedCoeffs := [op(''[simplify](''['/'](S1Coeffs, S1Coeffs[1])))];
EqsB := ''['='](ReducedCoeffs, subs([k = kh, R = Rh, C = Ch], ReducedCoeffs));
S1 := [solve(EqsB)];
  BVDPsols := remove(proc (s) options operator, arrow; true in map(proc (e) options operator, arrow; 'as
  BVDPsols := op(BVDPsols);

```

Create a special version of GTS, that takes in expressions like S1 to make solve run faster.

```

*TypeTools:-AddType(hname, proc (n) local r;
  r := n;
  if not n::name then return false end if;
  while not r::symbol do r := op(0, r) end do;
  evalb(substring(r, -1 .. -1) = ':-h') end proc);
GTS := proc (H, F, In, N0, Na, Nd, BVDPsols) local H0, Ha, Hd, hatsubs, H0hat, Hdha;
  Nvars := nops(F);
  H0 := listLieDer2(H, subs(In[1], F), N0);
  H0 := subs([x[1] = 0, x[2] = 0, op(In[1]), y[0] = 0], H0);
  Ha := listLieDer2(H, subs(In[1], F), Na);
  Ha := subs(y = ya, Ha);
  Ha := subs(ya[0] = yd[0], Ha);
  Hd := listLieDer2(H, subs(In[2], F), Nd);
  Hd := subs(y = yd, Hd);
  hatsubs := [k = kh, C = Ch, R = Rh, x = xh];
  H0hat := subs(hatsubs, H0);

```

```
Hdhat := subs(hatsubs, Hd);
Hahat := subs(hatsubs, Ha);
[op('~~['-'](Ha, Hahat)), op('~~['-'](Hd, Hdhat)), op('~~['-'](H0, H0hat)), op(BVDPsols)] end
```

Run the above and solve its output.

```
S2 := GTS(H1, Vector(F1), [[Inp = C[T]], [Inp = 0]], 5, 5, 5, BVDPsols);
Sols := [solve(S2)]; S3 := remove(proc (s) options operator, arrow; true in map(proc (e) optio
```

S2[2] is not a meaningful solution as it requires a parameter  $kh[m]$  to be a function of a variable, and it can't as it must be constant and the variable will vary.

Use eliminate to view the remaining output in a more convenient way.

```
eliminate(S3[1], [xh[1], xh[2], Rh, kh[a], kh[m], Ch[T], kh[d]])
```

## A.5.2 BERCS Identifiability

Derive the model.

```
F1 := Vector([
k[a1]*C*(R[b]-x[1]-2*x[2])-k[d1]*x[1]-(diff(x[2](t), t)),
k[a2]*x[1]*(R[b]-x[1]-2*x[2])-2*k[d2]*x[2],
k[a1]*C*(R[m]-x[3])-k[d1]*x[3]
]);
F2 := subs(diff(x[2](t), t) = F1[2], F1);
Croc := -F2[1]-F2[2]-F2[3]+k[m]*(In1*C[T]-C);
QSS := solve(0 = Croc, C);
nQSS := collect(numer(QSS), k[d1]);
dQSS := collect(denom(QSS), k[a1]);
QSS := nQSS/dQSS;
F3 := subs(C = QSS, F2)
G := alpha*(x[1]+x[2]+x[3]); In := [[In1 = 1], [In1 = 0]]
```

Set up programs.

```

with(LinearAlgebra);
with(Groebner);
lieDer2 := proc (H, F) local N, V, vars, out;
N := nops(F);
vars := [seq(x[t], t = 1 .. N), seq(y[t], t = 0 .. N+1)];
V := map(proc (a, b) options operator, arrow; diff(b, a) end proc, vars, H);
out := DotProduct(Vector([seq(F(t), t = 1 .. N), seq(y[t+1], t = 0 .. N+1)]), Vector(V), conjugate = 1);

listLieDer2 := proc (H, F, N) local L, i, tmp;
L := [y[0]-H];
tmp := H;
for i to N do tmp := lieDer2(tmp, F);
L := [op(L), y[i]-tmp] end do end proc;

BVDP := proc (F, H, In, Na, Nd) local N, Ga, Gd, La, Ld, i, G2, dG, dG2, vars;
N := nops(F);
La := listLieDer2(H, subs(In[1], F), 1);
Ld := listLieDer2(H, subs(In[2], F), 1);
Ga := [seq(numer(La[i])*denom(La[i]), i = 1 .. numelems(La))];
Gd := [seq(numer(Ld[i])*denom(Ld[i]), i = 1 .. numelems(Ld))];
vars := [seq(x[t], t = 1 .. N)];
for i to Na-1 do G2 := Basis([op(Ga)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Ga := [op(G2), op(dG2)] end do;
for i to Nd-1 do G2 := Basis([op(Gd)], plex(op(vars)));
dG := [seq(lieDer2(G2[j], F), j = 1 .. numelems(G2))];
dG2 := [seq(numer(dG[i])*denom(dG[i]), i = 1 .. numelems(dG))];
Gd := [op(G2), op(dG2)] end do;
G2 := Basis([op(subs(y = ya, Ga)), op(subs(y = yd, Gd))], plex(op(vars), seq(ya[Na-i], i = 0 .. Na)),
NonZeroSols := proc (S) local S1, S2, m, n, Res;

```

```
S1 := '~'[convert](S, list);
S2 := '~'[rhs](Array(S1));
m, n := op(S2)[1 .. 2];
Res := seq(AllNonZero(S2[i]), i = m);
ListTools:-SearchAll(true, [Res]) end proc;
```

```
BVDPwronskian := proc (BVDP, MaxDiff) coeffs(BVDP, [seq(ya[i], i = 0 .. MaxDiff), seq(yd[i], i = 0 .. MaxDiff)], M
```

Run these programs and obtain the Groebner bases the BVDP is a part of.

```
S1 := BVDP(F3, G, In, 2, 2);
S2 := BVDP(F3, G, In, 2, 3);
```

The second elements of both of these are BVDPs. Obtain relationships between parameters that the BVDPs require.

```
unassign('S1Monos');
S1Coeffs := [coeffs(S1[2], [seq(ya[i], i = 0 .. 3), seq(yd[i], i = 0 .. 3)], S1Monos)];
'~'[print](S1Coeffs) [];
'~'[print](S1Monos) [];
unassign('S2Monos');
S2Coeffs := [coeffs(S2[2], [seq(ya[i], i = 0 .. 3), seq(yd[i], i = 0 .. 3)], S2Monos)];
'~'[print](S2Coeffs) [];
'~'[print](S2Monos) [];
ReducedCoeffs := [op('~'[simplify]('~'['/'](S1Coeffs, S1Coeffs[4]))), op('~'[simplify]('~'['/'](S2Coeffs, S2Coeffs[4])))];
EqsB := '~'['='](ReducedCoeffs, subs([k = kh, R = Rh, C = Ch], ReducedCoeffs));
S1 := [solve(EqsB)];
BVDPsols := remove(proc (s) options operator, arrow; true in map(proc (e) options operator, arrow; 'assuming'([is(
BVDPsols := op(BVDPsols);
```

Set up the TSA method, so it can assume these relationships.

```
TypeTools:-AddType(hname, proc (n) local r;
```

```

r := n;
if not n::name then return false end if;
while not r::symbol do r := op(0, r) end do;
evalb(substring(r, -1 .. -1) = ':-h') end proc);

GTS := proc (H, F, In, NO, Na, Nd, BVDPsols) local H0, Ha, Hd, hatsubs, H0hat, Hdhat, Hahat;
Nvars := nops(F);
H0 := listLieDer2(H, subs(In[1], F), NO);
H0 := subs([seq(x[i] = 0, i = 1 .. Nvars), op(In[1]), y[0] = 0], H0);
Ha := listLieDer2(H, subs(In[1], F), Na);
Ha := subs(y = ya, Ha);
Ha := subs(ya[0] = yd[0], Ha);
Hd := listLieDer2(H, subs(In[2], F), Nd);
Hd := subs(y = yd, Hd);
hatsubs := [k = kh, C = Ch, R = Rh, x = xh];
H0hat := subs(hatsubs, H0);
Hdhat := subs(hatsubs, Hd);
Hahat := subs(hatsubs, Ha);
Sols := [solve([op('~'['-'](Ha, Hahat)), op('~'['-'](Hd, Hdhat)), op('~'['-'](H0, H0hat))],
Sols := remove(proc (s) options operator, arrow;
true in map(proc (e) options operator, arrow;
'assuming'([is(rhs(e) <= 0)], [positive]) end proc, s) end proc, Sols);
Sols := select(proc (s) options operator, arrow;
andmap(hastype, remove(evalb, s), hname) end proc, Sols) end proc;
Dims := proc (S::(set('='))) options operator, arrow;
nops(select(evalb, S)) end proc

```

Run these programs using the largest numbers of derivatives that this function will take

without it causing stability issues. This will vary between computers but the following work on mine.

```
S2 := GTS(G, F3, In, 2, 1, 1, BVDPsols);
numelems(S2);
Dims(op(S2))-3-8
```

Set up programs for the ORC method.

```
Ha := '‘[+‘](-listLieDer2(G, subs(In[1], F3), 2), [y[0], y[1], y[2]]);
VectorCalculus:-Jacobian(Ha, [x[1], x[2], x[3]]);
subs([op(In[1]), x[1] = 0, x[2] = 0, x[3] = 0], %);
Rank(%)
```

H has full rank at the initial conditions. Now find H for the diss phase and then find lambda for both.

```
hatsubs := [k[a1] = kh[a1], k[a2] = kh[a2], k[d1] = kh[d1], k[d2] = kh[d2],
k[m] = kh[m], C[T] = Ch[T], R[b] = Rh[b], R[m] = Rh[m]];
hatsubs := subs(eliminateS2[1], hatsubs);
lambdasubs := [x = lambda]; '‘[‘=‘](Ha, subs(hatsubs, lambdasubs, Ha));
LambdaMapA := op(solve(%, [lambda[1], lambda[2], lambda[3]]));
Hd := '‘[+‘](-listLieDer2(G, subs(In[2], F3), 2), [y[0], y[1], y[2]]);
'‘[‘=‘](Hd, subs(hatsubs, lambdasubs, Hd));
LambdaMapD := op(solve(%, [lambda[1], lambda[2], lambda[3]]))
```

Notably in either case  $\text{lambda}(x[2])=x[2]$  for indistinguishable solutions; i.e.  $x[2]$  can be deduced from the output.

As a result the monomial coefficients of  $x[2]$  must be conserved on indistinguishable solutions. This gives us extra information about lambda.

```
collect(F3[2], x[2]);
x2Coeffs := [coeffs(%, x[2])];
‘‘[‘=‘](x2Coeffs, subs(hatsubs, lambdasubs, x2Coeffs));
```

Re-calculate `lambda` assuming that the relationships in `x2Coeffs` hold.

```
Ha := ‘‘[‘+‘](-listLieDer2(G, subs(In[1], F3), 2), [y[0], y[1], y[2]]);
‘‘[‘=‘](Ha, subs(hatsubs, lambdasubs, Ha));
Eqs := [op(%), op(‘‘[‘=‘](x2Coeffs, subs(hatsubs, lambdasubs, x2Coeffs)))]);
lambdaMapA2 := solve(%, [lambda[1], lambda[2], lambda[3], Rh[m]])
```

We know that the value of a parameter is time invariant so  $0 = dp/dt = (dp/dx)*(dx/dt)$ , and as the system doesn't initiate on the steady state,  $dp/dx = 0$ .

```
diff(lambdaMapA2[1, 4], x[1]);
simplify(%, size);
0 = numer(rhs(%));
collect(%, x[1]);
[coeffs(rhs(%), x[1])];
[solve(%)];
EqsA2 := remove(proc (s) options operator, arrow; true in map(proc (e) options oper
```

There are 2 cases which can be resubstituted into `lambda`.

```
LambdaMapA3 := ‘‘[‘‘[simplify]](subs(EqsA2[1], LambdaMapA2), size):
‘‘[‘‘[simplify]](subs(EqsA2[2], LambdaMapA2), size);
```

In the second case everything is already solved and the relationships printed in Subsection 5.3.6 are obtained, but in the first it is not. Now we consider the first case with the second equation from eq.3.22.

```
hatsubs;
```

```
VectorCalculus:-Jacobian('`[rhs](LambdaMapA3[1, 1 .. 3]), [x[1], x[2], x[3]]);
Eq2s := convert('`['-'(F3, Multiply(%, subs(hatsubs, F3))), list);
```

Evaluate the coefficients of each element of this equation.

```
Temp := '`['`[numer]](Eq2s);
C1s := coeffs(expand(Temp[1]), [x[1], x[2], x[3]]);
C2s := coeffs(expand(Temp[2]), [x[1], x[2], x[3]]);
C3s := coeffs(expand(Temp[3]), [x[1], x[2], x[3]])
```

Solve these to obtain parameter relationships. This again gives the relationships printed in Subsection 5.3.6

```
[solve([C1s, C2s, C3s, In1 = 1])];
remove(proc (s) options operator, arrow; true in map(proc (e) options operator, arrow; 'assuming'([is(rhs(e) <= 0)
```

## A.6 Sample facsimile program

```
%* ~~~~~;
%* 27-Jan-2014 11:22:48;
%* from Templatesng2for3x1monoclonalstart.fac;
%* and Templatesng2for3x1monoclonalend.fac;
%* with data from SPR_DATA SN607D8-1;
%* Ligands [1] Analytes [1 2 3];
%* on the computer minimal-e8b3e0f;
%* ERC (QSS) Model for SPR data ;
%* Homogeneous analyte hetrogeneous binding version ;
%* ~~~~~;

EXECUTE ;
OPEN 10 "plotd211212p11.out";
```

```

OPEN 11 "plotd211212p12.out";
OPEN 12 "plotd211212p1.txt";

**;

*SETHMAX 0.1 ;

%* Define the variables ;
%* Bijk -> bound conc ;
%*      all data is from one ligand channel i,
%*      analyte channel j (1--5),
%*      binding type k (1--2);
%*
%* cT1 is not varied and CT2 is unused
%*
%* This model includes a step function yijt that takes two values,
%* to model the bulk effect, yija and 0, this can be taken out
%* by deleting yija and yijd from
%* the SETVARY execution ;
%*

VARIABLE
B11 B21 B31
B12 B22 B32
;

* Define the parameters ;
PARAMETER
alpha 1000 dum 1
d1 0.1 d2 0.05 d3 0.025
d4 1 d5 1 d6 1
Kon11 Kon21
Kon12 Kon22
Koff11 Koff21

```

## APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
Koff12 Koff22
R1 600 R2 640 R3 680
R4 R5 R6
kM1 kM2 kM3
kM4 kM5 kM6
cT1 100
cT2 100
RSS
y11a 18.4 y12a 16.9 y13a 8.1
y11d 0 y12d 0 y13d 0
TEMP
;

PARAMETER
Y1 Y2 Y3
cIN 0
;

INTEGER #COUNT ;

* Open file, read in and assign parameters ;
COMPILE INSTANT;
OPEN 7 "parsd211212p1.txt";
*;
READ 7 RSS ;
* binding parameters;
READ 7 Kon11 ;
READ 7 Kon12 ;
READ 7 TEMP ;
READ 7 TEMP;
* unbinding parameters;
READ 7 Koff11 ;
READ 7 Koff12 ;
```

```
READ 7 TEMP ;
READ 7 TEMP ;
* chip parameters;
READ 7 TEMP ;
READ 7 TEMP;
* flow parameters;
READ 7 kM1 ;
READ 7 kM2 ;
READ 7 kM3 ;
READ 7 TEMP ;
READ 7 TEMP ;
READ 7 TEMP ;
* concentration parameters;
* READ 7 cT1 ;
* READ 7 cT2 ;
**;

* Assign initial conditions for variables ;
COMPILE INITIAL;
B11 = 0 ;
B21 = 0 ;
B31 = 0 ;
*;
B12 = 0 ;
B22 = 0 ;
B32 = 0 ;
**;

COMPILE GENERAL ;
```

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
Y1 = alpha*B11 + alpha*B12 + cIN*y11a + (1 - cIN)*y11d;
Y2 = alpha*B21 + alpha*B22 + cIN*y12a + (1 - cIN)*y12d;
Y3 = alpha*B31 + alpha*B32 + cIN*y13a + (1 - cIN)*y13d;
**;

%*EQUATION NOTES
%*notably these are for a single antibody binding in multiple ways;
%*for multiple antibodies binding in single ways change (cT1) for (1-cT1);
%*in the second binding type, or develop a way of converting;
%*;

COMPILE EQUATIONS;
*first binding type;
'B11 = Kon11*d1*cT1*cIN*(R1-B11-B12)-B11*Koff11;
'B21 = Kon11*d2*cT1*cIN*(R2-B21-B22)-B21*Koff11;
'B31 = Kon11*d3*cT1*cIN*(R3-B31-B32)-B31*Koff11;
*second binding type;
'B12 = Kon12*d1*cT1*cIN*(R1-B11-B12)-B12*Koff12;
'B22 = Kon12*d2*cT1*cIN*(R2-B21-B22)-B22*Koff12;
'B32 = Kon12*d3*cT1*cIN*(R3-B31-B32)-B32*Koff12;
**;

COMPILE CHANGE ;
IF (cIN) *, *, 10 ;
cIN = 2 ;
LABEL 10 ;
cIN = cIN - 1 ;
**;

DATA 0.027;
TIME Y1 Y2 Y3 ;
RANGE 673.3 637.6 596.1;
```

```
0.0 14.2 11.1 15.0;
0.9 13.8 10.4 16.8;
1.8 15.2 8.5 13.5;
2.7 14.3 10.4 17.2;
3.6 15.2 8.1 14.2;
4.5 15.0 10.1 16.8;
5.4 12.0 10.9 14.4;
6.3 13.8 9.9 14.8;
7.2 14.9 10.1 14.1;
8.1 12.4 4.0 10.9;
9.0 0.0 0.0 0.0;
9.9 33.3 28.9 18.8;
10.8 58.8 45.8 34.4;

*timepoints have been removed from the datablock
*to reduce the length of the program;

1981.8 531.7 510.5 491.5;
1982.7 532.3 510.1 493.6;
1983.6 529.0 509.6 493.6;
1984.5 532.2 510.4 495.0;
1985.4 531.4 509.9 493.7;
1986.3 531.9 509.3 492.4;
1987.2 528.9 509.8 493.5;
1988.1 529.6 508.2 493.1;
1989.0 531.7 510.6 493.2;
1989.9 531.0 508.7 494.1;
1990.8 530.8 509.6 493.4;
1991.7 529.6 509.7 492.0;
1992.6 530.8 508.9 492.1;

**;

SETVARY
Kon11 Kon12
Koff11 Koff12
```

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
R1 R2 R3
kM1 kM2 kM3
;

COMPILE INSTANT ;
WRITE 1=10, " PRINT STREAM NO. " % ;
WRITE 1, "TIME Y1 Da1 Y2 Da2 Y3 Da3" % ;
#COUNT=0 ;
**;

COMPILE PRINT;
* Output routine called during the final phase;
#COUNT = #COUNT + 1 ;
DO 10 FOR #2=#COUNT-1 ;
WRITE 1=10, ((E14,6)) TIME, Y1, VOBS<0,#2>, Y2, VOBS<1,#2>, Y3,
  VOBS<2,#2> ;
LABEL 10;
**;

WHENEVER TIME = 9 965.7 CALL CHANGE RESTART ;
**;
*previously 7.2 and 981;

BEGIN;

SETNOFIT;
WHENEVER TIME = TOBS % CALL PRINT;
**;

COMPILE INSTANT;
* calculate RSS;
RSS = 0 ;
DO 1 FOR #1 = 0 (1) 3 ;
```

```

DO 2 FOR #2 = 0 (1) 2215 ;
    RSS = RSS + RESIDARRAY<#1,#2>*RESIDARRAY<#1,#2> ;
    LABEL 2 ;
LABEL 1 ;
* Print values of parameters;
REWIND 7;
WRITE 1=7, (E14,6) RSS " RSS" % ;
* binding parameters;
WRITE 1, (E14,6) Kon11 " Kon11 in 1/concentration unit seconds" % ;
WRITE 1, (E14,6) Kon12 " Kon12 in 1/concentration unit seconds" % ;
WRITE 1, (E14,6) Kon11 " Kon11 in 1/concentration unit seconds" % ;
WRITE 1, (E14,6) Kon12 " Kon12 in 1/concentration unit seconds" % ;
* ;
WRITE 1, (E14,6) Koff11 " Koff11 in 1/seconds" % ;
WRITE 1, (E14,6) Koff12 " Koff12 in 1/seconds" % ;
WRITE 1, (E14,6) Koff11 " Koff11 in 1/seconds" % ;
WRITE 1, (E14,6) Koff12 " Koff12 in 1/seconds" % ;
* machine parameters;
WRITE 1, (E14,6) R1 " R1 in ng/mm^2" % ;
WRITE 1, (E14,6) R2 " R2 in ng/mm^2" % ;
WRITE 1, (E14,6) R3 " R3 in ng/mm^2" % ;
WRITE 1, (E14,6) R1 " R1 in ng/mm^2" % ;
WRITE 1, (E14,6) R2 " R2 in ng/mm^2" % ;
WRITE 1, (E14,6) R3 " R3 in ng/mm^2" % ;
WRITE 1, (E14,6) kM1 " kM1 in ng/concentration unit mm^2 seconds" % ;
WRITE 1, (E14,6) kM2 " kM2 in ng/concentration unit mm^2 seconds" % ;
WRITE 1, (E14,6) kM3 " kM3 in ng/concentration unit mm^2 seconds" % ;
WRITE 1, (E14,6) kM1 " kM1 in ng/concentration unit mm^2 seconds" % ;
WRITE 1, (E14,6) kM2 " kM2 in ng/concentration unit mm^2 seconds" % ;
WRITE 1, (E14,6) kM3 " kM3 in ng/concentration unit mm^2 seconds" % ;
WRITE 1, (E14,6) cT1 " cT1 is a dimensionless quantity" % ;
*WRITE 1, (E14,6) cT2 " cT2 is a dimensionless quantity" % ;
WRITE 1, (E14,6) y11a " y11a in ng/mm^2" % ;

```

```
WRITE 1, (E14,6) y12a " y12a in ng/mm^2" % ;
```

```
WRITE 1, (E14,6) y13a " y13a in ng/mm^2" % ;
```

```
**;
```

```
BEGIN;
```

```
STOP;
```

## A.7 Sample facsimile Logfile

```
FACSIMILE RELEASE 4.40.102 NVC SM D    DATE 03/05/05
```

```
COPYRIGHT (C) 2003 MCPA Software Ltd.
```

```
FACSIMILE software is the property of
```

```
MCPA Software Ltd.
```

```
Running ...
```

```
QSS2d211212p1.fac
```

```
ARRAY SIZES REQUESTED -
```

```
INTEGER  1160038, DOUBLE  512620, REAL  5000, CHARACTER  140728
```

```
***** MAXIMUM ARRAY SIZES
```

```
** NPARS  500000    REAL SCALARS, ARRAYS, AND WORKSPACE
```

```
** NINTS  500000    INTEGER SCALARS, ARRAYS AND WORKSPACE
```

```
** NCODE  40000     LOCATIONS FOR CODE
```

```
** NAMES  2000      ENTRIES IN DICTIONARY
```

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
** NTEXT      20000   LOCATIONS FOR TEXT (INCLUDING NAMES)
** MATSZ       500    MATRIX FOR INVERT OR SOLVE
** NCURV      2000   DATA CURVES
** NVARY       100   VARIED PARAMETERS
** NOBSM      18000  OBSERVED VALUES
** NPTTN      600000 SPARSITY PATTERN WORKSPACE
***** END OF LIST OF SIZES
```

```
!!!! NOTES FILE MISSING
```

```
* ~~~~~;
* 07-Apr-2014 18:31:34;
* from TemplatesSS2for3x1monoclonalstartfac;
* and TemplatesSS2for3x1monoclonalendfac;
* with data from SPR_DATA SN607D8-1;
* Ligands [1] Analytes [1 2 3];
* on the computer minimal-e8b3e0f;
* ERC (QSS) Model for SPR data ;
* Homogeneous analyte version ;
* ~~~~~;
```

```
EXECUTE ;
OPEN 10 "pltd211212p11.out";
OPEN 11 "plotd211212p12.out";
OPEN 12 "plotd211212p1.txt";
**;
```

```
*SETHMAX 0.1 ;
```

```
* Define the variables ;
* Bijk -> bound conc ;
*      all data is from one ligand channel i,
*      analyte channel j (1--5),
```

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
*      binding type k (1--2);
* Y4-6 was removed or commented out, due to data problems
* as a result there are unused parameters;
*
* cT1 and (100-cT1) were added and varied;
*
* This model includes a step function yijt that takes two values,
* to model the bulk effect, yija and 0, this can be taken out
* by deleting yija and yijd from
* the SETVARY execution ;
*
*Analyte dilutions appear to be all 0.1 mg/ml;
*As a result units off Ka1 are ml/(mg s) not mm3/ng s
*as they were in the budapest paper, so need to be x1000
* and /1000,000 before they can be compared;
*
*Data are taken from LiA3 i in 1,2,3 in the
*processed data 4th feb 2011.xls;
*
*-----;
```

VARIABLE

```
B11 B21 B31 B41 B51 B61
B12 B22 B32 B42 B52 B62
;
```

```
* Define the parameters ;
```

PARAMETER

```
alpha 1000 dum 1
d1 0.1 d2 0.05 d3 0.025
d4 1 d5 1 d6 1
Kon11 Kon21
Kon12 Kon22
```

```
Koff11 Koff21
Koff12 Koff22
R1 600 R2 640 R3 680
R4 R5 R6
kM1 kM2 kM3
kM4 kM5 kM6
cT1 100
cT2 100
RSS
y11a 18.4 y12a 21.9 y13a 20
y11d 0 y12d 0 y13d 0
TEMP
;

PARAMETER
Y1 Y2 Y3
cIN 0
;

INTEGER #COUNT ;

* Open file, read in and assign parameters ;
COMPILE INSTANT;
LOCN INSTRUCTION

74:OPEN 7 "parsd211212p1.txt";
78:*;
78:READ 7 RSS ;
83:* binding parameters;
83:READ 7 Kon11 ;
88:READ 7 Kon12 ;
93:READ 7 TEMP ;
98:READ 7 TEMP;
```

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
103:* unbinding parameters;
103:READ 7 Koff11 ;
108:READ 7 Koff12 ;
113:READ 7 TEMP ;
118:READ 7 TEMP ;
123:* chip parameters;
123:READ 7 TEMP ;
128:READ 7 TEMP ;
133:READ 7 TEMP ;
138:READ 7 TEMP ;
143:READ 7 TEMP ;
148:READ 7 TEMP;
153:* flow parameters;
153:READ 7 km1 ;
158:READ 7 km2 ;
163:READ 7 km3 ;
168:READ 7 TEMP ;
173:READ 7 TEMP ;
178:READ 7 TEMP ;
183:* concentration parameters;
183:* READ 7 cT1 ;
183:* READ 7 cT2 ;
183:**;
```

```
NO. OF VARIABLES = 12
NO. IMPLICIT      = 0
NO. OF ERRORS     = 0
```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	46	499954	500000
REAL SCALARS AND ARRAYS	62	499938	500000
INDEXES AND SECRETS	99	400	499
CODE ARRAY	70	39930	40000

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

NAME DICTIONARY	86	1914	2000
TEXT ARRAY	173	39827	40000

1.0000000e+27  
1.7782794e-03  
1.7782794e-03  
1.7782794e-03  
1.7782794e-03  
1.7782794e-05  
5.6234133e-02  
1.7782794e-05  
5.6234133e-02  
3.5481339e-02  
7.0794578e-02  
3.1622777e-02  
3.5481339e-02  
7.0794578e-02  
3.1622777e-02  
5.6234133e+00  
5.6234133e+00  
5.6234133e+00  
3.1622777e+01  
3.1622777e+01  
3.1622777e+01

\* Assign initial conditions for variables ;

COMPILE INITIAL;

LOCN INSTRUCTION

74:B11 = 0 ;  
80:B21 = 0 ;  
86:B31 = 0 ;  
92:\*

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

92:B12 = 0 ;

98:B22 = 0 ;

104:B32 = 0 ;

110:\*\*;

NO. OF VARIABLES = 12

NO. IMPLICIT = 0

NO. OF ERRORS = 0

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	46	499954	500000
REAL SCALARS AND ARRAYS	63	499937	500000
INDEXES AND SECRETS	99	400	499
CODE ARRAY	111	39889	40000
NAME DICTIONARY	87	1913	2000
TEXT ARRAY	173	39827	40000

COMPILE GENERAL ;

LOCN INSTRUCTION

115:Y1 = alpha\*B11 + alpha\*B12 + cIN\*y11a + (1 - cIN)\*y11d;

148:Y2 = alpha\*B21 + alpha\*B22 + cIN\*y12a + (1 - cIN)\*y12d;

181:Y3 = alpha\*B31 + alpha\*B32 + cIN\*y13a + (1 - cIN)\*y13d;

214:\*\*;

NO. OF VARIABLES = 12

NO. IMPLICIT = 0

NO. OF ERRORS = 0

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	46	499954	500000
REAL SCALARS AND ARRAYS	64	499936	500000

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

INDEXES AND SECRETS	99	400	499
CODE ARRAY	215	39785	40000
NAME DICTIONARY	88	1912	2000
TEXT ARRAY	173	39827	40000

COMPILE EQUATIONS;  
 LOCN INSTRUCTION

```
!!!! EQUATIONS FIXES MAX. NO. OF VARIABLES AS 12
219:*notably these are for a single antibody binding in multiple ways;
219:*for multiple antibodies binding in single ways change (cT1) for (1-cT1)
219:*in the second binding type, or develop a way of converting;
219:*;
219:*first binding type;
219:*;
219:'B11 = (Kon11*d1*cT1*cIN*(R1-B11-B12)-B11*Koff11)
219:/(1+(Kon11/kM1)*(R1-B11-B12));
272:'B21 = (Kon11*d2*cT1*cIN*(R2-B21-B22)-B21*Koff11)
272:/(1+(Kon11/kM2)*(R2-B21-B22));
325:'B31 = (Kon11*d3*cT1*cIN*(R3-B31-B32)-B31*Koff11)
325:/(1+(Kon11/kM3)*(R3-B31-B32));
378:*;
378:*second binding type;
378:*;
378:'B12 = (Kon12*d1*cT1*cIN*(R1-B11-B12)-B12*Koff12)
378:/(1+(Kon12/kM1)*(R1-B11-B12));
431:'B22 = (Kon12*d2*cT1*cIN*(R2-B21-B22)-B22*Koff12)
431:/(1+(Kon12/kM2)*(R2-B21-B22));
484:'B32 = (Kon12*d3*cT1*cIN*(R3-B31-B32)-B32*Koff12)
484:/(1+(Kon12/kM3)*(R3-B31-B32));
537:**;
```

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

NO. OF VARIABLES = 12

NO. IMPLICIT = 0

NO. OF ERRORS = 0

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	131	499869	500000
REAL SCALARS AND ARRAYS	258	499742	500000
INDEXES AND SECRETS	99	400	499
CODE ARRAY	538	39462	40000
NAME DICTIONARY	97	1903	2000
TEXT ARRAY	173	39827	40000

COMPILE CHANGE ;

LOCN INSTRUCTION

542:IF (cIN) \*, \*, 10 ;

549:cIN = 2 ;

555:LABEL 10 ;

555:cIN = cIN - 1 ;

564:\*\*;

NO. OF VARIABLES = 12

NO. IMPLICIT = 0

NO. OF ERRORS = 0

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	131	499869	500000
REAL SCALARS AND ARRAYS	259	499741	500000
INDEXES AND SECRETS	99	400	499
CODE ARRAY	565	39435	40000
NAME DICTIONARY	98	1902	2000



APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

3 CURVES  
2215 TIME  
6642 OBSERVATIONS

POINTS

A TIME POINT INCREMENT OF 1 AND OFFSET OF 0 SPECIFIED  
3 DEPENDENT VARIABLES SELECTED  
2215 TIME POINTS AND 6642 OBSERVATIONS USED

\* compile instant and compile print were deleted;

SETVARY

Kon11 Kon12

Koff11 Koff12

R1 R2 R3

kM1 kM2 kM3

;

COMPILE INSTANT ;

LOCN INSTRUCTION

569:WRITE 1=10, " PRINT STREAM NO. " % ;

578:WRITE 1, "TIME Y1 Da1 Y2 Da2 Y3 Da3" % ;

585:#COUNT=0 ;

588:\*\*;

NO. OF VARIABLES = 12

NO. IMPLICIT = 0

NO. OF ERRORS = 0

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	147	499853	500000
REAL SCALARS AND ARRAYS	24659	475341	500000

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

INDEXES AND SECRETS	99	400	499
CODE ARRAY	565	39435	40000
NAME DICTIONARY	112	1888	2000
TEXT ARRAY	216	39784	40000

COMPILE PRINT;

LOCN INSTRUCTION

```

569:* Output routine called during the final phase;
569:#COUNT = #COUNT + 1 ;
578:DO 10 FOR #2=#COUNT-1 ;
589:WRITE 1=10, ((E14,6)) TIME, Y1, VOBS<0,#2>, Y2, VOBS<1,#2>, Y3,
589: VOBS<2,#2> ;
627:LABEL 10;
637:**;

```

```

NO. OF VARIABLES = 12
NO. IMPLICIT     = 0
NO. OF ERRORS   = 0

```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	150	499850	500000
REAL SCALARS AND ARRAYS	24659	475341	500000
INDEXES AND SECRETS	100	399	499
CODE ARRAY	638	39362	40000
NAME DICTIONARY	115	1885	2000
TEXT ARRAY	216	39784	40000

WHENEVER TIME = 9.9 966.6 CALL CHANGE RESTART ;

\*\*;

\*previously 7.2 and 981;

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

BEGIN;

SPACE FOR	IN USE	FREE	TOTAL
REAL SCALARS AND ARRAYS	171047	328953	500000

SIZE NEEDED FOR ARRAY WSFIT = 146374

OPTIMISATION RUN WITH 6642 OBSERVATIONS AND 10 VARIED PARAMETERS

TIME	0.0000000E+00	TCPU	0.00	0.00							
STP	0	TRY	0	DER	0	JAC	0	LUD	0	RST	0

AFTER 0 ITERATIONS AND 1 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
2.2523E+12	-46.6543	0.9998	0.0212
2.7078E+12	-46.3195	0.9998	0.0212
3.0511E+12	-45.6941	0.9997	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 8.01115E+12

MEAN ABSOLUTE CORRELATION INDEX IS 46.22263

VARY( 1):KON11	=	1.7783E-03
VARY( 2):KON12	=	1.7783E-03
VARY( 3):KOFF11	=	1.7783E-05
VARY( 4):KOFF12	=	5.6234E-02
VARY( 5):R1	=	6.0000E+02
VARY( 6):R2	=	6.4000E+02
VARY( 7):R3	=	6.8000E+02
VARY( 8):KM1	=	5.6234E+00
VARY( 9):KM2	=	5.6234E+00
VARY( 10):KM3	=	5.6234E+00

1 SIMULATION RUNS COMPLETED

2 SIMULATION RUNS COMPLETED

3 SIMULATION RUNS COMPLETED

4 SIMULATION RUNS COMPLETED  
5 SIMULATION RUNS COMPLETED  
6 SIMULATION RUNS COMPLETED  
7 SIMULATION RUNS COMPLETED  
8 SIMULATION RUNS COMPLETED  
9 SIMULATION RUNS COMPLETED  
10 SIMULATION RUNS COMPLETED

AFTER 0 ITERATIONS AND 11 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
2.2515E+12	-46.6559	0.9998	0.0212
2.7065E+12	-46.3221	0.9998	0.0212
3.0489E+12	-45.6980	0.9997	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 8.00689E+12  
MEAN ABSOLUTE CORRELATION INDEX IS 46.22532

VARY( 1):KON11 = 1.7783E-03  
VARY( 2):KON12 = 1.8233E-03  
VARY( 3):KOFF11 = 1.8233E-05  
VARY( 4):KOFF12 = 5.6234E-02  
VARY( 5):R1 = 6.0000E+02  
VARY( 6):R2 = 6.4000E+02  
VARY( 7):R3 = 6.8000E+02  
VARY( 8):KM1 = 5.6234E+00  
VARY( 9):KM2 = 5.6234E+00  
VARY( 10):KM3 = 5.7658E+00

11 SIMULATION RUNS COMPLETED  
12 SIMULATION RUNS COMPLETED  
13 SIMULATION RUNS COMPLETED  
14 SIMULATION RUNS COMPLETED  
15 SIMULATION RUNS COMPLETED  
16 SIMULATION RUNS COMPLETED  
17 SIMULATION RUNS COMPLETED  
18 SIMULATION RUNS COMPLETED

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

19 SIMULATION RUNS COMPLETED

20 SIMULATION RUNS COMPLETED

AFTER 10 ITERATIONS AND 21 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
7.5278E+04	-41.7504	0.9993	0.0212
8.5159E+04	-38.0133	0.9994	0.0212
8.0996E+04	-29.5476	0.9993	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.41433E+05

MEAN ABSOLUTE CORRELATION INDEX IS 36.43711

VARY( 1):KON11	=	1.3227E-03
VARY( 2):KON12	=	1.8327E-03
VARY( 3):KOFF11	=	2.6857E-05
VARY( 4):KOFF12	=	5.5239E-02
VARY( 5):R1	=	7.1514E-01
VARY( 6):R2	=	6.9469E-01
VARY( 7):R3	=	6.9369E-01
VARY( 8):KM1	=	1.9193E-01
VARY( 9):KM2	=	9.2889E-02
VARY( 10):KM3	=	7.0287E-02

21 SIMULATION RUNS COMPLETED

22 SIMULATION RUNS COMPLETED

23 SIMULATION RUNS COMPLETED

24 SIMULATION RUNS COMPLETED

25 SIMULATION RUNS COMPLETED

26 SIMULATION RUNS COMPLETED

27 SIMULATION RUNS COMPLETED

28 SIMULATION RUNS COMPLETED

29 SIMULATION RUNS COMPLETED

30 SIMULATION RUNS COMPLETED

AFTER 20 ITERATIONS AND 31 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
------	------	------	------

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
5.9939E+04 -40.8353 0.9993 0.0212
7.4825E+04 -37.4998 0.9994 0.0212
1.0020E+05 -32.9112 0.9993 0.0212
TOTAL RESIDUAL SUM OF SQUARES IS      2.34969E+05
MEAN ABSOLUTE CORRELATION INDEX IS 37.08212
VARY( 1):KON11                        = 1.4073E-03
VARY( 2):KON12                        = 1.8752E-03
VARY( 3):KOFF11                       = 3.3312E-05
VARY( 4):KOFF12                       = 5.1930E-02
VARY( 5):R1                           = 7.0965E-01
VARY( 6):R2                           = 7.0683E-01
VARY( 7):R3                           = 7.3782E-01
VARY( 8):KM1                          = 1.7062E-01
VARY( 9):KM2                          = 1.2653E-01
VARY(10):KM3                          = 6.2661E-02

31 SIMULATION RUNS COMPLETED
32 SIMULATION RUNS COMPLETED
33 SIMULATION RUNS COMPLETED
34 SIMULATION RUNS COMPLETED
35 SIMULATION RUNS COMPLETED
36 SIMULATION RUNS COMPLETED
37 SIMULATION RUNS COMPLETED
38 SIMULATION RUNS COMPLETED
39 SIMULATION RUNS COMPLETED
40 SIMULATION RUNS COMPLETED

AFTER 30 ITERATIONS AND 41 RUNS, INDIVIDUAL CURVE STATISTICS ARE
      RSQS      CORI      AUCO      AUCR
7.6263E+03 -7.2599 0.9981 0.0212
1.0401E+04 -2.1273 0.9988 0.0212
1.8801E+04  2.5696 0.9988 0.0212
TOTAL RESIDUAL SUM OF SQUARES IS      3.68288E+04
MEAN ABSOLUTE CORRELATION INDEX IS 3.98559
```

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

VARY( 1):KON11 = 1.6244E-03  
VARY( 2):KON12 = 1.9251E-03  
VARY( 3):KOFF11 = 4.8943E-05  
VARY( 4):KOFF12 = 5.0168E-02  
VARY( 5):R1 = 6.2121E-01  
VARY( 6):R2 = 5.9621E-01  
VARY( 7):R3 = 5.9459E-01  
VARY( 8):KM1 = 4.0732E-01  
VARY( 9):KM2 = 1.3648E-01  
VARY( 10):KM3 = 1.5097E-01

41 SIMULATION RUNS COMPLETED  
42 SIMULATION RUNS COMPLETED  
43 SIMULATION RUNS COMPLETED  
44 SIMULATION RUNS COMPLETED  
45 SIMULATION RUNS COMPLETED  
46 SIMULATION RUNS COMPLETED  
47 SIMULATION RUNS COMPLETED  
48 SIMULATION RUNS COMPLETED  
49 SIMULATION RUNS COMPLETED  
50 SIMULATION RUNS COMPLETED

AFTER 40 ITERATIONS AND 51 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
5.6269E+03	8.6478	0.9978	0.0212
5.2442E+03	12.2801	0.9980	0.0212
8.2999E+03	14.7010	0.9977	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 1.91711E+04

MEAN ABSOLUTE CORRELATION INDEX IS 11.87628

VARY( 1):KON11 = 2.1062E-03  
VARY( 2):KON12 = 1.9913E-03  
VARY( 3):KOFF11 = 5.9873E-05  
VARY( 4):KOFF12 = 4.7054E-02  
VARY( 5):R1 = 6.0728E-01

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

VARY( 6):R2 = 5.7917E-01  
VARY( 7):R3 = 5.6467E-01  
VARY( 8):KM1 = 4.0470E-01  
VARY( 9):KM2 = 1.3194E-01  
VARY( 10):KM3 = 1.4651E-01

51 SIMULATION RUNS COMPLETED  
52 SIMULATION RUNS COMPLETED  
53 SIMULATION RUNS COMPLETED  
54 SIMULATION RUNS COMPLETED  
55 SIMULATION RUNS COMPLETED  
56 SIMULATION RUNS COMPLETED  
57 SIMULATION RUNS COMPLETED  
58 SIMULATION RUNS COMPLETED  
59 SIMULATION RUNS COMPLETED  
60 SIMULATION RUNS COMPLETED

AFTER 50 ITERATIONS AND 61 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
1.9859E+03	-3.0714	0.9953	0.0212
1.0898E+03	-2.5375	0.9927	0.0212
1.7286E+03	3.5605	0.9896	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 4.80421E+03  
MEAN ABSOLUTE CORRELATION INDEX IS 3.05645

VARY( 1):KON11 = 2.4105E-03  
VARY( 2):KON12 = 2.1375E-03  
VARY( 3):KOFF11 = 1.3477E-04  
VARY( 4):KOFF12 = 4.1374E-02  
VARY( 5):R1 = 6.2669E-01  
VARY( 6):R2 = 5.9887E-01  
VARY( 7):R3 = 5.8367E-01  
VARY( 8):KM1 = 1.6932E-02  
VARY( 9):KM2 = 9.2812E-02  
VARY( 10):KM3 = 1.9663E-01

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

61 SIMULATION RUNS COMPLETED  
62 SIMULATION RUNS COMPLETED  
63 SIMULATION RUNS COMPLETED  
64 SIMULATION RUNS COMPLETED  
65 SIMULATION RUNS COMPLETED  
66 SIMULATION RUNS COMPLETED  
67 SIMULATION RUNS COMPLETED  
68 SIMULATION RUNS COMPLETED  
69 SIMULATION RUNS COMPLETED  
70 SIMULATION RUNS COMPLETED

AFTER 60 ITERATIONS AND 71 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
1.1890E+03	-1.2383	0.9922	0.0212
5.5294E+02	-0.8924	0.9857	0.0212
1.2190E+03	1.1332	0.9852	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.96094E+03  
MEAN ABSOLUTE CORRELATION INDEX IS 1.08796

VARY( 1):KON11 = 2.2618E-03  
VARY( 2):KON12 = 2.6717E-03  
VARY( 3):KOFF11 = 1.5724E-04  
VARY( 4):KOFF12 = 3.2238E-02  
VARY( 5):R1 = 6.3325E-01  
VARY( 6):R2 = 6.0587E-01  
VARY( 7):R3 = 5.9630E-01  
VARY( 8):KM1 = 3.1813E-03  
VARY( 9):KM2 = 3.6190E-01  
VARY( 10):KM3 = 6.8379E-01

71 SIMULATION RUNS COMPLETED  
72 SIMULATION RUNS COMPLETED  
73 SIMULATION RUNS COMPLETED

RESIDUAL SUM OF SQUARES IS 2712.0929160

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

COMPARED WITH EXPECTED RANGE 6408.1677 6859.6210  
FOR 6632 DEGREES OF FREEDOM  
PROBABLY OBSERVATIONAL ERRORS WERE OVERESTIMATED  
CONVERGENCE MAY BE INADEQUATE - RE-ENTERING WITH TIGHTER CRITERION

SIZE NEEDED FOR ARRAY WSFIT = 146374

OPTIMISATION RUN WITH 6642 OBSERVATIONS AND 10 VARIED PARAMETERS

AFTER 0 ITERATIONS AND 1 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
1.0502E+03	-0.8161	0.9912	0.0212
4.9983E+02	-2.4313	0.9841	0.0212
1.1620E+03	7.2167	0.9844	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.71209E+03  
MEAN ABSOLUTE CORRELATION INDEX IS 3.48803

VARY( 1):KON11 = 2.2423E-03  
VARY( 2):KON12 = 2.9138E-03  
VARY( 3):KOFF11 = 1.6320E-04  
VARY( 4):KOFF12 = 3.3646E-02  
VARY( 5):R1 = 6.3473E-01  
VARY( 6):R2 = 6.0786E-01  
VARY( 7):R3 = 5.9691E-01  
VARY( 8):KM1 = 2.7261E-03  
VARY( 9):KM2 = 7.0765E-01  
VARY( 10):KM3 = 1.6394E+00

1 SIMULATION RUNS COMPLETED  
2 SIMULATION RUNS COMPLETED  
3 SIMULATION RUNS COMPLETED  
4 SIMULATION RUNS COMPLETED  
5 SIMULATION RUNS COMPLETED  
6 SIMULATION RUNS COMPLETED  
7 SIMULATION RUNS COMPLETED

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

8 SIMULATION RUNS COMPLETED

9 SIMULATION RUNS COMPLETED

10 SIMULATION RUNS COMPLETED

AFTER 0 ITERATIONS AND 11 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCD	AUCR
1.0155E+03	0.1078	0.9909	0.0212
5.2077E+02	-2.6646	0.9846	0.0212
9.3077E+02	4.2165	0.9804	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.46702E+03

MEAN ABSOLUTE CORRELATION INDEX IS 2.32959

VARY( 1):KON11 = 2.2990E-03

VARY( 2):KON12 = 2.9876E-03

VARY( 3):KOFF11 = 1.6733E-04

VARY( 4):KOFF12 = 3.3646E-02

VARY( 5):R1 = 6.3473E-01

VARY( 6):R2 = 6.0786E-01

VARY( 7):R3 = 5.9691E-01

VARY( 8):KM1 = 2.7261E-03

VARY( 9):KM2 = 7.0765E-01

VARY( 10):KM3 = 1.6809E+00

11 SIMULATION RUNS COMPLETED

RESIDUAL SUM OF SQUARES IS 2467.0215749

COMPARED WITH EXPECTED RANGE 2620.5588 2805.1763

FOR 6632 DEGREES OF FREEDOM

PROBABLY OBSERVATIONAL ERRORS WERE OVERESTIMATED

CONVERGENCE MAY BE INADEQUATE - RE-ENTERING WITH TIGHTER CRITERION

SIZE NEEDED FOR ARRAY WSFIT = 146374

OPTIMISATION RUN WITH 6642 OBSERVATIONS AND 10 VARIED PARAMETERS

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

AFTER 0 ITERATIONS AND 1 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
1.0155E+03	0.1078	0.9909	0.0212
5.2077E+02	-2.6646	0.9846	0.0212
9.3077E+02	4.2165	0.9804	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.46702E+03  
MEAN ABSOLUTE CORRELATION INDEX IS 2.32959

VARY( 1):KON11 = 2.2990E-03  
VARY( 2):KON12 = 2.9876E-03  
VARY( 3):KOFF11 = 1.6733E-04  
VARY( 4):KOFF12 = 3.3646E-02  
VARY( 5):R1 = 6.3473E-01  
VARY( 6):R2 = 6.0786E-01  
VARY( 7):R3 = 5.9691E-01  
VARY( 8):KM1 = 2.7261E-03  
VARY( 9):KM2 = 7.0765E-01  
VARY( 10):KM3 = 1.6809E+00

- 1 SIMULATION RUNS COMPLETED
- 2 SIMULATION RUNS COMPLETED
- 3 SIMULATION RUNS COMPLETED
- 4 SIMULATION RUNS COMPLETED
- 5 SIMULATION RUNS COMPLETED
- 6 SIMULATION RUNS COMPLETED
- 7 SIMULATION RUNS COMPLETED
- 8 SIMULATION RUNS COMPLETED
- 9 SIMULATION RUNS COMPLETED
- 10 SIMULATION RUNS COMPLETED

AFTER 0 ITERATIONS AND 11 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
1.0027E+03	1.1289	0.9908	0.0212
5.8294E+02	-2.6158	0.9861	0.0212
7.7182E+02	0.6952	0.9763	0.0212

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

TOTAL RESIDUAL SUM OF SQUARES IS 2.35742E+03  
MEAN ABSOLUTE CORRELATION INDEX IS 1.47995  
VARY( 1):KON11 = 2.3572E-03  
VARY( 2):KON12 = 3.0632E-03  
VARY( 3):KOFF11 = 1.7157E-04  
VARY( 4):KOFF12 = 3.3646E-02  
VARY( 5):R1 = 6.3473E-01  
VARY( 6):R2 = 6.0786E-01  
VARY( 7):R3 = 5.9691E-01  
VARY( 8):KM1 = 2.7261E-03  
VARY( 9):KM2 = 7.0765E-01  
VARY( 10):KM3 = 1.7234E+00

11 SIMULATION RUNS COMPLETED

RESIDUAL SUM OF SQUARES IS 2357.4197076

COMPARED WITH EXPECTED RANGE 2383.7587 2551.6938

FOR 6632 DEGREES OF FREEDOM

PROBABLY OBSERVATIONAL ERRORS WERE OVERESTIMATED

ANALYSIS OF VARIANCE PROCEEDS ON ASSUMPTION OF GOOD FIT

BUT INSPECT RESIDUALS FOR EVIDENCE OF SYSTEMATIC ERROR

AFTER 0 ITERATIONS AND 1 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
1.0027E+03	1.1289	0.9908	0.0212
5.8294E+02	-2.6158	0.9861	0.0212
7.7182E+02	0.6952	0.9763	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.35742E+03

MEAN ABSOLUTE CORRELATION INDEX IS 1.47995

VARY( 1):KON11 = 2.3572E-03  
VARY( 2):KON12 = 3.0632E-03  
VARY( 3):KOFF11 = 1.7157E-04  
VARY( 4):KOFF12 = 3.3646E-02  
VARY( 5):R1 = 6.3473E-01

---

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

VARY( 6):R2 = 6.0786E-01  
VARY( 7):R3 = 5.9691E-01  
VARY( 8):KM1 = 2.7261E-03  
VARY( 9):KM2 = 7.0765E-01  
VARY( 10):KM3 = 1.7234E+00

1 SIMULATION RUNS COMPLETED  
2 SIMULATION RUNS COMPLETED  
3 SIMULATION RUNS COMPLETED  
4 SIMULATION RUNS COMPLETED  
5 SIMULATION RUNS COMPLETED  
6 SIMULATION RUNS COMPLETED  
7 SIMULATION RUNS COMPLETED  
8 SIMULATION RUNS COMPLETED  
9 SIMULATION RUNS COMPLETED  
10 SIMULATION RUNS COMPLETED  
11 SIMULATION RUNS COMPLETED

PARAMETER NUMBER,	TYPE,	NAME,	VALUE
1	PRI.	KON11	2.3572E-03
2	PRI.	KON12	3.0694E-03
3	PRI.	KOFF11	1.7191E-04
4	PRI.	KOFF12	3.3646E-02
5	PRI.	R1	6.3601E-01
6	PRI.	R2	6.0786E-01
7	PRI.	R3	5.9691E-01
8	PRI.	KM1	2.7261E-03
9	PRI.	KM2	7.0765E-01
10	PRI.	KM3	1.7269E+00

RESIDUAL SUM OF SQUARES = 2.3508E+03 FOR 6632 DEGREES OF FREEDOM

SINGULAR VALUES OF DEPENDENCE MATRIX AND STRUCTURE OF COMPONENTS

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

	VALUES	1527.690	1513.422	1499.613	218.799	127.810	66.218
		21.366	16.358	0.111	0.012		
KON11		-0.09972	0.09309	0.04894	-0.86557	0.39817	0.26184
		0.03532	0.03795	0.00014	0.00037		
KON12		0.00846	-0.00674	-0.00246	-0.14238	0.12559	-0.77239
		0.30077	0.52610	0.00110	0.00116		
KOFF11		0.09051	-0.05034	-0.00672	-0.44388	-0.88770	-0.06268
		-0.01477	0.00607	0.00004	0.00006		
KOFF12		-0.00783	0.00638	0.00266	0.16052	-0.12645	0.55806
		0.56997	0.56731	0.00069	0.00062		
R1		-0.83620	-0.54110	-0.06843	0.00421	-0.05501	-0.01527
		-0.00378	0.00022	0.00000	0.00000		
R2		-0.33901	0.62116	-0.70173	0.02377	-0.07445	-0.02619
		-0.00418	-0.00173	0.00002	-0.00001		
R3		-0.40935	0.55685	0.70742	0.07854	-0.11424	-0.05098
		-0.00804	-0.00469	-0.00002	-0.00001		
KM1		-0.00142	-0.00080	-0.00014	-0.03131	0.00745	-0.12639
		0.76362	-0.63235	-0.00116	-0.00192		
KM2		-0.00001	0.00001	-0.00001	-0.00029	0.00012	-0.00034
		-0.00046	0.00249	-0.90080	-0.43423		
KM3		-0.00001	0.00001	0.00002	-0.00025	0.00011	-0.00007
		-0.00061	0.00123	0.43424	-0.90080		

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

PERMUTE SEQUENCE OF VARIED PARAMETERS FROM

TO

DATA DO NOT DETERMINE THE FOLLOWING 2 VARIED PARAMETERS

SO DEFINE  $Q(I)=PARAM(I)/VALUE(I)$  AS INDEPENDENT NUMBERS (CURRENTLY ALL = 1.0)

I	PARAM(I)	VALUE(I)
9	KM2	7.0765E-01
10	KM3	1.7269E+00

WELL-DETERMINED PARAMETERS VARY AS FOLLOWING POWERS OF THE  $Q(I)$

I =            9            10

1	KON11	-0.0003	-0.0003
2	KON12	-0.0015	-0.0006
3	KOFF11	-0.0001	0.0000
4	KOFF12	-0.0009	-0.0003
5	R1	0.0000	0.0000
6	R2	0.0000	0.0000
7	R3	0.0000	0.0000



APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

AFTER 0 ITERATIONS AND 9 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
9.9312E+02	-2.8955	0.9907	0.0212
5.8462E+02	-1.8743	0.9861	0.0212
7.6709E+02	1.4812	0.9761	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.34482E+03  
MEAN ABSOLUTE CORRELATION INDEX IS 2.08366

VARY( 1):KON11 = 2.3572E-03  
VARY( 2):KON12 = 3.0755E-03  
VARY( 3):KOFF11 = 1.7225E-04  
VARY( 4):KOFF12 = 3.3646E-02  
VARY( 5):R1 = 6.3601E-01  
VARY( 6):R2 = 6.0786E-01  
VARY( 7):R3 = 5.9691E-01  
VARY( 8):KM1 = 2.7315E-03

9 SIMULATION RUNS COMPLETED

RESIDUAL SUM OF SQUARES IS 2344.8244108  
COMPARED WITH EXPECTED RANGE 6410.1336 6861.6550  
FOR 6634 DEGREES OF FREEDOM  
PROBABLY OBSERVATIONAL ERRORS WERE OVERESTIMATED  
CONVERGENCE MAY BE INADEQUATE - RE-ENTERING WITH TIGHTER CRITERION

SIZE NEEDED FOR ARRAY WSFIT = 119724

OPTIMISATION RUN WITH 6642 OBSERVATIONS AND 8 VARIED PARAMETERS

AFTER 0 ITERATIONS AND 1 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
9.9312E+02	-2.8955	0.9907	0.0212
5.8462E+02	-1.8743	0.9861	0.0212
7.6709E+02	1.4812	0.9761	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.34482E+03

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

MEAN ABSOLUTE CORRELATION INDEX IS 2.08366  
VARY( 1):KON11 = 2.3572E-03  
VARY( 2):KON12 = 3.0755E-03  
VARY( 3):KOFF11 = 1.7225E-04  
VARY( 4):KOFF12 = 3.3646E-02  
VARY( 5):R1 = 6.3601E-01  
VARY( 6):R2 = 6.0786E-01  
VARY( 7):R3 = 5.9691E-01  
VARY( 8):KM1 = 2.7261E-03

1 SIMULATION RUNS COMPLETED  
2 SIMULATION RUNS COMPLETED  
3 SIMULATION RUNS COMPLETED  
4 SIMULATION RUNS COMPLETED  
5 SIMULATION RUNS COMPLETED  
6 SIMULATION RUNS COMPLETED  
7 SIMULATION RUNS COMPLETED  
8 SIMULATION RUNS COMPLETED

AFTER 0 ITERATIONS AND 9 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
9.8863E+02	-2.6645	0.9906	0.0212
5.8570E+02	-1.5034	0.9862	0.0212
7.6503E+02	1.8765	0.9760	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.33936E+03

MEAN ABSOLUTE CORRELATION INDEX IS 2.01479

VARY( 1):KON11 = 2.3572E-03  
VARY( 2):KON12 = 3.0817E-03  
VARY( 3):KOFF11 = 1.7260E-04  
VARY( 4):KOFF12 = 3.3646E-02  
VARY( 5):R1 = 6.3601E-01  
VARY( 6):R2 = 6.0786E-01  
VARY( 7):R3 = 5.9691E-01  
VARY( 8):KM1 = 2.7315E-03

9 SIMULATION RUNS COMPLETED

RESIDUAL SUM OF SQUARES IS 2339.3569848  
COMPARED WITH EXPECTED RANGE 2265.6976 2425.2904  
FOR 6634 DEGREES OF FREEDOM  
ANALYSIS OF VARIANCE PROCEEDS ON ASSUMPTION OF GOOD FIT  
BUT INSPECT RESIDUALS FOR EVIDENCE OF SYSTEMATIC ERROR

AFTER 0 ITERATIONS AND 1 RUNS, INDIVIDUAL CURVE STATISTICS ARE

RSQS	CORI	AUCO	AUCR
9.8863E+02	-2.6645	0.9906	0.0212
5.8570E+02	-1.5034	0.9862	0.0212
7.6503E+02	1.8765	0.9760	0.0212

TOTAL RESIDUAL SUM OF SQUARES IS 2.33936E+03  
MEAN ABSOLUTE CORRELATION INDEX IS 2.01479

VARY( 1):KON11	= 2.3572E-03
VARY( 2):KON12	= 3.0817E-03
VARY( 3):KOFF11	= 1.7260E-04
VARY( 4):KOFF12	= 3.3646E-02
VARY( 5):R1	= 6.3601E-01
VARY( 6):R2	= 6.0786E-01
VARY( 7):R3	= 5.9691E-01
VARY( 8):KM1	= 2.7261E-03

1 SIMULATION RUNS COMPLETED  
2 SIMULATION RUNS COMPLETED  
3 SIMULATION RUNS COMPLETED  
4 SIMULATION RUNS COMPLETED  
5 SIMULATION RUNS COMPLETED  
6 SIMULATION RUNS COMPLETED  
7 SIMULATION RUNS COMPLETED  
8 SIMULATION RUNS COMPLETED  
9 SIMULATION RUNS COMPLETED

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

PARAMETER NUMBER,	TYPE,	NAME,	VALUE
1	PRI.	KON11	2.3572E-03
2	PRI.	KON12	3.0878E-03
3	PRI.	KOFF11	1.7294E-04
4	PRI.	KOFF12	3.3646E-02
5	PRI.	R1	6.3601E-01
6	PRI.	R2	6.0786E-01
7	PRI.	R3	5.9691E-01
8	PRI.	KM1	2.7261E-03

RESIDUAL SUM OF SQUARES = 2.3344E+03 FOR 6634 DEGREES OF FREEDOM

SINGULAR VALUES OF DEPENDENCE MATRIX AND STRUCTURE OF COMPONENTS

	VALUES	1529.626	1513.580	1498.975	218.692	128.386	66.408
		21.688	16.207				
KON11		-0.09115	0.10178	0.04932	-0.86243	0.40355	0.26369
		-0.03650	-0.03745				
KON12		0.00789	-0.00752	-0.00251	-0.14338	0.12636	-0.77260
		-0.30302	-0.52404				
KOFF11		0.08620	-0.05834	-0.00722	-0.44956	-0.88487	-0.06130
		0.01472	-0.00632				
KOFF12		-0.00729	0.00711	0.00270	0.16153	-0.12729	0.55787
		-0.56761	-0.56938				
R1		-0.88039	-0.46700	-0.05918	0.00383	-0.05531	-0.01535
		0.00378	-0.00020				

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```

R2          -0.28973  0.64513 -0.70214  0.02322 -0.07494 -0.02637
            0.00417  0.00178

R3          -0.35372  0.59317  0.70782  0.07774 -0.11518 -0.05132
            0.00817  0.00467

KM1         -0.00142 -0.00073 -0.00007 -0.02960  0.00621 -0.12253
            -0.76443  0.63224
    
```

FITTED VALUES, ACCURACIES AND CONFIDENCE LIMITS

NO.	NAME	VALUE	SDLN	5 PERCENT	95 PERCENT
1	KON11	2.3572E-03	0.0042	2.3411E-03	2.3734E-03
2	KON12	3.0878E-03	0.0220	2.9780E-03	3.2017E-03
3	KOFF11	1.7294E-04	0.0043	1.7172E-04	1.7418E-04
4	KOFF12	3.3646E-02	0.0265	3.2213E-02	3.5144E-02
5	R1	6.3601E-01	0.0005	6.3549E-01	6.3652E-01
6	R2	6.0786E-01	0.0006	6.0727E-01	6.0845E-01
7	R3	5.9691E-01	0.0009	5.9605E-01	5.9778E-01
8	KM1	2.7261E-03	0.0312	2.5897E-03	2.8697E-03

CORRELATION MATRIX COMPONENTS

COLUMN	1	2	3	4	5	6	7	8
ROW 1	1.000	0.221	-0.341	0.487	-0.442	-0.625	-0.822	-0.101
ROW 2	0.221	1.000	0.031	0.847	0.007	-0.061	-0.123	-0.383
ROW 3	-0.341	0.031	1.000	-0.020	0.533	0.582	0.583	-0.098
ROW 4	0.487	0.847	-0.020	1.000	-0.151	-0.260	-0.384	-0.198

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```
ROW 5 -0.442 0.007 0.533 -0.151 1.000 0.448 0.508 -0.142

ROW 6 -0.625 -0.061 0.582 -0.260 0.448 1.000 0.652 -0.034

ROW 7 -0.822 -0.123 0.583 -0.384 0.508 0.652 1.000 -0.009

ROW 8 -0.101 -0.383 -0.098 -0.198 -0.142 -0.034 -0.009 1.000
```

!!!! COMMAND INTERPRETER RE-ENTERED

SETNOFIT;

WHENEVER TIME = TOBS % CALL PRINT;

\*\*;

COMPILE INSTANT;

LOCN INSTRUCTION

642:\* calculate RSS;

642:RSS = 0 ;

648:DO 1 FOR #1 = 0 (1) 7 ;

656: DO 2 FOR #2 = 0 (1) 2215 ;

664: RSS = RSS + RESIDARRAY<#1,#2>\*RESIDARRAY<#1,#2> ;

677: LABEL 2 ;

707:LABEL 1 ;

712:\* Print values of parameters;

712:REWIND 7;

714:WRITE 1=7, (E14,6) RSS " RSS" % ;

735:\* binding parameters;

735:WRITE 1, (E14,6) Kon11 " Kon11 in 1/concentration unit seconds" % ;

754:WRITE 1, (E14,6) Kon12 " Kon12 in 1/concentration unit seconds" % ;

773:WRITE 1, (E14,6) Kon11 " Kon11 in 1/concentration unit seconds" % ;

792:WRITE 1, (E14,6) Kon12 " Kon12 in 1/concentration unit seconds" % ;

APPENDIX A. SAMPLES OF MAPLE AND FACSIMILE CODE

---

```

811:* ;
811:WRITE 1, (E14,6) Koff11 " Koff11 in 1/seconds" % ;
830:WRITE 1, (E14,6) Koff12 " Koff12 in 1/seconds" % ;
849:WRITE 1, (E14,6) Koff11 " Koff11 in 1/seconds" % ;
868:WRITE 1, (E14,6) Koff12 " Koff12 in 1/seconds" % ;
887:* machine parameters;
887:WRITE 1, (E14,6) R1 " R1 in ng/mm^2" % ;
906:WRITE 1, (E14,6) R2 " R2 in ng/mm^2" % ;
925:WRITE 1, (E14,6) R3 " R3 in ng/mm^2" % ;
944:WRITE 1, (E14,6) R1 " R1 in ng/mm^2" % ;
963:WRITE 1, (E14,6) R2 " R2 in ng/mm^2" % ;
982:WRITE 1, (E14,6) R3 " R3 in ng/mm^2" % ;
1001:WRITE 1, (E14,6) kM1 " kM1 in ng/concentration unit mm^2 seconds" % ;
1020:WRITE 1, (E14,6) kM2 " kM2 in ng/concentration unit mm^2 seconds" % ;
1039:WRITE 1, (E14,6) kM3 " kM3 in ng/concentration unit mm^2 seconds" % ;
1058:WRITE 1, (E14,6) kM1 " kM1 in ng/concentration unit mm^2 seconds" % ;
1077:WRITE 1, (E14,6) kM2 " kM2 in ng/concentration unit mm^2 seconds" % ;
1096:WRITE 1, (E14,6) kM3 " kM3 in ng/concentration unit mm^2 seconds" % ;
1115:WRITE 1, (E14,6) cT1 " cT1 is a dimensionless quantity" % ;
1134:*WRITE 1, (E14,6) cT2 " cT2 is a dimensionless quantity" % ;
1134:
1134:**;

```

```

NO. OF VARIABLES = 12
NO. IMPLICIT      = 0
NO. OF ERRORS    = 0

```

SPACE FOR	IN USE	FREE	TOTAL
INTEGER SCALARS AND ARRAYS	305	499695	500000
REAL SCALARS AND ARRAYS	173114	326886	500000
INDEXES AND SECRETS	102	397	499
CODE ARRAY	638	39362	40000
NAME DICTIONARY	129	1871	2000



# Appendix B

## Publications

### B.1 Paper in Human Immunology

## ARTICLE IN PRESS

Human Immunology xxx (xxxx) xxx–xxx



ELSEVIER

Contents lists available at ScienceDirect

Human Immunology

journal homepage: [www.elsevier.com/locate/humimm](http://www.elsevier.com/locate/humimm)

## Direct quantitative measurement of the kinetics of HLA-specific antibody interactions with isolated HLA proteins

Sunil Daga<sup>a,b,i,s</sup>, Harry Moyse<sup>c</sup>, David Briggs<sup>d</sup>, David Lowe<sup>e</sup>, Neil Evans<sup>c</sup>, James Jones<sup>e</sup>, Rico Buchli<sup>g</sup>, Curtis McMurtrey<sup>f</sup>, Arend Mulder<sup>h</sup>, William Hildebrand<sup>f</sup>, Frans Claas<sup>h</sup>, Robert Higgins<sup>a,i</sup>, Daniel A. Mitchell<sup>f,1</sup>, Daniel Zehnder<sup>a,i,1</sup>

<sup>a</sup> Renal Department, UHCW NHS Trust, Coventry, UK<sup>b</sup> Renal Department, St. James's University Hospital, Leeds, UK<sup>c</sup> School of Engineering, University of Warwick, Coventry, UK<sup>d</sup> H & I department, NHSBT, Birmingham, UK<sup>e</sup> H & I department, Royal Liverpool University Hospital, Liverpool, UK<sup>f</sup> University of Oklahoma Health Science Center, Oklahoma City, OK, United States<sup>g</sup> Pure Protein LLC, Oklahoma City, OK, United States<sup>h</sup> Immunohaematology and Blood Transfusion, Leiden University Medical Center, 2333 ZA Leiden, The Netherlands<sup>1</sup> Warwick Medical School, University of Warwick, Coventry, UK

## ARTICLE INFO

## Keywords:

HLA-specific antibodies

Affinity

Surface plasmon resonance

Binding kinetics

Epitope

Eplet

## ABSTRACT

HLA specific antibodies vary in their pathogenicity and this is likely to be the net effect of constant chain usage, quantity, specificity, and affinity. Here we have measured the affinity of human monoclonal antibodies for a range of HLA proteins. Purified antibodies and ligands allowed dynamic interactions to be measured directly by surface plasmon resonance. Physicochemical differences between pairs of ligands were quantified using electrostatic mismatch and hydrophobic mismatch scores.

All antibodies were characterized by fast on-rates and slow off rates but with a wide range of association rates ( $k_{on}$ ,  $3.63\text{--}24.25 \times 10^5$  per mol per second) and dissociation rates ( $k_{off}$ ,  $0.99\text{--}10.93 \times 10^{-3}$  per second). Dissociation constants ( $K_D$ ) ranged from  $5.9 \times 10^{-10}$  M to  $3.0 \times 10^{-8}$  M. SN320G6 has approximately a twenty-fold greater affinity for HLA A2 compared with SN607D8, but has a similar affinity for HLA-A2 and B57. In contrast, SN607D8 has greater than a twofold greater affinity for HLA-A2 compared with A68. Similarly, WKID12 has about a threefold greater affinity for HLA-B27 compared with B7. The higher affinity interactions correlate with the specificity of stimulating antigen. This is the first study to directly measure the binding kinetics and affinity constants for human alloantibodies against HLA.

## 1. Introduction

HLA specific antibodies are a relative barrier to solid organ transplantation. Higher levels, which typically precipitate a cytotoxic crossmatch, can cause hyperacute rejection [1]. Lower levels may be tolerated or be associated with acute rejection and/or poor outcome [2–6]. Post-transplantation, the appearance of donor HLA-specific antibodies (DSA) is often, but not always, associated with acute or chronic rejection and transplant glomerulopathy [3,7–10]. The ability of an HLA-specific antibody or antibody mixture (e.g. serum) to activate the classical pathway of the complement system appears to be related antibody pathogenicity, although this functionality is not necessarily

proportional to measured antibody level [11–14]. Antigen specificity (or the degree of expression of the target antigen) is a further important criterion related to pathogenicity as, for example, transplantation can be successful in the presence of cytotoxic positive crossmatches due to HLA Class II antibodies [15].

The specificity of HLA antibodies is defined by reaction patterns against different alleles sharing specific amino acid residues [16]. These are generally considered to be the nominal epitopes allowing binding of the antibody and tend to be reduced to the minimal number of required residues, sometimes to a single amino acid. It has been suggested that the critical interactions for binding involve those residues within a 3 Å diameter patch, termed an eplet [17,18]. However, the exact

\* Corresponding author at: Renal Department, St. James's University Hospital, LTH NHS Trust, Leeds, UK.

E-mail address: [sunildaga@nhs.net](mailto:sunildaga@nhs.net) (S. Daga).

<sup>1</sup> Equal contributions.

<http://dx.doi.org/10.1016/j.humimm.2017.10.012>

Received 13 June 2017; Received in revised form 11 October 2017; Accepted 26 October 2017

0198-8859/ © 2017 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.

Please cite this article as: Daga, S., Human Immunology (2017), <http://dx.doi.org/10.1016/j.humimm.2017.10.012>

## ARTICLE IN PRESS

S. Daga et al.

Human Immunology xxx (xxxx) xxx-xxxx

conformation and orientation of the interaction has yet to be shown for any HLA protein-anti-HLA antibody complex [19]. The footprint of the antibody complementarity determining regions (CDRs) can be up to 900 Å<sup>2</sup> in surface area [20,21], and therefore a larger number of amino acid residues within the HLA protein, perhaps up to 25, will likely be involved in determining the nature of the interaction, imparting both specificity and affinity. Importantly, the reactivity of HLA proteins with antibody can vary even when sharing the critical epitope/epit. It is speculated that the difference in reactivity is due to a difference in release of binding energy which is proportional to the total surface area of interaction [22,23]. The immunogenicity of the HLA antigen, defined as its ability to evoke/stimulate a *de novo* antibody response can also vary and is believed to be related to electrostatic and hydrophobic properties of amino acid residues within polymorphic regions of the HLA protein [24–26].

Current solid phase single antigen bead assays measure mean fluorescence intensity (MFI) and this readout is often used to estimate amounts of antibodies, although this has not been validated. High value MFIs will have components relating to both higher absolute antibody concentrations and to stronger antibody binding (affinity). Resolving the distinction between concentration and affinity is important because affinity measurements are likely to provide a greater insight into antibody function. During the immune response, the affinity of antibodies typically increases by affinity maturation. This is generated via somatic hypermutation of immunoglobulin gene segments, leading to clonally selected variations in amino acid sequences that favor increased binding to target antigens [27]. Only one study has previously determined the affinity of HLA antibodies, and that was of mouse monoclonal antibodies against human HLA-A2 protein expressed on human B-cell lines. That study shows marked differences in affinity between F<sub>ab</sub> fragments from two separate monoclonal antibodies (determined to be in the micromolar and nano-molar range, respectively), which were due to difference in the dissociation rates [28]. The authors used a saturation binding technique with radio-labeled antibodies. Our study is the first to directly measure the affinities of human allo-antibodies against HLA.

Through the use of real time biosensor techniques (by surface plasmon resonance), we have been able to measure binding kinetics and the affinity constants of antibodies against different HLA proteins that share epitopes/epitopes but differ in adjacent residues. This approach was used to determine quantitatively the importance of epitope configurations in relation to affinity. Here, we used this technique to measure the affinity of clinically relevant anti-HLA antibodies.

## 2. Material and methods

### 2.1. Biotinylated soluble recombinant HLA proteins

Biotinylated soluble recombinant HLA proteins (sHLA) were provided by Pure Protein LLC (Oklahoma City, OK, USA) [29]. Products of the following genes were studied; HLA-A\*01:01, A\*02:01, A\*68:01, B\*57:01, B\*07:02, and B\*27:05. We will refer to these proteins as HLA-A1, A2, A68, B57, B7, and B27, respectively, for simplicity. The concentration of the purified biotinylated molecules was determined using the Micro BCA protein assay kit (Thermo Fisher Scientific, Rockford, IL) using bovine gamma globulin (Sigma; Poole, UK) as a protein standard.

### 2.2. Human monoclonal HLA-specific antibodies

Human monoclonal HLA-specific antibodies (mAbs) were secreted by human hybridomas produced by Epstein-Barr virus (EBV) transformation of B-lymphocytes from HLA antibody positive multi-parous women (sensitized during pregnancy), followed by fusion and sub-cloning of antibody producing EBV lines (Table 1). SN607D8 and SN203G6 were derived from the same multiparous donor (HLA type A\*24 A\*29, B\*07, B\*44, C\*07, C\*16) sensitized during pregnancy

**Table 1**

Characteristics of IgG human monoclonal HLA-specific antibodies. Epitope designations use the nomenclature from the HLA epitope registry ([www.epregistry.br](http://www.epregistry.br)) and are based on reaction pattern rather than a known binding specificity. \*Corresponding to the alleles from which the proteins were derived.

mAb	Immunizing HLA	Bead assay-defined HLA specificities*	Epitope designation	Isotype
SN607D8	A2	A*02:01/03/06, A*68:01/02, A*69:01	142TKH	IgG1, κ
SN203G6	A2 and/or B57	A*02:01/03/06, B*57:01/03, B*58:01	62GE	IgG1, λ
WK1D12	B27	B*27:05/08, B*07:02, B*13:02, B*40:01/02/ 06, B*81:01	163EW	IgG1, κ

(child's HLA type A\*02, A\*29, B\*44, B\*57). WK1D12 was derived from a female donor (HLA type A\*01, B\*08, C\*07) sensitized during pregnancy (child's HLA type A\*01, A\*11, B\*08, B\*27, C\*01, C\*07).

Antibody-containing supernatants were dialyzed (8000 MWCOSpectra/Por® Dialysis) against PBS overnight. The supernatants were further processed by affinity chromatography using HLA protein immobilized on Sepharose beads [29] and ion-exchange chromatography using Q-Sepharose. HLA-A2 (from A\*02:01) and B7 (from B\*07:02) were separately coupled to cyanogen bromide activated Sepharose and the resulting columns were used to enrich mAbs SN203G6 and SN607D8, and WK1D12 respectively. Overall this approach yielded highly purified monoclonal IgG and avoided contamination with residual bovine IgG and other serum components from the hybridoma culture medium. The homogeneity of the purified IgG was confirmed via SDS-PAGE using 4–12% gradient polyacrylamide gels. The concentration of purified antibodies was determined using Lowry's assay [30]. CDC defined antibody specificities were confirmed in a single antigen bead assay (One Lambda, Canoga Park, CA) as per manufacturer's instructions. Dose response curves from antibody doubling dilution series were used to calculate effective concentrations that gave 50% of maximal MFI signal (EC-50). Hill-type EC-50 values were calculated using curve fitting performed in MATLAB.

### 2.3. Surface plasmon resonance

Sensorgrams were obtained using the Bio-Rad Proteon XPR36 biosensor platform [13,14]. 2.5 µg/ml of biotinylated HLA protein was immobilized on neutravidin-coated sensor chips (Protein™ NLC Sensor Chip #1765021) with a flow rate of 25 µl/min over 300 s. Purified monoclonal HLA-specific antibodies were flowed over the chip at a rate of 25 µl/min (slowest rate for the equipment to allow maximum binding) for 960 s to obtain equilibrium phase (where the sensorgram of binding phase plateaus) and at 37 °C in the Association phase to represent physiological binding temperatures. Following this, the running buffer (PBS containing 0.05% Tween 20) was passed over for 960 s at a flow rate of 25 µl/min for the dissociation phase. Interactions of all three HLA-specific antibodies were studied in duplicate over a range of six different concentrations (3.125–100 nM) against the same surface concentration of immobilized HLA protein.

The data obtained from the sensorgrams were modelled using an implementation of differential evolution [31] on MATLAB software (Mathworks®) with the simulation tool FACSIMILE (MCPA Software) [32]. Together these allow for curve fitting and simultaneous determination of kinetic Association ( $k_{on}$ ) rates, dissociation ( $k_{off}$ ) rates and the Dissociation Constant  $K_D$  (as the ratio of the dissociation,  $k_{off}$ , and Association,  $k_{on}$ , rate constants).

### 2.4. Molecular differences between the allele pairs

The PyMOL Molecular Graphics System, Version 1.1 Schrödinger, LLC was used to model the crystallographic structures of HLA-A2 (PDB

ID: 3MRK), A68 (PDB ID: 4HWZ), B27 (PDB ID: 10GT identifier), B57 (PDB ID: 2YPK), and B7 (PDB ID: 3VCL). These structural diagrams were used to show molecular differences between each allele pair shown. We considered the area within a 15 Å radius around the designated epitope on the HLA molecule examined (as compared to the whole protein), since this corresponds to the maximum radius of the antibody recognition site comprising the cognate antibody paratope [18]. Within this putative contact area on the HLA molecule the number of amino acid differences, electrostatic mismatch score (EMS), and hydrophobic mismatch score (HMS) [26,24], were estimated to quantify physicochemical differences between pairs of HLA proteins. Two sets of comparisons were made, including and excluding differences located within the peptide binding groove.

### 3. Results

#### 3.1. Determination of binding kinetics for human monoclonal HLA-specific antibody-HLA protein interactions

The sensorgrams illustrating the interactions between human HLA-mAbs and immobilized HLA-proteins are distinct, specific and reproducible (Fig. 1). Differences in signal (defined by response units, RU) and kinetics ( $k_{on}$  and  $k_{off}$  values) were observed for different antigen-antibody interactions (data given in Table 2). Consistent with the known specificities of these antibodies (Table 1) SN607D8 showed binding to HLA-A2 and A68 and not to HLA-A1; SN230G6 bound to HLA-A2 and B57 and not to HLA A1; WK1D12 bound to HLA-B7 and B27 and not to HLA-A1.

Modelling of the sensorgram data was optimised using MATLAB together with the simulation tool FACSIMILE, with the kinetic parameters determined across both phases using the full dynamic data. The quality of the fit of the model to the data to provide the kinetic constants was assessed visually (Fig. 2) and quantitatively within MATLAB/FACSIMILE via the residual sum of squares. The fitting process yielded fits with low residual sum of squares in which typically the

estimated SPR signal remained within the sensorgram signal throughout the experiment (Fig. 2). The binding kinetics modelled varied greatly between the three antibodies against the HLA proteins which they are known to recognize (Table 2). All are characterized by fast on-rates and slow off rates but there is a wide range in the magnitude of these parameters for the reactions tested. The Association rates ( $k_{on}$ ) varied between 3.63 and  $24.25 \times 10^5$  per mol per second and the dissociation rates ( $k_{off}$ ) varied between 0.99 and  $10.93 \times 10^{-3}$  per second. The dissociation constant ( $K_D$ ) values obtained were in the range of 0.1–10 nM (Table 2).

SN230G6 has an affinity (inverse of  $K_D$ ) for both HLA-A2 and B57 around an order of magnitude greater than that of SN607D8 for HLA-A2 or A68; against just HLA-A2, the former exhibits a twenty-fold greater affinity compared with the latter. Fig. 1 shows that SN230G6 has a faster on-rate, and reaches a higher equilibrium level than SN607D8. The greater affinity of SN230G6 for HLA-A2 is also seen with a dose-response experiment via bead assay: SN607D8 requires about a ten-fold greater concentration to reach half-maximal binding (EC-50) compared with SN230G6 (Table 2). EC-50 analysis was not able to distinguish any difference in the binding of these two antibodies to HLA-A2 compared to either A68 or B57 (SN607D8 and SN230G6, respectively). In contrast, the biosensor assay showed that SN607D8 has approximately 2.5 times the affinity for HLA-A2 compared with A68. SN230G6, on the other hand, has similar affinities for HLA-A2 and B57 (1.33-fold difference). We also found that WK1D12 has about a threefold greater affinity for HLA-B27 compared to B7. Thus, in the two cases where the stimulating HLA for the antibody is known, the affinity is greater for that allele compared to a different allele but one which shares the critical binding motif.

#### 3.2. Influence of molecular difference on binding affinities

Fig. 3 shows the structural differences between the HLA pairs within a radius of 15 Å about the notional epitope or eplet that associates with the reactivity of each antibody. This corresponds to the binding

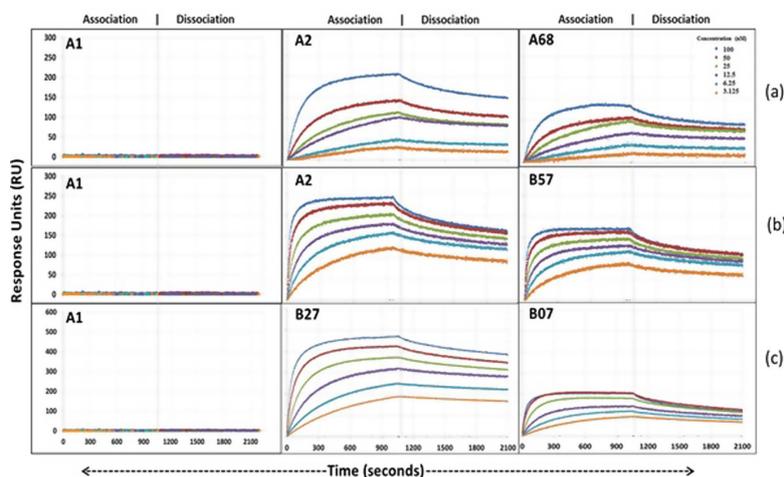


Fig. 1. Binding kinetics of three of human mAbs to HLA. a. SN607D8; b. SN230G6; c. WK1D12. The sensorgrams show antibody binding at six concentrations as shown in the top right panel. Note the scale of the y axis is the same for a and b (max 300 RU) and different for c (max 600RU).

ARTICLE IN PRESS

S. Daga et al.

Human Immunology xxx (xxxx) xxx-xxxx

**Table 2**  
Difference in kinetic rates and dissociation constants for mAbs interaction with different alleles. Each shown  $k_{on}$ ,  $k_{off}$ , and  $K_D$  value is from two separate binding experiments. The original sensitizing antigens are shown in bold. ND = not done. EC-50 values were calculated using dose response curve fitting performed in MATLAB.

mAb	Target HLA	Bead assay		Sensorgram analysis				
		EC-50	EC-50 ratio	$k_{on}$ ( $M^{-1} s^{-1}$ )	$k_{off}$ ( $s^{-1}$ )	$K_D$ (M)	Average $K_D$ (M)	Ratio of $K_D$
SN607D8	A2	2		$4.6 \times 10^5$	$5.6 \times 10^{-3}$	$1.22 \times 10^{-8}$	$1.2 \times 10^{-8}$	
	A68	1.6	0.8	$4.05 \times 10^5$	$4.9 \times 10^{-3}$	$1.21 \times 10^{-8}$	$3.0 \times 10^{-8}$	2.47
				$3.23 \times 10^5$	$9.45 \times 10^{-3}$	$2.93 \times 10^{-8}$		
$4.05 \times 10^5$	$12.4 \times 10^{-3}$	$3.07 \times 10^{-8}$						
SN230G6	A2	0.18		$23.2 \times 10^5$	$1.31 \times 10^{-3}$	$5.63 \times 10^{-10}$	$5.9 \times 10^{-10}$	
	B57	0.15	0.83	$19.6 \times 10^5$	$1.20 \times 10^{-3}$	$6.13 \times 10^{-10}$	$7.8 \times 10^{-10}$	1.33
				$28.7 \times 10^5$	$1.88 \times 10^{-3}$	$6.56 \times 10^{-10}$		
$19.8 \times 10^5$	$1.78 \times 10^{-3}$	$9.02 \times 10^{-10}$						
WK1D12	B27	ND		$9.8 \times 10^5$	$1.02 \times 10^{-3}$	$1.04 \times 10^{-9}$	$1.04 \times 10^{-9}$	
	B7	ND	ND	$9.1 \times 10^5$	$0.95 \times 10^{-3}$	$1.04 \times 10^{-9}$	$3.12 \times 10^{-9}$	3.01
				$14.2 \times 10^5$	$3.45 \times 10^{-3}$	$2.43 \times 10^{-9}$		
$9.8 \times 10^5$	$3.72 \times 10^{-3}$	$3.81 \times 10^{-9}$						

footprint as described by Duquesnoy et al. [18]. Within this area ( $900 \text{ \AA}^2$ ) we counted the number of amino acid residue differences, and the EMS and HMS for each pair (data given in Table 3). All the designated epitopes/epitopes lie adjacent to the peptide-binding groove and it is not certain how, or indeed if substitutions within the groove could contribute to the binding of these alloantibodies. Therefore, we have calculated the differences between the HLA pairs with and without the internal binding groove residues. These results are given in Table 3. It is immediately obvious that none of these physical differences correlate with the affinity differences between any of the HLA pairs and their respective antibodies. The largest number of amino acid differences of the antigen pairs examined is seen between HLA-A2, and B57, this is unsurprising as antigens encoded by different HLA gene loci would be reasonably expected to show increased variation. However, the difference in SN230G6 binding to these individual antigens is lowest. SN607D8 and WK1D12 exhibit similar differences in affinity for their tested ligands but there is a twofold difference in the  $\Delta$ HMS and  $\Delta$ EMS values.

4. Discussion

This paper describes the first use of a real time, label free biosensor technique to characterize biophysical properties of human, monoclonal, HLA-specific antibodies. We were able to show distinct  $K_D$  values for different HLA antibody/HLA allele combinations, in the order of 0.1–10.0 nM. We have shown that the signal obtained displays the expected HLA specificity and we can discern a distinct signal for low nanomolar concentrations of these human monoclonal HLA-specific

antibodies. The dissociation constants obtained for mAbs in this study are similar to previously reported values for target ligands [33,34,28]. SPR has been used to detect and study the binding interactions of dsDNA-specific antibodies in patients with SLE [35] while a recent study of ninety-nine patients showed the avidity of dsDNA antibodies correlates with disease activity of lupus nephritis [36]. It is therefore reasonable to assume that the characteristics of the HLA antibodies we describe here relate to their pathogenicity and are therefore likely to be of clinical relevance. Further to this, a given antibody may differ in its pathogenicity depending on its target allele because its affinity will vary accordingly. Duquesnoy et al. [22] have shown that the same monoclonal antibody can be associated with different degrees of complement-dependent cytotoxicity depending on the HLA allele expressed on the target cell. Those authors speculated that this involved target affinity differences and here we prove that such affinity differences do exist.

The two human antibodies (SN230G6 and SN607D8) which bind to HLA-A2 were derived from the same multiparous woman. SN607D8 is clearly the consequence of HLA-A2 immunization whereas SN230G6 could be the result of HLA-A2 or B57 immunization, or possibly by both as they were co-expressed by the sensitizing individual. The affinity of SN230G6 for HLA-A2 (epitope 62GE) is 20-fold greater than of SN607D8 for HLA-A2 (epitope 142TKH). Kushiata et al. [37], measured the CDC activity of these mAbs and showed a ten times higher CDC reactivity by SN230G6 compared to SN607D8 against the same target cell. We showed a similar difference of their EC-50 values in a single antigen bead assay (Table 2). This EC-50 analysis also

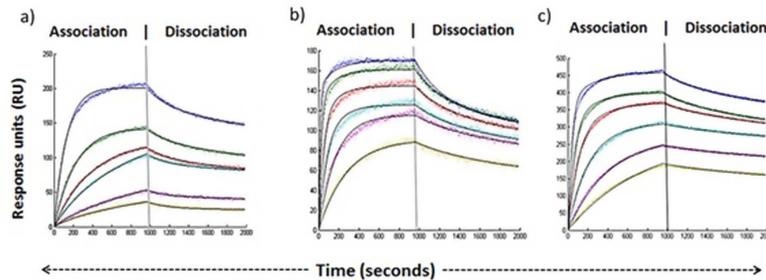


Fig. 2. Data fitting examples (solid black lines) shown (SN607D8 binding to HLA-A2 (a) SN230G6 to HLA-B57 (b) and WK1D12 to HLA-B27 (c)).

S. Daga et al.

Human Immunology xxx (xxxx) xxx-xxxx

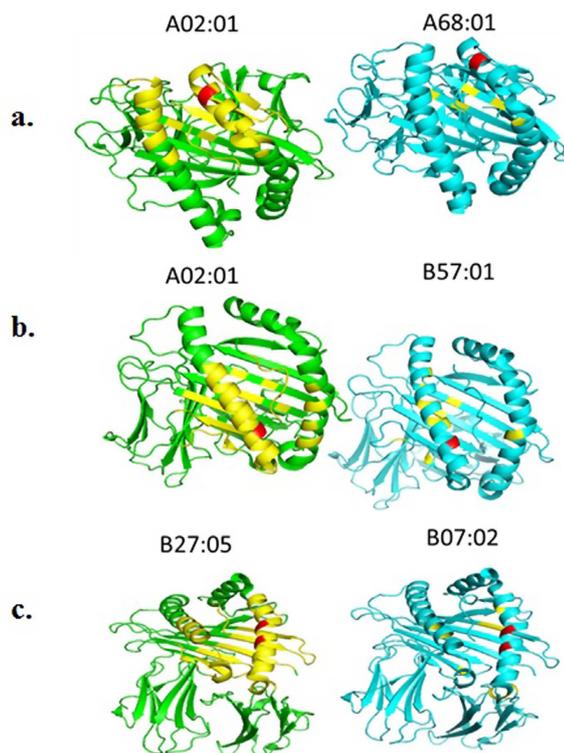


Fig. 3. Structural differences between HLA pairs, each recognized by the same antibody (a. SN607D8; b. SN230G6; c. WK1D12). Red positions indicate the amino acid residues that associate with the specificity of the antibody and define the notional epitope. On the left-hand ribbon diagrams, yellow shows the area within a radius of 15 Å from each such epitope and suggest a binding footprint of the antibody. Yellow residues on the right marks the amino acid differences within this footprint for each HLA pair. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3  
Molecular and physiochemical differences between alleles in the region described by a 15 Å radius around the presumed epitope. \*\*All includes those amino acid residues within the peptide binding groove and exposed excludes these when comparing between the two antigens listed on each row of the table.

mAb	HLA pairs	Number of amino acid residue differences between each antigen pair*		Total ΔHMS		Total ΔEMS		Ratio of K <sub>D</sub>
		All	Exposed	All	Exposed	All	Exposed	
SN607D8	A2, A68	5	1	15	13.4	12.09	12	2.47
SN230G6	A2, B57	22	12	35.7	20.1	35.7	22.7	1.33
WK1D12	B27, B7	15	9	40.2	25.1	38.6	27	3.01

demonstrates very starkly that the amount of serum antibody that binds to each bead is not a simple reflection of its serum concentration; it is also affinity dependent. The affinity difference for HLA-A2 is related to differences in kinetic rates; the Association rate is faster and the dissociation rate slower with SN230G6. The variation in affinity of a single mAb, SN607D8, for the two serologically related antigens, HLA-A2 and A68, could be due to interaction with different residues in the immediate vicinity of the distinguishing epitope which in turn may alter the spatial orientation and steric accessibility of the epitope from one antigen to the other. Thus, both the best fit and surface area involved in the overall interaction may determine the stability of binding [19].

Although the difference in dissociation constant is relatively small (within one order of magnitude), the kinetic parameters (Association and dissociation rates) differed more widely. Therefore, for comparisons, dissociation constants alone may not give a full picture. Close inspection of our data reveals that while the on-rates for binding of one antibody to different HLAs may be similar, the overall affinity may be more dependent on the dissociation rate, and this may be markedly slower for the target HLA that corresponds to the immunizing antigen (Fig. 4). These are useful observations as current solid phase assays do not distinguish these differences in binding characteristics which may be important in determining biological reactivity of the HLA-specific

## ARTICLE IN PRESS

S. Daga et al.

Human Immunology xxx (xxxx) xxx-xxxx

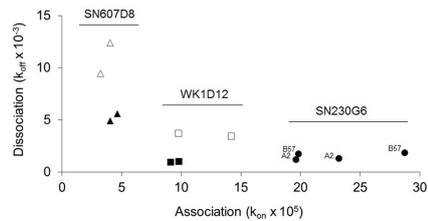


Fig. 4. Relationships between association (on) and dissociation rates (off) for the three mAbs and their target ligands, as follows:  $\Delta$ SN607D8 + HLA A2;  $\triangle$ SN607D8 + HLA A68;  $\square$ WK1D12 + HLA B7;  $\blacksquare$ WK1D12 + HLA B27;  $\bullet$ SN230G6 + HLA A2 and HLA B57. Where the target HLA and the HLA present during immunization are the same the symbols in the figure are filled (Respectively HLA A2 and B57 reactions with SN230G6 are identified). Each of the duplicate results are shown.

antibodies and their downstream effects. Thus, the higher affinity associated with a stimulating antigen suggests that a repeat antigen mismatch transplant might impart a greater risk than if only the epitope were repeated in an antibody incompatible transplant.

The high degree of sequence homology between alleles (and between the products of different loci) results in cross-reactivity due to shared immunogenic epitopes and eplets. We have shown that despite sharing a critical sequence, which confers specificity, substitutions in other parts of the target molecule can result in large affinity differences. It was not obvious which of the mismatched residues or their physicochemical properties accounted for these differences and this is most likely because we do not know the orientation of binding or the exact interface surfaces. This is also illustrated by the similar affinity of SN230G6 for HLA-A2 and B57, products of different loci, which implies that there might be an almost identical antibody-accessible surface on these two molecules. We are unable to resolve whether or not non-exposed residues in the peptide binding groove contributed to the binding kinetics of any of the antibodies studied, either directly or by virtue of the peptides presented. Although SN607D8 binds differently to HLA-A2 and A68, the only differences between these alleles within 15A of the presumed critical eplet (142TKH) are within the peptide binding groove. Taken together, this suggests that the model by which antibody binding is precisely centered around an eplet [18] may not be correct, at least for these antibodies. Ultimately these issues are likely to be resolved by structural imaging experiments.

In conclusion, the binding assay described here possesses advantages over current assays since it captures a broader, real-time picture of the binding events and levels, compared with simpler endpoint assays such as Luminex single antigen bead assay, ELISA and flow cytometry. There are, of course difficulties with SPR, the requirement for highly purified antibodies because the sensor chip surface is very sensitive to non-specific binding. By purifying the antibodies, we have been able to directly measure the binding kinetics of HLA antibodies and these measurements are likely to be critical in the understanding of the pathogenicity of these antibodies. Another exciting opportunity lies in the ability of SPR experiments to obtain thermodynamic data (such as free energy released) for these interactions, yielding further parameters of potential physiological value plus deepening our understanding of antibody function and biophysics. Our experiments have also underlined the limitations of our knowledge of how alloantibodies bind to their HLA targets. Although we can now show whether or not a particular antibody binds differently to its known ligands and measure the magnitude of any variation, we do not yet know completely the molecular basis underlying such differences.

## Acknowledgements

The research was funded by CLRN-NIHR West-Midlands (South) and Coventry Renal Research Fund, UHCW NHS Trust.

## References

- [1] R. Patel, P.I. Terasaki, Significance of the positive cross match test in Kidney Transplantation, *N. Engl. J. Med.* 280 (1969) 735–739.
- [2] A. Gupta, V. Iveson, M. Varagunam, S. Bodger, P. Sinnott, R.C. Thuraisingham, Pretransplant donor-specific antibodies in cytotoxic negative crossmatch kidney transplants: are they relevant? *Transplantation* 85 (2008) 1200–1204.
- [3] R. Higgins, D. Lowe, M. Hathaway, C. Williams, F.T. Lam, H. Kashi, L.C. Tan, C. Imray, S. Fletcher, K. Chen, N. Krishnan, R. Hamer, S. Daga, M. Edey, D. Zehnder, D. Briggs, Human leukocyte antigen antibody-incompatible renal transplantation: excellent medium-term outcomes with negative cytotoxic crossmatch, *Transplantation* 92 (2011) 900–906.
- [4] N. Singh, A. Djarnali, D. Lorentzen, J.D. Pirsch, G. Levenson, N. Neidlinger, B. Voss, J.R. Torrebalba, R.M. Hofmann, J. Odorico, L.A. Fernandez, H.W. Sollinger, M. Samaniego, Pretransplant donor-specific antibodies detected by single-antigen bead flow cytometry are associated with inferior kidney transplant outcomes, *Transplantation* 90 (2010) 1079–1084.
- [5] M. Willcombe, F. Brookes, E. Santos-Nunez, J. Galliford, A. Ballou, A. Mclean, C. Roufosse, H.T. Cook, A. Dorling, A.N. Warrens, T. Cairns, D. Taube, Outcome of patients with preformed donor-specific antibodies following alemtuzumab induction and tacrolimus monotherapy, *Am. J. Transplant.* 11 (2011) 470–477.
- [6] R. Higgins, D. Lowe, S. Daga, M. Hathaway, C. Williams, F.T. Lam, H. Kashi, L.C. Tan, C. Imray, S. Fletcher, N. Krishnan, P. Hart, D. Zehnder, D. Briggs, Pregnancy-induced HLA antibodies respond more vigorously after renal transplantation than antibodies induced by prior transplantation, *Hum. Immunol.* 76 (2015) 546–552.
- [7] R. Higgins, D. Lowe, M. Hathaway, F. Lam, H. Kashi, L.C. Tan, C. Imray, S. Fletcher, K. Chen, N. Krishnan, R. Hamer, D. Zehnder, D. Briggs, Rises and falls in donor-specific and third-party HLA antibody levels after antibody incompatible transplantation, *Transplantation* 87 (2009) 882–888.
- [8] N.S. Krishnan, R.M. Higgins, F.T. Lam, H. Kashi, S. Jobson, K. Ramaiyan, M. Rahman, A. Morris, HA-1 mismatch has significant effect in chronic allograft nephropathy in clinical renal transplantation, *Transplant. Proc.* 39 (2007) 1439–1445.
- [9] R.A. Montgomery, B.E. Lonze, K.E. King, E.S. Kraus, L.M. Kucirka, J.E. Locke, D.S. Warren, C.E. Simpkins, N.N. Dagher, A.L. Singer, A.A. Zachary, D.L. Segev, Desensitization in HLA-incompatible Kidney Recipients and Survival, *N. Engl. J. Med.* 365 (2011) 318–326.
- [10] R.M. Higgins, S. Daga, D.A. Mitchell, Antibody-incompatible kidney transplantation in 2015 and beyond, *Nephrol. Dial. Transplant.* 30 (2015) 1972–1978.
- [11] N. Khovanova, S. Daga, T. Shaikhina, N. Krishnan, J. Jones, D. Zehnder, D. Mitchell, R. Higgins, D. Briggs, D. Lowe, Subclass analysis of donor HLA-specific IgG in antibody-incompatible renal transplantation reveals a significant association of IgG with rejection and graft failure, *Transplant. Int.* (2015).
- [12] C. Lefebvre, D. Viglietti, C. Bentlejewski, J.P. Duong van Huyen, D. Verney, O. Aubert, J. Verine, X. Jouven, C. Legendre, D. Glotz, A. Loupy, A. Zeevi, IgG Donor-specific anti-human HLA antibody subclasses and kidney allograft antibody-mediated injury, *J. Am. Soc. Nephrol.* (2015).
- [13] A. Loupy, C. Lefebvre, D. Verney, C. Prugger, J.-P. Duong van Huyen, N. Mooney, C. Suberbielle, V. Frémeaux-Bacchi, A. Méjean, F. Desgrandchamps, D. Anglicheu, D. Nochy, D. Charon, J.-P. Empana, M. Delahousse, C. Legendre, D. Glotz, G.S. Hill, A. Zeevi, X. Jouven, Complement-binding anti-HLA antibodies and kidney-allograft survival, *N. Engl. J. Med.* 369 (2013) 1215–1226.
- [14] T. Shaikhina, D. Lowe, S. Daga, D. Briggs, R. Higgins, N. Khovanova, Machine learning for predictive modelling based on small data in biomedical engineering, *IFAC-PapersOnLine* (2015) 469–474.
- [15] C.J. Taylor, J.R. Chapman, A. Ting, P.J. Morris, Characterization of Lymphocytotoxic antibodies causing a positive crossmatch in Renal Transplantation-Relationship to primary and reagent outcome, *Transplantation* 48 (1989) 953–958.
- [16] N. El-Awar, P.I. Terasaki, J. Cai, C.-T. Deng, M. Ozawa, A. Nguyen, M. Lias, N. Conger, Epitopes of HLA-A, B, C, DR, DQ, DP and MICA antigens, *Clin. Transplant.* (2009) 295–321.
- [17] M. Marrari, R.J. Duquesnoy, Correlations between Terasaki's HLA class II epitopes and HLA/matchmaker-defined eplets on HLA-DR and -DQ antigens, *Tissue Antigens* 74 (2009) 134–146.
- [18] R.J. Duquesnoy, M. Marrari, A. Mulder, F.H.J. Claas, J. Mostecky, I. Balazs, Structural aspects of human leukocyte antigen class I epitopes detected by human monoclonal antibodies, *Hum. Immunol.* 73 (2012) 267–277.
- [19] A.R. Tambur, F.H.J. Claas, HLA epitopes as viewed by antibodies: what is it all about? *Am. J. Transplant.* 15 (2015) 1148–1154.
- [20] D.R. Davies, S. Sheriff, E.A. Padlan, Antibody-antigen complexes, *J. Biol. Chem.* 263 (1988) 10541–10544.
- [21] M. Marrari, J. Mostecky, A. Mulder, F. Claas, I. Balazs, R.J. Duquesnoy, Human monoclonal antibody reactivity with human leukocyte antigen class I epitopes defined by pairs of mismatched eplets and self-eplets, *Transplantation* 90 (2010) 1468–1472.
- [22] R.J. Duquesnoy, M. Marrari, L. Jelenik, A. Zeevi, F.H.J. Claas, A. Mulder, Structural aspects of HLA class I epitopes reacting with human monoclonal antibodies in Ig-

- binding, C1q-binding and lymphocytotoxicity assays, *Hum. Immunol.* 74 (2013) 1271–1279.
- [23] R.J. Duquesnoy, Human leukocyte antigen epitope antigenicity and immunogenicity, *Curr. Opin. Organ Transplant.* 19 (2014) 428–435.
- [24] D.H. Mallon, J.A. Bradley, P.J. Winn, C.J. Taylor, V. Kosmoliaptis, Three-dimensional structural modelling and calculation of electrostatic potentials of HLA Bw4 and Bw6 epitopes to explain the molecular basis for alloantibody binding: toward predicting HLA antigenicity and immunogenicity, *Transplantation* 99 (2015) 385–390.
- [25] D.H. Mallon, J.A. Bradley, C.J. Taylor, V. Kosmoliaptis, Structural and electrostatic analysis of HLA B-cell epitopes: inference on immunogenicity and prediction of humoral alloresponses, *Curr. Opin. Organ Transplant.* 19 (2014) 420–427.
- [26] V. Kosmoliaptis, T.R. Dafforn, A.N. Chaudhry, D.J. Halsall, J.A. Bradley, C.J. Taylor, High-resolution, three-dimensional modeling of human leukocyte antigen class I structure and surface electrostatic potential reveals the molecular basis for alloantibody binding epitopes, *Hum. Immunol.* 72 (2011) 1049–1059.
- [27] L. Boursier, W. Su, J. Spencer, Imprint of somatic hypermutation differs in human immunoglobulin.
- [28] J.P. Ways, P. Parham, The binding of monoclonal antibodies to cell-surface molecules, *J. Biochem.* 216 (1983) 423–432.
- [29] C. McMurrey, D. Lowe, R. Buchill, S. Daga, D. Royer, A. Humphrey, S. Cate, S. Osborn, A. Mojsilovic, R. VanGundy, W. Bardet, A. Duty, D. Mojsilovic, K. Jackson, P. Stastny, D. Briggs, D. Zehnder, R. Higgins, W. Hildebrand, Profiling antibodies to class II HLA in transplant patient sera, *Hum. Immunol.* 75 (2014) 261–270.
- [30] O.H. Lowry, N.J. Rosebrough, A.L. Farr, R.J. Randall, Protein measurement with the Folin phenol reagent, *J. Biol. Chem.* 193 (1951) 265–275.
- [31] K.V. Price, R.M. Storn, J. Lampinen, A. Differential Evolution – A Practical Approach to Global Optimization, Springer-Verlag, Berlin, 2005.
- [32] N.D. Evans, H.A.J. Moyses, D. Lowe, D. Briggs, R. Higgins, D. Mitchell, M.J. Chappell, Structural identifiability of surface binding reactions involving heterogeneous analyte: application to surface plasmon resonance experiments, *Automatica* 49 (2013) 48–57.
- [33] T. Bravman, V. Bronner, K. Lavie, A. Notcovich, G.A. Papalia, D.G. Myska, Exploring “one-shot” kinetics and small molecule analysis using the ProteOn XPR36 array biosensor, *Anal. Biochem.* 358 (2006) 281–288.
- [34] L. Heinrich, N. Tissot, D.J. Hartmann, R. Cohen, Comparison of the results obtained by ELISA and surface plasmon resonance for the determination of antibody affinity, *J. Immunol. Methods* (2009).
- [35] A. Buhl, J.H. Metzger, N.H.H. Heegaard, P. Von Landenberg, M. Fleck, P.B. Lippa, Novel biosensor-based analytic device for the detection of anti-double-stranded DNA antibodies, *Clin. Chem.* 53 (2007) 334–341.
- [36] S. Andrejevic, I. Jeremic, M. Sefik-Bukilica, M. Nikolic, B. Stojimirovic, B. Bonaci-Nikolic, Immunoserological parameters in SLE: high-avidity anti-dsDNA detected by ELISA are the most closely associated with the disease activity, *Clin. Rheumatol.* 32 (2013) 1619–1626.
- [37] F. Kushihata, J. Watanabe, A. Mulder, F. Claas, J.C. Scornik, Human leukocyte antigen antibodies and human complement activation: role of IgG subclass, specificity, and cytotoxic potential, *Transplantation* 78 (2004) 995–1001.

**B.2 Paper at 8th IFAC Symposium on Biological and Medical Systems**

8th IFAC Symposium on Biological and Medical Systems  
The International Federation of Automatic Control  
August 29-31, 2012, Budapest, Hungary



## Estimation of antibody binding affinities in incompatible blood type renal transplants from surface plasmon resonance

H.A.J. Moyle\* D. Lowe\*\* D. Briggs\*\*\* R. Higgins\*\*\*\* A. Bentall\*\*\*\*\*  
S. Ball\*\*\*\*\* D. Mitchell\*\* M.J. Chappell\*\*\*\*\* D. Zehnder\*\* N.D. Evans\*\*\*\*\*

\*MOAC, University of Warwick, Coventry, CV4 8JL  
UK (e-mail: h.a.j.moyle@warwick.ac.uk).

\*\* CSRI, University of Warwick, Coventry, CV4 7AL UK

\*\*\* Histocompatibility Laboratory, NHS Blood transplant, Birmingham, B15 2SG

\*\*\*\*Transplant unit, University hospitals Coventry and Warwickshire, Coventry CV2 2DX UK

\*\*\*\*\* University hospital Birmingham, Birmingham CV2 2DX UK

\*\*\*\*\*School of Engineering, University of Warwick, Coventry, CV4 7AL

**Abstract:** Blood group (ABO) incompatible transplants carry an increased risk of rejection. This risk could be dramatically reduced by the removal and suppression of the antibody types that would attack the donor organ. A prerequisite for this removal is an experimental procedure that can estimate the binding affinities of multiple antibodies from patient blood samples. This paper presents the usage of surface plasmon resonance (SPR) experiments with a pre existing mathematical model and a recently created expanded version of it that can estimate multiple antibody binding affinities from parallel experiments. SPR experiments were conducted on purified patient antibody samples of the IgG and IgA isotype, as well as mixed antibody from the other isotypes. The result of these experiments was analyzed with both mathematical models, and the expanded model was demonstrated to give a vastly improved fit. Estimates of antibody binding affinity were compared between samples and the non IgG/IgA protein sample was seen to have the highest binding affinity.

**Keywords:** parameter estimation, structural identifiability, surface plasmon resonance, renal transplant

### 1. INTRODUCTION

One strategy to reduce a patient's risk of renal transplant rejection is the use of tailored immunosuppressant drugs. Before such drugs can be used, a better understanding of the abundance of each type of antibody that a transplant can tolerate is needed. This paper presents the use of mathematical models to evaluate the dynamics of antibody binding in surface plasmon resonance (SPR) experiments with fractions of patient blood. These models and their fits are important tools for identifying and characterising potential risks in individual patients undergoing blood type (ABO) incompatible transplantation, and as a result creating tailored immunosuppressant drugs (Higgins et al., 2007).

Currently a patient receiving a kidney transplant undergoes crossmatch to test for compatibility with a potential donor. Such a test fails if the patient has pre-formed donor specific antibodies (DSAs) associated with the ABO blood group or human leukocyte antigen tissue type (HLA). In such cases pre-transplant antibody removal can be used to allow the operation to proceed (Higgins et al., 2009). However, this approach carries a large immunological risk, with heightened chances of infection, dysfunction, and rejection.

One way of decreasing these risks would be to create a system of experiments and parameter estimation procedures

to enable identification of the set of antibody binding characteristics for each individual. Once these were known, immunosuppression can potentially be tailored to the patient's individual needs. This immunosuppression would act specifically on the antibodies that attack the transplant, lowering and maintaining them at safe levels. This approach has the potential to lower rejection rates and, as a result, could improve the availability of and reduce waiting times for donor organs.

A starting point for this kind of procedure would be its use on refined patient blood samples. As a result this paper deals with such use on refined patient samples.

SPR and parallel SPR experiments are methods already established for determining binding affinities of cultured antibodies (Englebienne, 1998) and (MacKenzie et al., 1996). Parallel SPR experiments were used rather than the established method for measuring ABO specific antibody levels, namely haemagglutination (HA), because of problems with the reproducibility of results (Aikawa et al., 2003).

Mathematical models have been developed that can be used to estimate parameters and establish binding affinities using data from SPR experiments. The simplest models include a single differential equation, which assumes that over a short period of time the analyte becomes well mixed, and the rate

8th IFAC Symposium on Biological and Medical Systems  
August 29-31, 2012, Budapest, Hungary

at which the reactants bind is solely dependent on the laws of mass action. However, it has been demonstrated that the interactions measured are determined not only by the availability of both reactants but the effects of transport processes (Myszka et al., 1998). As a result the use of these models in parameter fitting can introduce systematic errors (Chaiken, Rosé and Karlsson, 1992). To avoid this problem Myszka et al. (1998) and (Edwards, 1999) consider a boundary layer between the well mixed analyte and the ligand where reactions take place, and develop a model that assumes the analyte on the boundary layer is in a quasi steady state.

These models consist of a single ordinary differential equation with bound concentration as the state variable. All of these models were designed to deal with experiments where only one binding reaction is taking place. However, even within an individual antibody isotype, from a single patient, antibodies are polyclonal and as a result have a range of different affinities. In Evans et al. (unpublished) existing models were expanded to include the effects of two antibody types in a single mixture. This paper also provides full structural identifiability analyses of these models, and demonstrates that they could feasibly be used for parameter estimation if the total concentration of analyte in the mixture is known.

## 2. COMPARTMENTAL MODELS

Two models are fitted to the data to give estimates of sample binding affinities. These are the effective rate constant (ERC) model from Myszka (1998) and an expanded form of this model from Evans et al (unpublished) that allows two types of antibody to bind simultaneously. These are presented in terms of the parameters and variables in Table 1.

The original ERC model is:

$$\dot{B}_1(t) = \frac{k_{a1} D I_1 (R - B_1) - k_{d1} B_1}{1 + \left(\frac{k_{a1}}{k_M}\right) (R - B_1)} \quad (1)$$

This assumes the analyte is of a single binding type, and is therefore termed the *homogenous model*.

The expanded model is:

$$\dot{B}_1(t) = \frac{k_{a1} D I_1 (R - B_1 - B_2) - k_{d1} B_1}{1 + (k_{a1}/k_M)(R - B_1 - B_2)} \quad (2)$$

$$\dot{B}_2(t) = \frac{k_{a2} D I_2 (R - B_1 - B_2) - k_{d2} B_2}{1 + (k_{a2}/k_M)(R - B_1 - B_2)}$$

This assumes the analyte is of two binding types, and is therefore termed the *heterogeneous model*.

To create distinct periods over which both association and dissociation can be observed, the machine operates through two stages: the association phase  $[t_0, t_1]$  in which analyte is pumped across the ligand, and the dissociation phase,  $> t_1$ ,

where buffer is pumped across the ligand. This process is modelled using the variable input concentration defined by:

$$I_i(t) = \begin{cases} DC_i: t \in [t_0, t_1] \\ 0: t \notin [t_0, t_1] \end{cases} \quad (3)$$

The state variables of these models are related to the output by:

$$y(t) = \alpha \sum_{i=1}^n B_i(t) \quad (4)$$

where  $n$  is the number of assumed binding types.

With appropriate outcomes from the structural identifiability analyses this output relation allows the unknown parameters in the models to be uniquely determined from the experiment used to collect the data.

Table 1. Model parameters and variables.

Symbol	Meaning
$B_i(t)$	Density of the analytes $i^{\text{th}}$ component at time $t$ bound to the chip in units of $\text{ng}/\text{mm}^2$
$k_{ai}, k_{di}$	Constants of association and dissociation in units of $\text{mm}^3/\text{ng s}$ and $1/\text{s}$ respectively
$K_M$	Transport coefficient between the analyte and the surface of ligand in units of $\text{mm}^3/\text{ng s}^2$
$R$	Maximum density of bound analyte on the chip in units of $\text{ng}/\text{mm}^2$
$I_i(t)$	Inlet concentration of analyte with binding type $i$ at time $t$ in units of $\text{ng}/\text{mm}^3$
$D$	Dilution in the channel (without units)
$C_i$	Concentration of antibodies with binding type $i$ in analyte sample in units of $\text{ng}/\text{mm}^3$
$[t_0, t_1]$	The association phase, the period of time in which analyte is pumped through
$y(t)$	Measured bound analyte on the chip at time $t$ in the response units (RU) of the sensorgrams
$\alpha$	Conversion factor between the units of the model and the response units ( $1\text{RU} = 10^{-3} \text{ng mm}^{-2}$ )

The goal of this research is not simply to model SPR experiments or to fit parameters to models, but to develop a method, including mathematical and experimental methods, which can determine the affinity with which antibodies from a patient blood sample can bind to donor antigen. This binding affinity is represented with the symbol  $A_i$  and defined by:

$$A_i = \frac{k_{ai}}{k_{di}} \quad (5)$$

## 3. EXPERIMENTATION AND PARAMETER ESTIMATION

SPR experiments were conducted with the ProteOn XPR36 platform. Three analytes were used: IgG, IgA, and non-IgG/A protein. These were taken from patient serum and separated.

These analytes were simultaneously passed over three ligands consisting of A trisaccharide linker with surface concentrations of 0.3 mg/ml, 0.2 mg/ml and 0.1 mg/ml.

8th IFAC Symposium on Biological and Medical Systems  
August 29-31, 2012, Budapest, Hungary

The association phase began at 7.2s and lasted until 981 s. The experiment lasted a further 666 s, giving a total time of 1647 s.

Parameter estimation was conducted using a differential evolution algorithm, implemented using a user-contributed MATLAB Toolbox, coupled with FACSIMILE for Windows 4 (MCPA software, UK).

As three different antibody components were used in these experiments three different sets of parameters were estimated. As each antibody component was analysed in parallel at three different dilutions across three chips, there are parameters shared by these parallel experiments to be estimated: three maximum bound densities ( $R$ ) and transport coefficients ( $K_M$ ). These parameters are labelled with a subscript corresponding to their dilution, in the order 0.3 mg/ml, 0.2 mg/ml then 0.1mg/ml:  $K_{M1}, K_{M2}, K_{M3}$  and  $R_1, R_2, R_3$ .

3.1 Homogeneous model

The fit of the homogenous model (1) to the data is presented in Figure 1. It is a poor visual match for the data in parts a) and c). For these the bound concentration in the early part of the association phase is underestimated, in the mid part is overshoot, and in the last part tends to a value much lower than seen in the experiments. In the early part of the dissociation phase, the bound concentration from the experiments reduces rapidly but the model output does not, and predicts higher than observed bound concentrations in each channel and less than observed in the later part of the dissociation phase.

The homogeneous model gives a better visual match for the data in part b). It appears that the experiments reached saturation in each channel at a much lower bound concentration than the experiments in parts a) and c), and as a result has a low signal to noise ratio. In the early part of the association phase the model bound concentration increases rapidly until it reaches the saturation point where this increase abruptly stops; in the experiment a similar increase is seen except that the rate of increase decreases as the bound concentration approaches saturation. In the early part of the dissociation phase the model bound concentration falls to zero much faster than the experimental bound concentration.

Estimated parameter values are presented in Table 2 along with a corresponding value for the standard deviation of their natural logarithm (SDLN) in brackets. FACSIMILE works in terms of internal parameters that are the natural logarithm of the parameters used here and presents these as a measure of confidence in its estimates.

Some parameters are provided with NA rather than a SDLN value. This signifies that the parameter was not-well-determined by the data, and has an SDLN of greater than 0.2. If this is the case the parameter value given was chosen to minimise residual sum of squares(RSS).

From these fits the binding affinity can be calculated for each analyte:  $3 \times 10^{-2} \text{mm}^3/\text{ng}$  for IgG,  $2.8 \times 10^{-4} \text{mm}^3/\text{ng}$  for IgA and  $9.5 \times 10^{-2} \text{mm}^3/\text{ng}$  for non IgG/A protein. The RSS for this model summed across all experiments is  $8.5 \times 10^5$ .

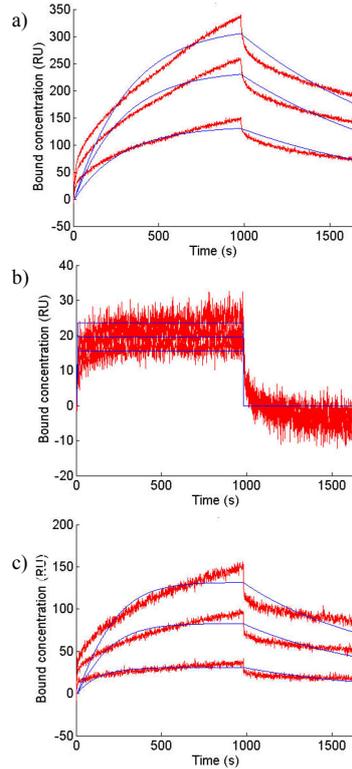


Figure 1. Fits of the homogeneous model (1) (blue lines) to data from parallel experiments (red lines) with trisaccharide linker densities of 0.3 mg/ml, 0.2 mg/ml and 0.1 mg/ml. Analyte in (a) is IgG antibody, (b) is IgA antibody, (c) is non IgG/IgA protein.

Table 2 parameters estimated with the homogeneous model with SDLNs in parenthesis.

	IgG	IgA	non-IgG/A protein
$k_{a1}$	$2.7 \times 10^{-3}(0.01)$	$3.1 \times 10^{-3}(NA)$	$2.1 \times 10^1(NA)$
$k_{d1}$	$8.8 \times 10^{-3}(NA)$	$1.1 \times 10^{-3}(NA)$	$2.2 \times 10^2(0.03)$
$K_{M1}$	$9.5 \times 10^{-3}(0.48)$	$7.1 \times 10^{-3}(NA)$	$3.6 \times 10^{-6}(0.01)$
$K_{M2}$	$2.0 \times 10^{-3}(NA)$	$1.2 \times 10^{-3}(0.08)$	$2.1 \times 10^{-6}(0.01)$
$K_{M3}$	$1.5 \times 10^{-3}(NA)$	$6.3 \times 10^{-3}(0.06)$	$1.3 \times 10^{-6}(0.02)$
$R_1$	$4.1 \times 10^1(0.01)$	$8.8 \times 10^1(0.09)$	$1.4 \times 10^1(0.06)$
$R_2$	$3.1 \times 10^1(0.02)$	$7.4 \times 10^1(0.09)$	$9.2 \times 10^2(0.02)$
$R_3$	$1.8 \times 10^1(0.01)$	$6.0 \times 10^1(0.09)$	$3.4 \times 10^2(0.06)$
RSS	$7.4 \times 10^5$	$4.2 \times 10^4$	$6.4 \times 10^4$

8th IFAC Symposium on Biological and Medical Systems  
August 29-31, 2012, Budapest, Hungary

### 3.2 Heterogeneous model

The fit of the heterogeneous model (2) to the data is presented in Figure 2. Whilst it offers a dramatically better visual match it has some disparities similar to the homogeneous model. The model fits much closer to the data in parts a) and c). For these two data sequences it accurately estimates the bound concentration throughout the association phase, except in the very early part where it overshoots. The experimental bound concentration in the dissociation phase is reproduced accurately in part c), but in part a) the model appears to be tending asymptotically towards different values than the experiments.

In part b) where there were saturation problems, this model gives a very similar estimate to that provided by the homogeneous model. This model, however, predicts the slowing of the increase in bound concentration as it reaches saturation in the association phase, and again as it decays to zero in the dissociation phase.

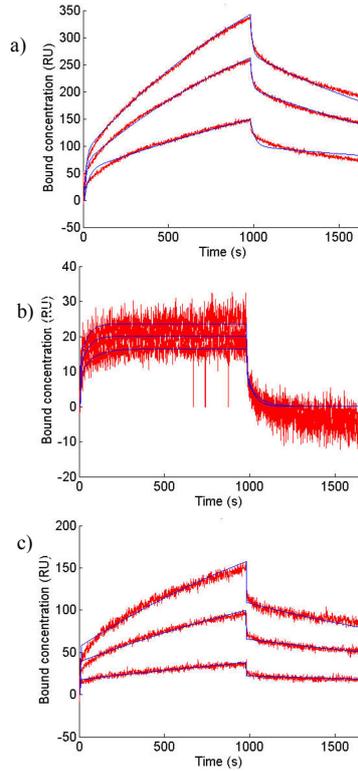
Estimated parameters are presented in Table 3.

**Table 3** parameters estimated with the heterogeneous model with SDLNs in parenthesis.

	IgG	IgA,	non-IgG/A protein
$k_{a1}$	$5.2 \times 10^{-3}$ (0.06)	$1.6 \times 10^{-6}$ (NA)	$3.7 \times 10^1$ (0.09)
$k_{a2}$	$4.1 \times 10^{-3}$ (0.07)	$9.4 \times 10^{-6}$ (NA)	$4.5 \times 10^{-3}$ (0.03)
$k_{d1}$	$7.8 \times 10^{-4}$ (0.01)	$6.0 \times 10^{-3}$ (NA)	$8.4 \times 10^{-1}$ (0.04)
$k_{d2}$	$5.4 \times 10^{-2}$ (0.01)	$1.9 \times 10^{-2}$ (0.06)	$6.7 \times 10^2$ (0.04)
$K_{M1}$	$1.6 \times 10^4$ (0.07)	$2.7 \times 10^4$ (0.31)	$1.7 \times 10^0$ (0.02)
$K_{M2}$	$1.5 \times 10^4$ (0.07)	$1.7 \times 10^4$ (0.27)	$2.1 \times 10^{-5}$ (0.03)
$K_{M3}$	$8.9 \times 10^4$ (0.06)	$9.7 \times 10^5$ (0.23)	$1.0 \times 10^{-5}$ (0.02)
$R_1$	$1.4 \times 10^0$ (0.02)	$1.3 \times 10^0$ (0.08)	$2.3 \times 10^{-1}$ (0.01)
$R_2$	$1.0 \times 10^0$ (0.02)	$1.1 \times 10^0$ (0.08)	$1.2 \times 10^{-1}$ (0.01)
$R_3$	$5.3 \times 10^{-1}$ (0.02)	$8.8 \times 10^{-1}$ (0.08)	$4.1 \times 10^{-2}$ (0.08)
$C_1$	$8.0 \times 10^0$ (0.06)	$2.2 \times 10^1$ (0.03)	$6.9 \times 10^0$ (0.02)
$C_2$	$9.2 \times 10^1$ (-)	$7.8 \times 10^1$ (-)	$1.0 \times 10^2$ (-)
RSS	$2.5 \times 10^3$	$3.5 \times 10^4$	$1.0 \times 10^4$

From these fits we can calculate the binding affinity estimated by this model for each analyte:  $6.7 \times 10^{-2} \text{mm}^3/\text{ng}$  and  $7.6 \times 10^{-4} \text{mm}^3/\text{ng}$  for IgG,  $2.6 \times 10^{-4} \text{mm}^3/\text{ng}$  and  $4.9 \times 10^{-4} \text{mm}^3/\text{ng}$  for IgA,  $4.4 \times 10^1 \text{mm}^3/\text{ng}$  and  $6.7 \times 10^{-6} \text{mm}^3/\text{ng}$  for non IgG/A protein.

These results are supported by the structural identifiability analysis of Evans et al. (unpublished). These show that if the total concentration of analyte is known, and the indexes of the parameters are ordered then only models with identical parameters will produce an identical output. In the case of this model, the indexes of the parameters are chosen so the reactant with lower concentration is considered binding type 1 and the reactant with higher concentration is considered binding type 2.



**Figure 2.** Fits of the heterogeneous model (2) (blue lines) to data from parallel experiments (red lines) with trisaccharide linker densities of 0.3mg/ml, 0.2mg/ml and 0.1mg/ml. Analyte in (a) is IgG antibody, (b) is IgA antibody, (c) is non IgG/IgA protein.

These estimates of binding affinity, even within a single isotype, are spread by orders of magnitude. This allows the model to better duplicate the rapid association and dissociation seen at the very beginning of each phase, or the much slower association and dissociation seen throughout the rest of the phase (compare Figure 1 and Figure 2).

The RSS for this model summed across all experiments is  $4.8 \times 10^4$ .

### 3.3 Comparison

The simulated output of the homogeneous model captures some of the dynamics of the data (Figure 1), although it was unable to match the rapid changes in bound concentration at the start of either phase for the IgG experiment (a) or the non

8th IFAC Symposium on Biological and Medical Systems  
August 29-31, 2012, Budapest, Hungary

IgG/IgA experiment (c), and overestimated the speed that that the experiment would reach saturation in part (b).

The heterogeneous model (Figure 2) dramatically improved on this. It was better able to match the rapid changes in bound concentration at the start of the phases, although it overestimated this rapid change in the IgG experiment (a) or the non IgG/IgA experiment (c).

The binding affinity estimates for both models are separated by orders of magnitude. This is to be expected since the homogeneous model assumes that all of the analyte in the experiment binds with the same affinity, and so will not pick up a small amount of analyte that binds with higher affinity than the rest. This is a problem even if it is used to fit data from experiments with antibody of one isotype, because the antibody will be polyclonal and contain a range of affinities. The heterogeneous model assumes that there are two binding types within the analyte and as a result is better able to deal with experiments with polyclonal data. This is reflected by the difference in RSS between the models.

There are considerable differences in the residual sum of squares of the models. The heterogeneous model has a RSS of 0.05% of the RSS for the homogenous model for the IgG experiment, 83% for the IgA experiment, and 16% for the non IgG/IgA experiment. This suggests that the heterogeneous model is an improvement as a model for each experiment, though this is much more prominent for IgG. This improvement is also reflected in the reduction in RSS for the model summed across all experiments; the total RSS for the heterogeneous model is 6% of the total RSS for the homogeneous model.

As there are large differences in the reduction of the RSS values between models this may be a motivating case for the use of the Akaike information criterion (AIC), or the Bayesian information criterion (BIC). A potential problem with this approach is that this may lead to poorer estimates for the affinities.

#### 4. CONCLUSIONS

Homogeneous and heterogeneous versions of a standard model for binding of an analyte to an immobilised ligand have been applied to SPR experiments performed on blood samples from ABO incompatible transplant patients. The heterogeneous model resulted in better fits for each of the experiments individually and when considered as a group.

The results from this set of experiments and estimates from the samples from a single transplant patient can by no means be generalised and used to compare IgG, IgA and non IgG/IgA antibodies in general. However, knowledge of the binding affinities of these antibodies against blood type A antigen could inform a patient's treatment, and the procedures that were used to obtain this information could be used for other patients.

The ranking of the analytes in order of highest affinity for the homogeneous model was IgA, non-IgG/A, then IgG. The

ranking for the heterogeneous model was non-IgG/A, IgG then IgA. This ordering has clinical significance, as it may be used to determine which antibody subtypes contain antibodies that could damage the transplant, and should be suppressed. However, it is unclear what patient antibody binding affinities a donor organ can endure without dysfunction and rejection. It may be the case that only one or two of these antibody groupings would need to be suppressed for this patient.

There are a number of reasons for the differences in analyte affinity between the models. The homogeneous model, assumes that all of the analyte in the experiment binds with the same affinity, as a result it will not pick up a small amount of analyte that binds more aggressively than the rest.

This is demonstrated by the results of the non-IgG/IgA protein fit. Included in that analyte was a small amount of IgM antibody as well as non antibody protein. If results from the homogeneous model were used as the basis for patient treatment, there would be a significant risk that small groups of antibody with higher affinities could exist in one of the fractions taken from patient sample. These could have their higher affinities unrecognised, and as a result not be suppressed, and risk damaging the transplant organ.

A possible problem with this is that antibodies of a single analyte may be expected to have many more than two different binding affinities. The heterogeneous version of the model could also be used to estimate parameters for experiments on an analyte containing any number of antibody binding affinities if there were suitable results from identifiability analyses. However, with this would come the problems of over-fitting and model selection.

#### REFERENCES

- Aikawa, A., Yamashita, M., Hadano, T., Ohara, T., Arai, K., Kawamura, T. and Hasegawa, A. (2003) 'ABO incompatible kidney transplantation --immunological aspect-', *Exp Clin Transplant*, vol. 1, no. 2, pp. 112-118.
- Chaiken, I., Rosè, S. and Karlsson, R. (1992) 'Analysis of macromolecular interactions using immobilized ligands.', *Anal Biochem*, vol. 201, no. 2, pp. 197-210.
- Edwards, D.A. (2001) 'The effect of a receptor layer on the measurement of rate constants.', *Bull Math Biol*, vol. 63, no. 2, pp. 301-327.
- Englebienne, P. (1998) 'Use of colloidal gold surface plasmon resonance peak shift to infer affinity constants from the interactions between protein antigens and antibodies specific for single or multiple epitopes.', *Analyst*, vol. 123, no. 7, pp. 1599-1603.
- Evans, N.D., Moyses, H.A.J., Lowe, D., Briggs, D., Higgins, R., Mitchel, D., Chappell, M.J. and Zehnder, D. (unpublished) 'Estimation of monoclonal IgM binding affinities from surface plasmon experiments', submitted to *Automatica*.
- Higgins, R., Hathaway, M., Lowe, D., Lam, F., Kashi, H., L.C., T., Imray, S., Fletcher, S., Zehnder, D., Chen, K., Krishnan, N., Hamer, R. and Briggs, D. (2007) 'Blood

**8th IFAC Symposium on Biological and Medical Systems**  
**August 29-31, 2012, Budapest, Hungary**

- levels of human leucocyte antigen antibodies after renal transplantation: resolution and rejection in the presence of circulating donor specific antibody', *Transplantation*, vol. 84, pp. 876-884.
- Higgins, R., Lowe, D., Hathaway, M., Lam, F., Kashi, H., Tan, L.C., Imray, C., Fletcher, S., Chen, K., Krishnan, N., Hamer, R., Zehnder, D. and Briggs, D. (2009) 'Rises and falls in donor-specific and third-party HLA antibody levels after antibody incompatible transplantation.', *Transplantation*, vol. 87, no. 6, pp. 882-888.
- MacKenzie, C.R., Hiram, T., Deng, S.J., Bundle, D.R., Narang, S.A. and Young, N.M. (1996) 'Analysis by surface plasmon resonance of the influence of valence on the ligand binding affinity and kinetics of an anti-carbohydrate antibody.', *J Biol Chem*, vol. 271, no. 3, pp. 1527-1533.
- Myszka, D.G., He, X., Dembo, M., Morton, T.A. and Goldstein, B. (1998) 'Extending the range of rate constants available from BIACORE: interpreting mass transport-influenced binding data.', *Biophys J*, vol. 75, no. 2, pp. 583-594.