

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

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

<http://go.warwick.ac.uk/wrap/3133>

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.

# **Refolding Of Protein Kinases**

Thesis submitted for the Degree of Doctor of Philosophy

**Richard H. Cowan**

Department of Biological Sciences, University of Warwick

September 2009

# **Contents**

	Page Number
Contents	i
List of Figures and Tables	vi
Acknowledgements	xiv
Declaration	xv
Summary	xvi
List of Abbreviations	xvii

## **Chapter 1. Introduction**

<b><i>1.1 Protein Folding</i></b>	<b>1</b>
1.1.1 Protein Folding and Common Protein Structures	2
1.1.2 Models and Mechanisms of Protein Folding	2
1.1.3 The Folding Energy Landscape	3
1.1.4 Protein Folding Intermediates	5
1.1.5 Techniques for Characterizing the Folding Pathway	7
1.1.6 Folding Probes	10
<b><i>1.2 Protein Refolding</i></b>	<b>14</b>
1.2.1 The Formation of Bacterial Inclusion Bodies	15
1.2.2 The Refolding Problem	17
1.2.3 The Preparative Refolding of Proteins	18
<b><i>1.3 Protein Kinases</i></b>	<b>19</b>
1.3.1 The Human Kinome	19
1.3.2 p38 $\alpha$ , a Previously Studied Kinase	22
1.3.3 AKT2	24
1.3.4 KIS	26
1.3.5 PhK	27
1.3.6 TTK, a Dual Specificity Kinase	28
<b><i>1.4 Aims</i></b>	<b>29</b>

## Chapter 2. Design, Testing and Evaluation of a Protein Refolding Screen using p38 $\alpha$ as a Model Protein Kinase.

<b>2.1 Introduction</b>	32
<b>2.2 Materials and Methods</b>	33
2.2.1 Materials	33
2.2.2 Transformation of <i>E.coli</i> with Plasmid DNA	34
2.2.3 Preparation of Plasmid DNA	34
2.2.4 Expression of p38 $\alpha$ as Inclusion Bodies	34
2.2.5 Isolation and Solubilisation of Inclusion Bodies	35
2.2.6 Production of Native p38 $\alpha$	36
2.2.7 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)	36
2.2.8 Mass Spectrometry	36
2.2.9 Design of a Refolding Screen for Kinases	37
2.2.10 Refolding of Human p38 $\alpha$	38
2.2.11 ePAGE of Refolded Protein	38
2.2.12 Silver Staining of ePAGE and SDS-PAGE gels	38
2.2.13 Denaturing Capillary Electrophoresis	39
2.2.14 Analytical Size Exclusion Chromatography	39
2.2.15 Detection of Binding Activity by Surface Plasmon Resonance	40
2.2.16 Large Scale Refolding	40
2.2.17 Circular Dichroism	41
2.2.18 iCycler Thermal Stability Measurements	41
<b>2.3 Results</b>	42
2.3.1 Expression of p38 $\alpha$ as Inclusion Bodies	42
2.3.2 Expression of p38 $\alpha$ in a Soluble Form	43
2.3.3 Comparison of the Monomeric State of the Two Forms of p38 $\alpha$	43
2.3.4 Design of a Protein Refolding Screen	44
2.3.5 Analytical Readouts for Refolding Screens	48
2.3.6 Refolding Screen for p38 $\alpha$	58



2.3.7	Thermal Melting Analysis of p38 $\alpha$ at Various pHs	61
2.3.8	Larger Scale Refolding	63
2.3.9	Far-UV Circular Dichroism of Native and Refolded p38 $\alpha$	64
2.3.10	Binding of Refolded p38 $\alpha$ to SB202190	65
<b>2.4</b>	<b><i>Discussion</i></b>	<b>67</b>

## **Chapter 3. Application of the Refolding Screen to Additional Protein Kinases and Refinement of the Screening Procedure**

<b>3.1</b>	<b><i>Introduction</i></b>	<b>70</b>
<b>3.2</b>	<b><i>Materials and Methods</i></b>	<b>72</b>
3.2.1	Materials	72
3.2.2	Gateway Cloning	72
3.2.2.1	Gateway Cloning – BP Reaction	73
3.2.2.2	Gateway Cloning – LR Reaction	73
3.2.3	Production of Kinase Panel Inclusion Bodies	74
3.2.4	Screening Refolding of the Kinase Panel	74
3.2.5	Fractional Factorial Screens for the Refolding of Protein Kinases	75
3.2.6	Cumulative Probability of Studentised Residuals	75
<b>3.3</b>	<b><i>Results</i></b>	<b>75</b>
3.3.1	Selection of the Kinase Panel	75
3.3.2	Production of an Expression Construct for PhK Kinase Domain	80
3.3.3	Expression of Kinase Panel Proteins as Inclusion Bodies	84
3.3.4	The Refolding of AKT2	85
3.3.5	The Refolding of KIS	90
3.3.6	The Refolding of PhK	96
3.3.7	The Refolding of TTK	102
3.3.8	Fractional Factorial Screen Design	109
3.3.9	Fractional Factorial Screen Results	114
<b>3.4</b>	<b><i>Discussion</i></b>	<b>126</b>

## **Chapter 4. Mutagenesis, Expression and Characterisation of Human TTK tryptophan Mutants**

<b>4.1 Introduction</b>	130
<b>4.2 Materials and Methods</b>	133
4.2.1 Materials	133
4.2.2 Site Directed Mutagenesis	133
4.2.3 Expression of Wild Type TTK and TTK Tryptophan Mutants	134
4.2.4 Purification of Wild Type TTK and TTK Tryptophan Mutants	134
4.2.5 Tryptophan Fluorescence of TTK	136
4.2.6 Circular Dichroism of TTK	136
4.2.7 Curve Fitting for Tryptophan Emission Spectra	136
<b>4.3 Results</b>	137
4.3.1 Wild Type TTK Expression Construct	137
4.3.2 Creation of TTK Tryptophan Mutants Using Site-Directed Mutagenesis	138
4.3.3 Expression of Wild Type TTK and TTK Tryptophan Mutants	141
4.3.4 Purification of Wild Type TTK and TTK Tryptophan mutants	142
4.3.5 Mass Spectrometry of TTK and TTK Tryptophan Mutants	147
4.3.6 Far-UV CD Spectra of TTK and TTK Tryptophan Mutants	151
4.3.7 Thermal Melting Analysis of TTK and TTK Tryptophan Mutants	153
4.3.8 Tryptophan Fluorescence Spectra of TTK and TTK Tryptophan Mutants	157
4.3.9 Spectral Broadening of Tryptophan Emission Spectra	159
<b>4.4 Discussion</b>	162

## **Chapter 5. Equilibrium Folding of Human TTK Protein Kinase Domain**

<b>5.1 Introduction</b>	165
-------------------------	-----

<b>5.2 Materials and Methods</b>	165
5.2.1 Materials	165
5.2.2 Equilibrium Folding and Unfolding	166
5.2.3 Data Normalisation	166
5.2.4 Analysis of Equilibrium Folding Curves	167
5.2.4.1 Two State Folding Transitions	167
5.2.4.2 Three State Folding Transitions	168
<b>5.3 Results</b>	170
5.3.1 Equilibrium Folding and Unfolding of Wild Type TTK	170
5.3.2 Equilibrium Folding and Unfolding of TTK <sup>W612F W622F</sup>	182
5.3.3 Equilibrium Folding and Unfolding of TTK <sup>W612F W622F W628F</sup>	193
5.3.4 Comparison of Unfolding and Refolding of TTK	206
<b>5.4 Discussion</b>	210

## Chapter 6. General Discussion

6.1 The Refolding Problem and Shared Protein Folds	215
6.2 The Refolding of Five Protein Kinases through a Broad Refolding Screen	216
6.3 The Refolding of Five Protein Kinases through a Fraction Factorial Screen	218
6.4 The Role of the Core Tryptophan of the Kinase Domain	219
6.5 The Formation and Nature of Intermediate States on the Folding Pathway of TTK and p38 $\alpha$ .	222
6.6 Super-imposable and Non Super-imposable Folding Curves	225
6.7 Simple Folding and Cooperative Folding	225
6.8 Does the Kinase Domain Fold Via a Common Pathway?	226
6.9 Future Work	228

<b>References</b>	229
-------------------	-----

## Appendix

Reprint of Cowan *et al*, 2008

## **Figures and Tables**

Page Number

### **Chapter 1. Introduction**

Figure 1.1	Illustration of the process of protein folding through an energy landscape	4
Figure 1.2	Typical CD spectra given by different types of secondary structure	11
Figure 1.3	Absorption and emission spectra of phenylalanine, tyrosine and tryptophan residues	12
Figure 1.4	Dendrogram of 491 protein kinase domains from 478 genes identified by Manning <i>et al</i> (2002)	21
Figure 1.5	Crystal structure of p38 $\alpha$	23
Figure 1.6	Activation pathway and key substrates of PKB	26

### **Chapter 2. Design, Testing and Evaluation of a Protein Refolding Screen using p38 $\alpha$ as a Model Protein Kinase**

Figure 2.1	Purity of p38 $\alpha$ inclusion body preparation assessed by reducing SDS-PAGE	43
Figure 2.2	Elution profile of native protein and inclusion body protein denatured with 8 M urea from analytical scale Superdex 200 column	44
Table 2.1	Layout of the initial protein kinase refolding screen	47
Figure 2.3	Example ePAGE gels indicating limit of detection for silver staining method and typical screen results	50
Figure 2.4	Capillary electrophoresis analysis of refolded p38 $\alpha$	52
Figure 2.5	Analytical size exclusion chromatography analysis of native folded and refolded p38 $\alpha$	54
Figure 2.6	Structure of UTDC used to detect concentration of correctly folded protein present in refolded samples.	55

Figure 2.7	Surface plasmon resonance analysis of refolded and native p38 $\alpha$ binding to immobilised TDC	56
Figure 2.8	Overview of the process of screening for conditions that enhance the refolding of p38 $\alpha$	57
Figure 2.9	Recoveries of soluble, monomeric and folded protein obtained when refolding inclusion bodies and soluble, denatured protein	59
Figure 2.10	Recovery of soluble, monomeric and folded protein when refolding inclusion body protein and denatured soluble protein in buffer alone	60
Figure 2.11	Thermal melting analysis of native p38 $\alpha$ in a series of buffers and pHs	62
Table 2.2	T <sub>m</sub> values for p38 $\alpha$ under various pH and buffer conditions	63
Table 2.3	Comparison of refolding recoveries of p38 $\alpha$ from screen runs and scaled-up experiments	64
Figure 2.12	Circular dichroism spectra of natively folded p38 $\alpha$ and inclusion body derived p38 $\alpha$ refolded in 0.1M CAPSO, 10 % Betaine, pH 9.5	65
Figure 2.13	Structure of SB202190 used to assess the activity of p38 $\alpha$ refolded on a 20 mg scale	66
Figure 2.14	Inhibition of the binding of refolded p38 $\alpha$ to immobilized UTDC by SB202190 detected with SPR	66

### **Chapter 3. Application of the Refolding Screen to Additional Protein Kinases and Refinement of the Screening Procedure**

Figure 3.1	Sequence alignment of the kinase domains of the members of the kinase panel	78
Figure 3.2	Crystal structures of the members of the kinase panel	79

Table 3.1	Selected physical properties of the members of the kinase constructs	80
Figure 3.3	BSRG1 digest of PhK minipreps following BP reaction	81
Figure 3.4	BSRG1 digest of PhK minipreps following LR reaction	83
Figure 3.5	Diagrammatic representation of the PhK gamma subunit kinase domain expression construct used.	84
Figure 3.6	Example screen readout data for AKT2 refolding screens	86
Figure 3.7	Recovery of refolded protein and response units obtained for the refolding of AKT2 in the kinase refolding screen	87
Figure 3.8	Recovery of refolded protein obtained for the refolding of AKT2 for conditions resulting in measurable recovery of monomeric protein	88
Figure 3.9	Correlation between monomeric protein recovery and response units from the SPR based folded protein assay for the refolding of AKT2	89
Figure 3.10	Example screen readout data for KIS refolding screens	91
Figure 3.11	Recovery of refolded protein and response units obtained for the refolding of KIS in the kinase refolding screen	92
Figure 3.12	Correlation between monomeric protein recovery and response units from the SPR based folded protein assay	93
Figure 3.13	Recovery and response units obtained for the refolding of KIS for successful conditions	94
Figure 3.14	Example screen readout data for PhK refolding screens	97
Figure 3.15	Recovery of refolded protein and response units obtained for the refolding of PhK in the kinase refolding screen	99
Figure 3.16	Correlation between monomeric protein recovery and response units from the SPR based folded protein assay for the refolding of PhK	100
Figure 3.17	Recovery and response units obtained for the refolding of PhK for successful conditions	101
Figure 3.18	Example screen readout data for TTK refolding screens	103
Figure 3.19	Recovery of refolded protein and response units obtained for the refolding of TTK in the kinase refolding screen	104

Figure 3.20	Correlation between monomeric protein recovery and response units from the SPR based folded protein assay for the refolding of TTK	105
Figure 3.21	Recovery and response units obtained for the refolding of TTK for successful conditions	106
Figure 3.22	Thermal unfolding transitions and $T_m$ s for TTK in various pH and buffer conditions	108
Table 3.2	$T_m$ values for TTK under various pH and buffer conditions	109
Table 3.3	Aliased model terms for the fraction factorial refolding screen	111
Table 3.4	Composition of the fractional factorial screen for protein kinase refolding	113
Figure 3.23	Refolding recoveries of the five kinases tested using the fractional factorial refolding screen	115
Table 3.5	Model F-values for the ANOVA analysis of the fractional factorial refolding screen of five protein kinases	117
Figure 3.24	Studentised residuals for the fit of fractional factorial screen model against actual data	118
Figure 3.25	Normal probability plot of the residuals for the fit of fractional factorial screen model against actual data	120
Table 3.6	Significant positive and negative factors in the refolding of AKT2 identified by the ANVOA of the results of a fractional factorial screen of AKT2 refolding	121
Table 3.7	Positive and negative factors identified as significant by ANOVA of the fractional factorial refolding screen performed on PhK	123
Table 3.8	Positive and negative factors identified as significant by ANOVA of the fractional factorial refolding screen performed on TTK	124
Table 3.9	Significant positive and negative factors identified in ANVOA of fractional factorial screens performed on five protein kinases	125

## Chapter 4. Mutagenesis, Expression and Characterisation of Human TTK tryptophan Mutants

Figure 4.1	Crystal structure of human TTK kinase domain with native tryptophan residues highlighted	130
Figure 4.2	Space filling representation of the local environment around tryptophan residues 612 and 622	131
Figure 4.3	Local environment of the core tryptophans of the kinase domains of p38 $\alpha$ and TTK	132
Figure 4.4	Diagrammatic representation of the TTK kinase domain expression construct used	138
Figure 4.5	The amino acid sequence of the cleaved human TTK kinase domain construct used	138
Figure 4.6	1% Agarose gel showing BSRG1 digest of TTK containing pDONOR221 plasmids produced by BP reaction	139
Table 4.1	Summary of the TTK tryptophan mutants created	140
Figure 4.7	A 1% agarose gel showing the double digestion of putative TTK <sup>W612F W622F</sup> with BstB1 and BSRG1	141
Figure 4.8	Purification of TTK <sup>WT</sup> by Immobilised Metal Affinity Chromatography	143
Figure 4.9	Purification of TTK by subtractive immobilised metal affinity Chromatography	145
Figure 4.10	Purification of TTK by cation exchange chromatography	146
Figure 4.11	ESI-TOF mass spectrometry of wild type TTK	148
Figure 4.12	ESI-TOF mass spectrometry of TTK <sup>W612F W622F</sup>	149
Figure 4.13	ESI-TOF mass spectrometry of TTK <sup>W612F W622F W628F</sup>	150
Table 4.2	Molecular mass of TTK and TTK tryptophan mutants as measured by ESI-MS and calculated from the sequence	151
Figure 4.14	CD spectra of TTK and tryptophan mutants	152
Table 4.3	Secondary structure content predictions for TTK and TTK tryptophan mutants	152



Figure 4.15	Thermal unfolding curve of TTK monitored by Sypro-Orange fluorescence	154
Figure 4.16	Thermal unfolding curves of TTK tryptophan mutants	155
Figure 4.17	Average $T_m$ for wild type TTK and TTK tryptophan mutants	156
Table 4.4	Mid points of thermal melting transitions for TTK kinase domain and tryptophan mutants	156
Figure 4.18	Tryptophan emission spectra of TTK and TTK tryptophan mutants	158
Table 4.5	$\lambda_{max}$ for different TTK constructs under native and denaturing conditions	159
Figure 4.19	Fitting of log normal distribution to fluorescence spectra to estimate parameters for curve	160
Table 4.6	Log normal distribution parameters for the fitting of Native and unfolded TTK <sup>W612F W622F W628F</sup> spectra to equation 4.1	160
Figure 4.20	Tryptophan fluorescence spectrum position-width analysis	161

## Chapter 5. Equilibrium Folding of Human TTK Protein Kinase Domain

Figure 5.1	Far-UV CD spectra of wild type TTK kinase domain at various denaturant concentrations	171
Figure 5.2	Unfolding and refolding of wild type TTK induced by GdnHCl and monitored by Far-UV CD	172
Figure 5.3	Fluorescence spectra of wild type TTK under conditions of various denaturant concentration	174
Figure 5.4	Refolding and unfolding of wild type TTK induced by GdnHCl and monitored by change in $\lambda_{max}$ of tryptophan fluorescence	176
Figure 5.5	Folding phase diagrams for wild type TTK	177
Figure 5.6	Overlay of Far-UV CD and tryptophan fluorescence monitoring of the folding of wild type TTK	179

Figure 5.7	Far-UV CD spectra of wild type TTK in native, intermediate and unfolded states	180
Figure 5.8	Tryptophan fluorescence spectra of wild type TTK in native, intermediate and unfolded states	181
Figure 5.9	Unfolding and refolding of TTK <sup>W612F W622F</sup> induced by GdnHCl and monitored by Far-UV CD	184
Figure 5.10	Fluorescence spectra of TTK <sup>W612F W622F</sup> under conditions of various denaturant concentration	186
Figure 5.11	Refolding and unfolding of TTK <sup>W612F W622F</sup> induced by GdnHCl and monitored by change in $\lambda_{\max}$ of tryptophan fluorescence	187
Figure 5.12	Folding phase diagrams for TTK <sup>W612F W622F</sup>	188
Figure 5.13	Overlay of Far-UV CD and tryptophan fluorescence monitoring of the folding of TTK <sup>W612F W622F</sup>	190
Figure 5.14	Far-UV CD spectra of TTK <sup>W612F W622F</sup> in native, intermediate and unfolded states	191
Figure 5.15	Tryptophan fluorescence spectra of TTK <sup>W612F W622F</sup> in native, intermediate and unfolded states	192
Figure 5.16	Unfolding and refolding of TTK <sup>W612F W622F W628F</sup> induced by GdnHCl and monitored by Far-UV CD	195
Figure 5.17	Fluorescence spectra of TTK <sup>W612F W622F W628F</sup> under conditions of various denaturant concentration	197
Figure 5.18	Refolding and unfolding of TTK <sup>W612F W622F W628F</sup> induced by GdnHCl and monitored by change in $\lambda_{\max}$ of tryptophan fluorescence	198
Figure 5.19	Folding phase diagrams for TTK <sup>W612F W622F W628F</sup>	199
Figure 5.20	Overlay of Far-UV CD and tryptophan fluorescence monitoring of the folding of TTK <sup>W612F W622F W628F</sup>	200
Figure 5.22	Far-UV CD spectra of TTK <sup>W612F W622F W628F</sup> in native, intermediate and unfolded states	202
Figure 5.22	Tryptophan fluorescence spectra of TTK <sup>W612F W622F W628F</sup> in native, intermediate and unfolded states	203

Table 5.1	Thermodynamic parameters determined for the refolding and unfolding of wild type TTK and TTK tryptophan mutants followed by far-UV CD analysis	205
Table 5.2	Thermodynamic parameters determined for the refolding and unfolding of wild type TTK and TTK tryptophan mutants followed by tryptophan fluorescence	205
Figure 5.23	Overlay of the folding transitions of wild-type TTK, TTK <sup>W612F W622F</sup> and TTK <sup>W612F W622F W628F</sup> measured by far-UV CD	207
Figure 5.24	Overlay of the folding transitions of wild-type TTK, TTK <sup>W612F W622F</sup> and TTK <sup>W612F W622F W628F</sup> measured by tryptophan fluorescence	209

## Chapter 6. General Discussion

Figure 6.1	Local environments of the core tryptophans of the kinase domain in the structures of the kinases of the kinase panel	221
------------	--	-----

## **Acknowledgements**

I would like to thank both of my supervisors on this project, Dr Teresa Pinheiro and Mr Rick Davies for their advice and support throughout this project. I would also like to thank the members past and present of the Structural Biology group at the University of Warwick, particularly Mr David Jenkins, Dr David Roper, Mr Phillip Robinson and Dr Narinder Sanghera for their help and instruction in the use of instruments and general advice. I would also like to thank all the members of the protein engineering section of Discovery Enabling Capabilities and Sciences at Alderley Park for their friendliness and support during this time. I would particularly like to thank Dr Gareth Davies, Mr Malcolm Anderson, Mr Paul Hawtin and Dr Kate Wickson for their instruction in the use of particular instruments and techniques.

I would also like to thank my immediate family who have supported me through my moves between Coventry and Macclesfield, and Mr John Vincent and Mrs Jen Vincent who generously allowed me to lodge with them for 6 months. Finally I would like to thank Tanya, without whose support, love, care and encouragement the writing of this thesis would not have been possible.

*Soli Deo Gloria!*

## **Declaration**

I hereby declare that the research submitted in this thesis was conducted by myself under the supervision of Dr Teresa Pinheiro at the Department of Biological Sciences, University of Warwick, and under the supervision of Mr Rick Davies at AstraZeneca, Alderley Park.

No part of this work has previously been submitted to be considered for a degree or other qualification. All sources of information have been specifically acknowledged in the form of references.

## **Summary**

The vast increase in the number of protein structures identified since the first high resolution protein structure was determined has shown that there are relatively few protein folds. The study of the folding of proteins has also expanded significantly since its inception. The contact-order of residues in the native structure has been implicated as important in folding. This has raised the question of whether the common folds observed in protein structures fold via common mechanisms.

The protein kinase domain is a large, pharmaceutically important and conserved protein fold, of which many examples fail to fold correctly when over-expressed as recombinant proteins in *Escherichia coli*. The kinase domain forms an excellent area in which to study the folding of a large conserved domain with varying sequence.

To study the refolding of the kinase fold a refolding screen was created, using p38 $\alpha$  as a model protein kinase. The results of this screen were compared to the results of refolding a further four protein kinases, AKT2, KIS, PhK and TTK to determine commonalities in the folding of protein kinases. The refolding of the five protein kinases was also examined using a fractional factorial screen which examined combinations of refolding additives. In the screening of the refolding of protein kinases no factors were identified which were common to the refolding of all five of the tested protein kinases.

The equilibrium folding of a single protein kinase, TTK, was also studied. The folding of TTK was determined to proceed via different pathways on folding and unfolding, with a co-operative unfolding pathway through a molten-globule intermediate, and a non-cooperative refolding pathway via an intermediate with different secondary structure content to the unfolding intermediate.

The difference between the folding properties of protein kinases determined in the screen and through the analysis of the equilibrium folding of TTK suggest that there may not be a common protein kinase folding mechanism.

## **Abbreviations**

AFM	Atomic force microscopy
ANOVA	Analysis of variation
ANS	1-anilino-8-naphthalene sulphonic acid
ATF-2	Activating transcription factor 2
ATP	Adenosine tri-phosphate
BSA	Bovine serum albumin
CAPSO	<i>N</i> -cyclohexyl-2-hydroxyl-3-aminopropanesulphonic acid
CD	Circular dichroism
CDK	Cyclin-dependent kinase
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulphonate
CHK2	Cell cycle checkpoint kinase 2
CK1	Casein kinase 1
CLK	CDC-like kinase
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDC	1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride
EDTA	Ethylene diamine tetra-acetic acid
EGTA	Ethylene glycol tetra-acetic acid
ESI-MS	Electrospray ionization mass spectrometry
ESI-TOF	Electrospray ionization time of flight
FGFR	Fibroblast growth factor receptor
FRET	Fluorescence resonance energy transfer
FT-IR	Fourier transform infrared
GdnHCl	Guanidine hydrochloride
GSK3	Glycogen synthase kinase 3

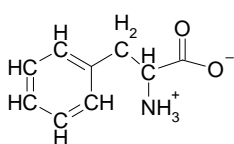
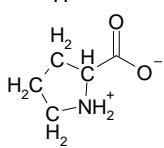
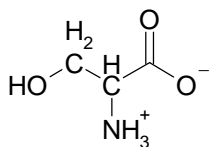
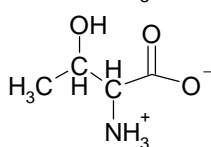
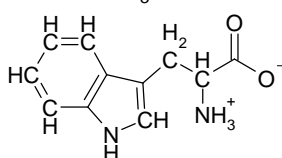
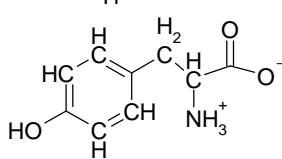
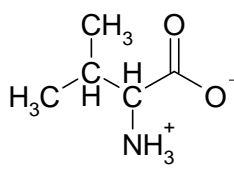
GST	Glutathione S-transferase
H/D exchange	Hydrogen / deuterium exchange
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulphonic acid
hFGF	Human fibroblast growth factor
h-GR	Human growth factor
HPLC	High-performance liquid chromatography
HSP	Heat shock protein
IFN	Interferon
KIS	Kinase interacting with stathmin
LC	Liquid chromatography
I	Intermediate
IMAC	Immobilised metal affinity chromatography
IPTG	isopropyl-beta-D-thiogalactopyranoside
MAP	Mitogen activated protein
MAPK	Mitogen activated protein kinase
MAPKAP2	MAPK activated-protein kinase-2
MES	2-( <i>N</i> -morpholino)ethanesulphonic acid
MKK3	mitogen-activated protein kinase kinase 3
MKK4	mitogen-activated protein kinase kinase 4
MKK6	mitogen-activated protein kinase kinase 6
MS	Mass spectrometry
Mps	Monopolar spindle
N	Native
NDSB	Non-detergent sulphobetaine
NHS	<i>N</i> -hydroxysulphosuccinimide
NiNTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance



OD	Optical density
PCR	Polymerase chain reaction
PDK	Phosphoinositide-dependent kinase
PEG	Polyethylene glycol
PH	Pleckstrin homology
PhK	Phosphorylase kinase
PI3K	Phosphoinositide 3-kinase
PIP3	Phosphatidylinositol (3,4,5)- trisphosphate
PKA	Protein kinase A
PKB	Protein kinase B
PKC	Protein kinase C
PKG	Protein kinase G
SDS	Sodium dodecylsulphate
SDS-PAGE	sodium dodecyl sulphate polyacryamide gel electrophoresis
SE	Standard error
siRNA	Small inhibitory RNA
SPR	Surface plasmon resonance
RNA	Ribonucleic acid
TDC	Target definition compound
TEV	Tobacco etch virus
TIM	Triosephosphate isomerase
TK	Tyrosine kinase
TKL	Tyrosine kinase like
T <sub>m</sub>	midpoint of thermal unfolding
TMAO	trimethylamine N-oxide
TrisHCl	Tris (hydroxymethyl) aminomethane hydrochloride
U	Unfolded
UTDC	Ureidoquinazoline target definition compound
UV	Ultra violet

Three letter and single letter amino acid abbreviations.

Amino Acid		Three letter code	Single letter code
Alanine		Ala	A
Arginine		Arg	R
Asparagine		Asn	N
Aspartic Acid		Asp	D
Cysteine		Cys	C
Glutamine		Gln	Q
Glutamic Acid		Glu	E
Glycine		Gly	G
Histidine		His	H
Isoleucine		Ile	I
Leucine		Leu	L
Lysine		Lys	K
Methionine		Met	M

Phenylalanine		Phe	F
Proline		Pro	P
Serine		Ser	S
Threonine		Thr	T
Tryptophan		Trp	W
Tyrosine		Tyr	Y
Valine		Val	V

---

# Chapter 1. Introduction

## 1.1 Protein Folding

The central dogma of molecular biology describes the process of the making of proteins, through a transcription of DNA into mRNA and the translation of this mRNA message into the primary sequence of the protein on the ribosome. Following the translation of the mRNA on the ribosome, the nascent polypeptide chain dissociates from the ribosome and then must adopt its final three dimensional shape.

When the process of the folding of a new protein is considered from the theoretical standpoint, the random search of the possible conformational space of the polypeptide chain would take an impossibly long time. If we assume that each residue in the protein can adopt two different conformations, then for a 100 residue protein, the total number of conformations that the protein chain could theoretically adopt is  $2^{100}$ , or  $1.27 \times 10^{30}$  conformations. If we assume that the protein takes 1 picosecond to sample each conformational state, then it would take, on average on the order of  $10^{18}$  seconds or approximately  $3 \times 10^{10}$  years for the protein to adopt the lowest energy state, termed the native state. As it is known that proteins of this size can fold in timescales of a millisecond or faster, it then follows that the folding of proteins must be directed in some fashion, and that proteins probably fold through specific pathways to the native state (Levinthal, 1968). The folding of bovine pancreatic ribonuclease was studied by Anfinsen (Anfinsen, 1973). The folding of the protein was studied *in vitro* and it was shown that the protein was capable of refolding from a denatured, oxidised state, in 8 M urea, to an active state with the correct disulphide bonds being formed. Given that the only component of the refolding reaction was the protein chain itself, it follows that the folding of the protein must be directed by the polypeptide chain itself, and not by other cellular factors.

Studies performed since this initial, key discovery have shown that the means by which the folding is directed by the sequence of the protein are diverse and complex. Local interactions between close and adjacent residues are important, but non-local interactions, between widely spaced residues are also important (Damaschun *et al.*, 1999; Navon *et al.*, 2001; Klein-Seetharaman *et al.*, 2002). The overall topology of the native protein (Baker, 2000), the position of hydrophobic residues (Dobson *et al.*,

1998) and disulphide bridges (Creighton, 2000) are also important factors in the folding of proteins.

### **1.1.1 Protein Folding and Common Protein Structures**

The developments in the areas of molecular biology and the biophysical analysis of proteins have led to a vast increase in the understanding of protein structure and in the number of protein structures known. The year of the writing of this thesis coincides with the 50<sup>th</sup> anniversary of the first high resolution protein structure, that of myoglobin in 1959 (Bodo *et al.*, 1959). The number of protein structures available in the Protein DataBank is now close to 60,000. In this vast expansion of the available data on the structure of proteins it has been noticed that proteins do not all adopt entirely different structures. Rather proteins are assembled from a number of smaller protein domains which have common folds despite a lack of conservation of the amino acid sequence. These families of folds include  $\beta$ -propellers and TIM barrels (Godzik, 1997). The number of protein folds has been estimated to be between 1000 (Wolf *et al.*, 2000; Wang, 1998) and 2700 (Xingsheng *et al.*, 2003).

With the observation that protein structures are conserved, the question has arisen as to the manner of folding of these conserved protein units. Some study has been performed to answer this question (Bueno *et al.*, 2006) the results of these experiments have been mixed. Some classes of proteins, notably topologically complex proteins such as knotted proteins (Gloss, 2007) have been observed to fold via common pathways, although this may be expected due to the complexity of the structure of these proteins; requiring the folding to proceed via the same pathway in order to form the topological knots in the structure (Mallam and Jackson, 2005). Studies on proteins with high sequence similarity have shown that the early events in folding are key to determining the structure arrived at (Scott and Daggett, 2007).

### **1.1.2 Models and Mechanisms of Protein Folding**

Several models have been proposed to describe how proteins can fold from an extended structure into a compact three dimensional structure based only on the information present in the sequence of the protein. Three major models have been proposed based upon the nature of the initial folding event. These models have been

suggested to explain the folding behaviour of different proteins, underlining that all proteins do not fold via a single, universal mechanism.

The first model proposes that a folding nucleus of a few local residues forms, and that the remaining structure of the protein forms around this nucleus. This model is termed the nucleation-growth model (Wetlaufer, 1973). From this model, a second model has been proposed which extends the folding nucleus to cover residues that are distributed throughout the protein. This is termed the nucleation-condensation model. The secondary structure then forms from the folding nucleus and is stabilized concurrent with its formation by tertiary interactions (Fersht, 1995a; Fersht, 1997; Fersht, 2000).

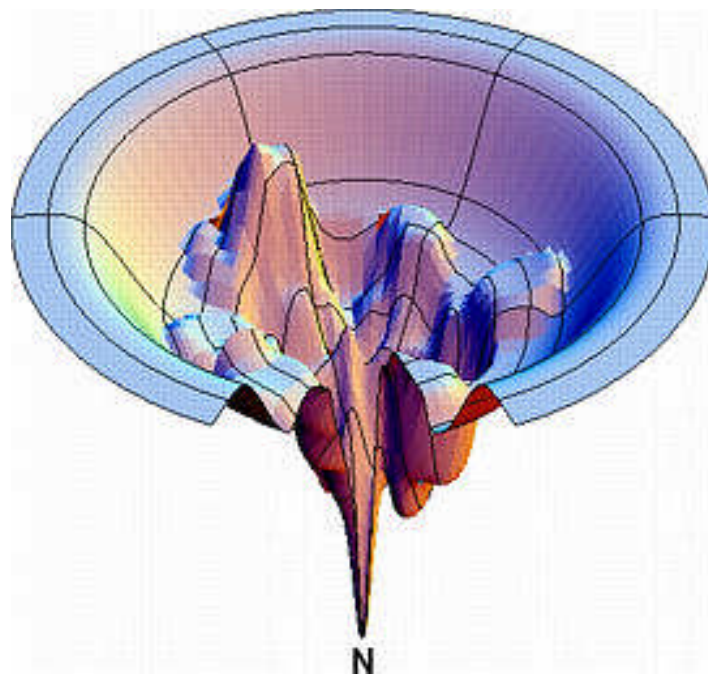
Most proteins contain a core of hydrophobic residues which are compacted together and exclude water from the core of the protein. It has been proposed that the formation of this hydrophobic core may be the important, initial event in the folding of many, particularly larger proteins. In this model the protein forms an initial collapsed state in which the hydrophobic residues are in the core (Chan *et al.*, 1998) and polar residues are on the surface of the protein and water is excluded from the core of the structure (Kim and Baldwin, 1990; Cheung *et al.*, 2002). Larger proteins tend to have more, or larger hydrophobic areas in their structures supporting the concept of the importance of hydrophobic collapse in their folding, although these proteins often have complex folding pathways, with several intermediate states which can accumulate (Dobson *et al.*, 1998).

The third model of protein folding stems from a recognition of the importance of protein secondary structure in the native state of proteins. This model suggests that a framework of secondary structure forms initially in different areas of the protein (Ptitsyn and Rashin, 1975). The individual secondary structure elements that have been formed then associate and are stabilized by tertiary interactions. The model is termed the framework model and has been suggested to describe the folding of staphylococcal nuclease (Wang *et al.*, 1995).

### **1.1.3 The Folding Energy Landscape**

Early work in the folding of proteins conceived of the folding pathways of proteins as narrow and defined, with folding intermediates and a few, limited, routes that the protein populated and adhered to in order to reach the native state. The protein folding process is now considered to be a much more stochastic process. The

fluctuations in the conformation of the protein required for the polypeptide chain mean that multiple polypeptide chains will not adopt the same conformations under the same conditions, but rather an ensemble of states will be present. The folding of the protein is then described as the polypeptide chain moving through an energy landscape from the unfolded protein which contains high levels of free energy and conformational diversity and freedom towards the most stable, lowest energy state, which is the native state (Vendruscolo *et al.*, 2003; Dobson, 2004). This concept is illustrated in Figure 1.1.



**Figure 1.1** Illustration of the process of protein folding through an energy landscape. The unfolded state is found at the top of the funnel. The position of the native state is indicated by “N” (Figure taken from Dill and Chan, 1997).

The topology of the energy landscape is not constant for all proteins but varies between different proteins. Small proteins, such as are commonly studied for folding tend not to populate folding intermediates and have a simple folding pathway. These proteins would have a smooth folding funnel, which has a single energy barrier between the native and the unfolded state. Proteins which have a more complex folding process, with multi-step pathways and multiple intermediates have a much more complex folding funnel, with local minima in multiple places which correspond to the folding intermediates identified (Dill and Chan, 1997; Gruebele, 2002).

This view of protein folding has been illustrated through slight variations in experimental conditions for the *in vitro* folding experiments of proteins and also by the results of single molecule folding experiments. Variations in the conditions under which folding experiments are performed, for example variations in the pH of folding, change the measured parameters of the *in vitro* folding experiments (Sato and Raleigh, 2002). Similarly, in single molecule folding experiments, the folding measured changes from experiment to experiment (Neinhaus, 2009).

#### **1.1.4 Protein Folding Intermediates**

Whether proteins fold through an energy landscape, or as viewed earlier in studies of protein folding, there may exist in the folding process a number of intermediates in the folding of the protein. Folding intermediates are structures, or ensembles of particular types of structures that accumulate at particular stages in the folding of a protein. They are commonly associated with the folding of larger proteins. The formation of these intermediates is important in the folding of these proteins as it reduces the amount of random search required to obtain the native state (Berg *et al.*, 2002). When proteins are folded through an equilibrium process, compact and stable intermediates have also been shown to be formed (Hughson *et al.*, 1990). The formation of protein folding intermediates is also important in the mis-folding of proteins, whether in protein folding diseases, or in the accumulation of proteins as inclusion bodies in bacteria. This mis-folding of proteins often involves folding intermediates which cannot subsequently fold into the native state. Such intermediates are called non-productive or off-pathway intermediates.

The models of protein folding outlined previously (Section 1.1.2) predict different forms of protein folding intermediates depending on the model. The framework model, suggests a stepwise formation of the native state, where folding units form and no rearrangement of the structure occurs before each step is stabilized. This model suggests two types of intermediate that form on folding of the protein. The first intermediate type forms quickly, in less than 10 ms. This intermediate consists of a partially folded secondary structure which fluctuates between folded and unfolded in local areas. The second intermediate forms slower, on the order of 0.1 to 1 s and contains secondary structure which is more stable than in the first intermediate (Ptitsyn *et al.*, 1995).



The hydrophobic collapse model of protein folding suggests an equilibrium folding intermediate that is formed as the unfolded protein in a random coil conformation collapses into a smaller restricted space. This type of intermediate is termed a “Molten Globule” intermediate (Ptitsyn, 1987; Ptitsyn *et al.*, 1995). The second of the kinetic intermediates formed in the framework model has been shown to be similar to this intermediate (Dolgikh *et al.*, 1981). The molten globule intermediate has a secondary structure which is equivalent to the native state of the protein, but the tertiary structure of the protein is dynamic. The dimensions of the intermediate state have been shown to be much lower than the random coil unfolded state, but not much greater than the native state (Kim and Baldwin, 1990). The formation of this intermediate is thought to be very important in the folding of proteins, since it allows the polypeptide chain to be provided with a close to native backbone, whilst allowing the structure to be dynamic and to adopt the native state (Ptitsyn, 1987).

Given the importance of the molten globule state, and the fact that it is a typical state for many proteins under slightly denaturing conditions (Ptitsyn, 1992), the precise nature of this intermediate state has been closely studied by several different techniques. The presence of the compact, denatured state has been inferred for many proteins from size exclusion chromatography experiments under slightly denaturing conditions (Dolgikh *et al.*, 1981; Ptitsyn *et al.*, 1995). The presence of native like secondary structure has been inferred for many proteins using far-UV CD as well as H/D exchange NMR. The definitional lack of native-like tertiary structure was first demonstrated with cytochrome c (Ougushi and Wada, 1983) through the lack of certain near-UV CD features and the chemical shift of some of the protons in the NMR spectra.

The most studied protein that shows a molten-globule folding intermediate is probably guinea pig  $\alpha$ -lactalbumin. The intermediate state observed is stable under denaturing conditions of ~1 M GdnHCl (Ikeguchi *et al.*, 1986), and shows the key features of a molten globule intermediate. H/D exchange NMR has been used to identify the residues that are involved in the formation of the native-like secondary structure present in the intermediate state. Twenty amides in the structure were found to be protected from exchange in the partially folded intermediate state. These residues were found to be in the hydrophobic core of the protein and in the  $\alpha$ -helical regions of the native conformation. Far-UV CD spectra do suggest that more

secondary structure is formed than was found to be protected from solvent exchange by H/D exchange NMR. This suggests that some of the native like secondary structure identified by CD analysis of the intermediate state may be transiently formed (Baum *et al.*, 1989; Chyan *et al.*, 1993).

Some evidence has also been presented for the formation of molten-globule like intermediates in the kinetic folding of certain proteins. The folding of apomyoglobin was studied using far-UV CD and showed that a molten-globule like intermediate was present during the folding of the protein. A similar intermediate state to that discovered in equilibrium folding experiments was observed, containing the A, G and H helices. This intermediate is formed rapidly, in less than 5 ms (Jennings and Wright, 1993). Fragments of the whole protein were used to confirm the identity of the regions thought to be formed in the intermediate (Waltho *et al.*, 1993; Shin *et al.*, 1993a; Shin *et al.*, 1993b). These studies confirmed the role of the A, G and H helices in the early folding of apomyoglobin.

On the basis of this evidence, and the common finding of molten globule like intermediates in protein folding experiments, it has been proposed that this intermediate may be considered to be a general intermediate for protein folding (Ptitsyn *et al.*, 1990). It is also thought that the intermediate may be important in cellular processes that require partial unfolding of proteins, such as cross membrane transport (Pinheiro, 1994; Ren *et al.*, 1999; Matouschek, 2003) and chaperone interactions (Robinson *et al.*, 1994; Radford and Dobson, 1995).

### **1.1.5 Techniques for Characterizing the Folding Pathway**

To determine the nature of any folding intermediates it is necessary to examine the folding of the protein using a series of techniques to access different information on the folding pathway. Traditionally folding studies have been performed on a large number of protein molecules, which during the folding process adopt an ensemble of structures. More recent developments have allowed the study of the folding of single protein molecules increasing the understanding of the unfolding and refolding of complex proteins.

The study of the folding of an ensemble of molecules averages the folding of large number of molecules at once and averages the folded state of the ensemble of protein structures present under the conditions used. The folding of the ensemble of states can be examined in two different manners depending on what information is

required on the folding of the protein. The thermodynamics and the kinetics of the folding of proteins require different techniques to access. The thermodynamics of the folding of proteins is accessed using equilibrium folding experiments. In these experiments the protein being observed is perturbed from its initial conditions by a denaturing factor, such as temperature, pH or the presence of a chemical denaturant, and allowed to reach equilibrium under the new conditions. The extent of denaturation or renaturation (if starting with denatured protein) is then measured using one or more folding probes. The process is then repeated until the protein is fully denatured or renatured. The free energy of folding can then be determined from the behaviour of the folding probes.

Once the thermodynamics of folding have been established, it is then possible to establish the kinetics of folding. The speed of folding varies dramatically between different proteins and as a result a number of different techniques are required to follow the folding of proteins. Larger proteins, however, tend to fold more slowly than smaller proteins. Due to the complexity of their structures, or a dependence on slow changes such as cis-trans proline isomerisation (Wedemeyer *et al.*, 2002) in their folding, some proteins fold very slowly, for example collagen III (Engel and Bächinger, 2000). With these proteins it is possible to use simple instrumentation with a manual mixing of the protein and the buffer used to refold or unfold the protein. Similar folding probes are used with this type of folding experiment as are used in equilibrium folding experiments, since the time scales involved are fairly long.

For proteins which fold faster, on a timescale of seconds to milliseconds a more specialized technique is required. The mixing of the buffers required for the experiments must be performed much faster than is possible with manual mixing. This requires specialized mixers and relatively fast flow rates for the buffers used. The technique of stopped-flow is used to achieve these. Depending on the instrument used, the dead-time, that is the time during which data cannot be collected, may be as low as ~0.5 ms but is more commonly in the range of 2-3 ms. Some proteins fold faster than this dead-time. A reduction in temperature can be used to slow the folding of the protein, so that it can be captured. If the folding of the protein is still too fast for stopped-flow to capture, or a major phase of the folding is identified which is too fast to be captured, then further techniques must be used which allow for lower dead times.

Several techniques have been developed for the study of folding events which are too fast to be accessed by stopped-flow methods. The methods that have been adopted depend on the method of denaturation used. When the protein being studied has been denatured either by chemical denaturants, or by a change in the pH, then the technique of continuous-flow is used. This technique is similar to the stopped-flow technique, but the flow is allowed to continue through the observation channel during the observation, and the flow channels and mixers are smaller than those used for stopped-flow, with the dead-times being correspondingly lower. With pressure, or temperature being used to denature the protein, the methods of P-jump or T-jump have been adopted. These techniques have been used to study the fast folding of several proteins including prions (Jenkins *et al.*, 2009) and ribonuclease A (Font *et al.*, 2006). P-jump and T-jump measurements have a distinct advantage over continuous-flow methods in that they require much less protein and one can perform refolding and unfolding experiments on the same samples.

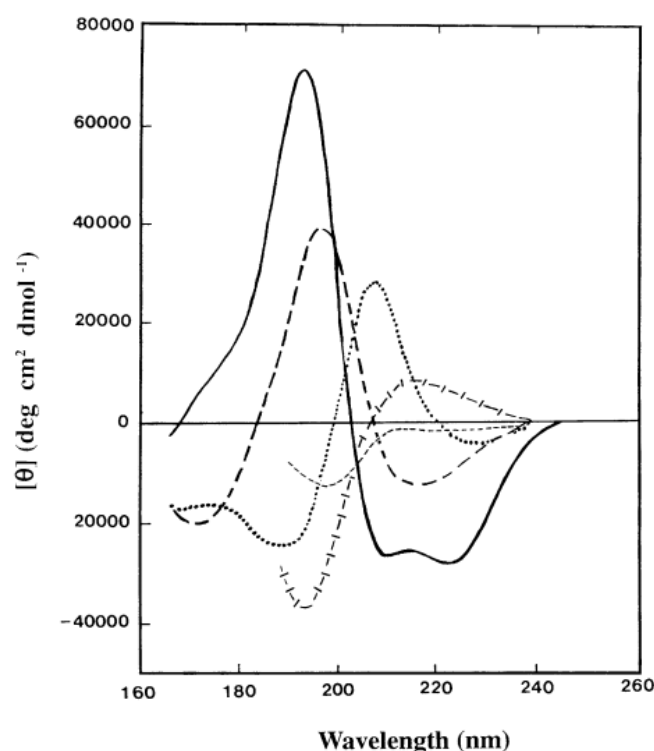
Single molecules studies are valuable in directly observing the folding of individual molecules. This technique uses a different method of unfolding the target protein. The protein being studied is unfolded by placing a force on the protein molecule to pull apart the structure. This technique is of particular value in studying the folding of large repeat proteins, such as the consensus ankyrin repeat protein N6C (Li *et al.*, 2006). The individual repeated domains of these proteins have often been studied by the techniques outlined above. Single molecule studies are particularly valuable in determining the extent of cooperativity between the unfolding and refolding of the repeated domains. Both optical traps and atomic force microscopy (AFM) have been used to capture and apply force to tested proteins. The information derived on the folding of proteins is less detailed than the information that is obtained with the techniques outlined above, although single molecule techniques do have several distinct advantages over equilibrium and kinetic folding measurements. Firstly the folding of a single molecule is directly being examined as opposed to a large number of molecules adopting an ensemble of conformational states. Secondly, the study of large repeat proteins is more easily performed than is the case with methods of the study of folding which involve large numbers of molecules. This is because the unfolding and refolding is directly observed through changes in the extension of the protein relative to the force applied, as opposed to the techniques used to observe the folding of proteins in other techniques, which may

require the modification of the protein to introduce reporter molecules or remove extra reporters from the protein. The technique is not sensitive to the concentration of protein used, which is advantageous compared to other techniques which show a concentration sensitivity for both successful folding and for their folding probes.

#### ***1.1.6 Folding Probes***

The study of the folding of proteins, through the methods outlined above, requires a probe to be used to report on the folding of the protein. This probe should describe the structure of the protein, or be sensitive to some feature of the protein, such as the distance between two areas of the protein. Combinations of different folding probes can be used to expand the amount of information obtained from the study of the folding of proteins.

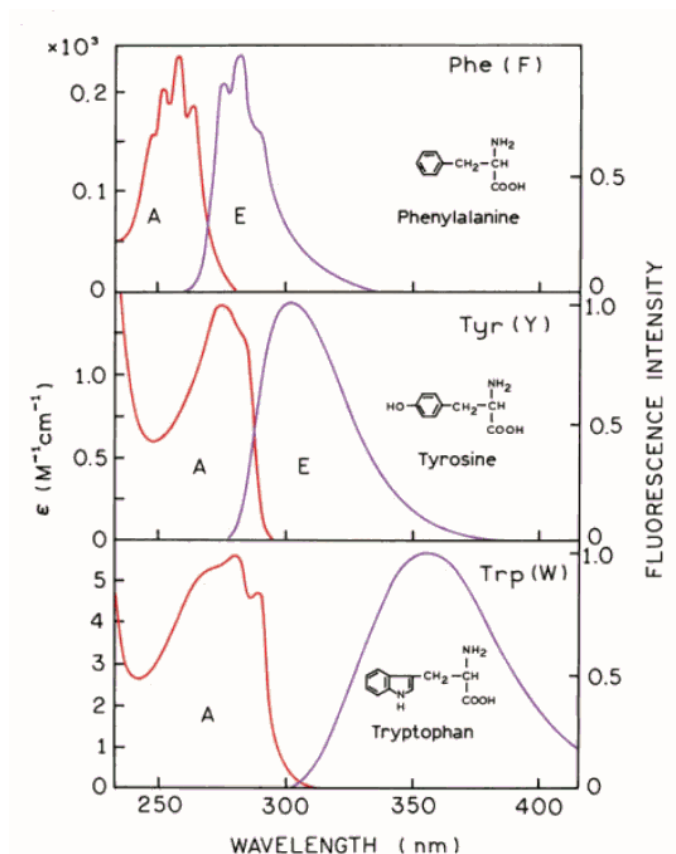
Features of the native protein can be exploited to probe the folding of proteins. Circular dichroism (CD) measures the difference between the absorption of left and right circularly polarized light. This difference is responsive to the secondary structure content of the protein. Different forms of secondary structure give characteristic spectra when the CD signal in the far-UV region is observed. These typical spectra are shown in figure 1.2.



**Figure 1.2:** Typical CD spectra given by different types of secondary structure. Solid Line  $\alpha$ -helix, long dashed line – anti-parallel  $\beta$ -sheet, dotted line – type I  $\beta$ -turn, cross-dashed line – extended  $3_1$ -helix poly(Pro) II helix, short dashed line – irregular structure. Figure adapted from Kelly *et al.* (2005).

The CD signal of proteins is dominated by the contribution to the signal from the  $\alpha$ -helical component at longer wavelengths of over 200 nm. The contribution of  $\beta$ -sheet is more dominant at wavelengths of under 200 nm. In folding experiments however, the high concentrations of denaturants or buffer components used to denature the protein interfere with the CD spectra at shorter wavelengths, causing a loss of information at these wavelengths. When temperature is used to denature the protein this loss of CD signal does not occur. Studies of the folding of proteins, therefore usually focus on the  $\alpha$ -helical component of the native protein.

The intrinsic fluorescence of proteins is also commonly used for studying their folding. Three of the twenty amino acids found in proteins are fluorescent, Phenylalanine, tyrosine and tryptophan. These residues fluoresce when excited with light at a wavelength in the near-UV region. The excitation and emission spectra of these residues are shown in Figure 1.3.



**Figure 1.3:** Absorption (A) and emission (E) spectra of phenylalanine, tyrosine and tryptophan residues. Adapted from Lakowicz (2006).

The differences in the excitation spectra of the three fluorescent residues allows different residues to be monitored by fluorescence. The fluorescence of the whole protein is monitored by excitation of these residues at 280 nm. This has the advantage of having a high fluorescence signal which allows the use of low protein concentrations in folding experiments. The fluorescence spectra obtained in this manner are complex, however, containing the contributions of many different fluorescent amino acids. The fluorescence of tryptophan residues, as seen in Figure 1.3, can be selectively excited by utilizing a long excitation wavelength, which minimizes or eliminates the contribution from other fluorescent residues. The fluorescence of tryptophan residues is responsive to their environment. When in a solvated environment the quantum yield of tryptophan fluorescence is reduced and the  $\lambda_{\text{max}}$  of fluorescence is shifted towards a longer wavelength compared to when the residue is found in a non-polar environment. Since tryptophan residues are commonly found in the hydrophobic core of proteins, then tryptophan residues form an excellent probe of protein folding. Tryptophan fluorescence is sometimes quenched by hydrogen bonds formed between the NH group in the indole ring of the

tryptophan side chain and the aromatic ring of phenylalanine residues (Nanda *et al.*, 2000). The release or re-acquisition of this quenching can also be used as a probe of the folding of the protein, in combination with the changes in tryptophan fluorescence on solvent exposure. The technique of tryptophan fluorescence is often combined with circular dichroism in the study of the folding of proteins, for example in the study of the folding of p38 $\alpha$  (Davies, 2004).

Modifications of proteins can be used to report directly on the folding of particular regions of proteins, particularly those that are difficult to access with other techniques. The technique of Fluorescence Resonance Energy Transfer (FRET). This requires the modification of the protein being studied to include two fluorophores, a donor and an acceptor. The donor fluorophore is excited by the incident light and transfers the energy to the acceptor fluorophores, by a resonance transfer, which then fluoresces. The fluorescence emission from such a protein, at the acceptor emission wavelength, when excited at the donor wavelength is inversely proportional to the distance between the donor and acceptor molecules raised to the sixth power (Stryer and Haugland, 1967). This proportionality can be used to examine the distance between the donor and acceptor molecules as a protein unfolds or refolds, assuming that the relative orientation of the fluorophores remains the same. This technique was used by Huang *et al.*, (2008) to study the folding of Engrailed homeodomain to imply the existence of an equilibrium folding intermediate that was populated at low concentrations of denaturant. Recent developments of this technique have allowed the deconvolution of the FRET signal into different populations of molecules with differing distances between the donor and acceptor molecules. This has given direct evidence for the number of states in the folding transitions, as shown by studies on chymotrypsin inhibitor 2 (Deniz *et al.*, 2000).

Probes that report indirectly on the folded state of the protein are used when more specific probes cannot be used for various reasons. Techniques involving mass spectrometry report indirectly on the folding state of the protein. Ivernizzi and Grandori (2007) used changes in the charge state distribution to report on the folding of  $\beta$ -lactoglobulin. The changes were used to identify the unfolding of the protein. Mass spectrometry can also be used to measure changes in the mass of a protein as a result of hydrogen deuterium exchange. The number of exchangeable protons changes as a protein unfolds, with more protons becoming exchangeable as the protein unfolds, as their backbone amides become exposed to the solvent. The



technique measures the average mass change as a result of exchange at different times or under conditions of partial denaturation and compares this to the change in mass of the native state and denatured state on exchange (Maier and Deinzer, 2005). The changes in mass cannot be mapped onto the residues of the protein, unlike the changes observed using H/D exchange NMR. Such a technique has been applied to bovine ubiquitin and lysozyme (Katta and Chait, 1993).

Other probes of folding that have been used include light scattering from the protein sample, both static light scattering and dynamic light scattering. Dynamic light scattering allows the estimation of the hydrodynamic radius of the protein. This technique is particularly applicable to equilibrium measurements of protein folding. This technique does, however, require high protein concentrations compared to other techniques. H/D exchange NMR is a powerful technique, capable of not only counting the number of exchangeable protons as with H/D exchange MS, but the exchangeable protons can be mapped onto the protein structure. This provides key information on the nature of any folding intermediates, as noted earlier in the studies on the folding of  $\alpha$ -lactalbumin, where the extent of structure present in the intermediate was shown utilizing H/D exchange NMR.

## ***1.2 Protein Refolding***

The expression of proteins in a host cell has become the key technique which has driven vast increases in the understanding of protein structure and enzyme kinetics and mechanisms, through the availability of large amounts of soluble proteins from easy to culture organisms. Although the information necessary for a protein to fold into its native conformation is present in the primary structure of proteins (Anfinsen, 1973), the over-expression of recombinant proteins in *E. coli* often results in the accumulation of the recombinant protein in structures known as inclusion bodies, which contain a highly concentrated, non-soluble form of the protein. To generate soluble protein for the study of protein structure, folding and enzyme kinetics it would be desirable to obtain soluble, correctly folded protein from inclusion bodies, since they represent a high purity, easy to isolate and relatively inexpensive source of recombinant protein. To do this, it is necessary to disaggregate the inclusion bodies into separated protein chains, and then induce these chains to adopt the correct conformation. This is a difficult problem and has been

solved by different methods for different proteins. To understand the difficulties in this procedure we must first understand how the inclusion bodies are formed, what is contained in inclusion bodies as well as their nature.

### **1.2.1 The Formation of Bacterial Inclusion Bodies**

Bacterial inclusion bodies are formed on the over-expression of some recombinant proteins in *E. coli*. Inclusion bodies can also be formed naturally under certain forms of cell stress, e.g. heat shock, but these types of inclusion bodies are not considered here. The formation of inclusion bodies for these proteins is dependent on the temperature of expression, or on the rate of expression, as determined by the concentration of the inducer of expression. Inclusion bodies can be visualized in *E. coli* cells by light microscopy. Inclusion bodies are refractile under phase contrast microscopy and grow continuously during the expression of the recombinant protein (Carrió *et al.*, 1998). Analysis of the size of inclusion bodies, and the amount of protein accumulating in the inclusion bodies indicates that the density of protein in the inclusion bodies does not change through the process of the building of inclusion bodies. That is, the conformation of the protein in the inclusion bodies is the same regardless of the size of the individual inclusion bodies or the length of the expression of the recombinant protein (Carrió *et al.*, 1998).

The bacterial inclusion bodies that are formed on the over-expression of recombinant proteins often contain more additional proteins as well as the recombinant protein, reducing the purity of the inclusion bodies. Several host proteins have been shown to be present in purified inclusion bodies in addition to the recombinant protein. The building of inclusion bodies of hFGF-2 includes the elongation factor Tu (EF-Tu). The chaperones DnaK and DnaJ were also found in hFGF-2 inclusion bodies in addition to certain metabolic enzymes from *E. coli* namely LpdA, GatY and TnaA (Rinas *et al.*, 2007). GroEL/ES has also been found in some bacterial inclusion bodies. The recombinant protein usually, however, accounts for more than 80% of the total protein in the inclusion bodies (Carrió *et al.*, 1998), making inclusion bodies a high purity source of recombinant protein.

In early studies of bacterial inclusion bodies, it was assumed that the inclusion bodies represented a disordered aggregate of mis-folded proteins. However, more recent studies on inclusion bodies have revealed that this view of bacterial inclusion bodies is misguided. Ami *et al.* (2006) studied the structure of the

proteins present in bacterial inclusion bodies using Fourier Transform Infrared (FT-IR) spectroscopy. In studying the secondary structure of human growth hormone (h-GH) and human interferon- $\alpha$  2b (IFN $\alpha$ -2b) inclusion bodies it was determined that a significant amount of secondary structure was present. This secondary structure resembled the native secondary structure. However, intermolecular  $\beta$ -sheet bridges were also present, which are responsible for the insoluble and densely packed nature of the inclusion bodies. The secondary structure component of the purified inclusion bodies was also different from the secondary structure component of thermal aggregates of the proteins studied. Other studies on bacterial inclusion bodies have established that the inclusion bodies of fluorescent proteins can fluoresce, indicating that these inclusion bodies contain proteins that are mostly folded (Garcia-Frutos *et al.*, 2005). In addition, inclusion bodies of over expressed enzymes have been shown to retain enzymatic activity (Tokatlidis *et al.*, 1991).

Bacterial inclusion bodies are often considered to be inert protein aggregates that grow by accumulation of mis-folded protein and may shrink by proteolysis of the inclusion body proteins. Carrió and Villaverde (2000) showed, however, that bacterial inclusion bodies are dynamic. Exploiting an expression system using a temperature sensitive phage repressor (cI857 phage repressor) to control expression they were able to express  $\beta$ -galactosidase and P22-tailspike polypeptide as inclusion bodies and subsequently halt expression whilst continuing the culture by means of a temperature shift. When this was performed, the amount of soluble protein increased, despite no further expression of the protein being performed. This indicates that the inclusion bodies formed were being disintegrated by cellular machinery and protein refolded.

This has given rise to a different view of inclusion bodies proposed by Ventura and Villaverde (2006) and Villaverde and Carrió (2003). This view understands inclusion bodies to be dynamic structures with the protein contained being conformationally diverse. They propose that inclusion bodies do not contain only mis-folded protein which is trapped in its mis-folded state, but rather, are transient reservoirs of protein which is under the quality control mechanisms of the bacterial cell, namely chaperones and proteases. The accumulation of proteins in inclusion bodies is driven by the balance between the translation of the protein from the mRNA and the speed and chaperone requirements of the folding of the polypeptide chain. If the rate of production of the protein exceeds the ability of the

cell to fold the protein, the excess partially folded protein will be stored in inclusion bodies to be folded later, when the production of the protein has lowered or ceased. There are no defined sequence features that lead to the accumulation of protein in inclusion bodies, as evidenced by similar mutations in different proteins resulting in different behaviour of the proteins regarding their accumulation as inclusion bodies.

### **1.2.2 The Refolding Problem**

The accumulation of recombinant protein as inclusion bodies presents an easy to access source of high purity protein. To make use of this source of protein however, the protein chains must be separated, and the protein induced to fold into its native conformation. Many proteins have been refolded from inclusion bodies for different purposes. Owen *et al.*, (1995) refolded the kinase domain of phosphorylase kinase from inclusion bodies, and crystallised the refolded protein to obtain the structure of the phosphorylase kinase domain.

Some proteins for therapeutic use are produced by refolding bacterial inclusion bodies, for example heterodimeric platelet-derived growth factor (Müller and Rinas, 1999). When refolding inclusion bodies on a larger scale like that used for the production of therapeutic proteins, studies by Mannall *et al.* (2007) have confirmed that the rate of dilution of the inclusion bodies from the unfolded state into the renaturation conditions used is key for obtaining high yields of refolded protein.

A survey of the renaturation conditions that have been reported in the REFOLD database (Chow *et al.*, 2006), indicates that the conditions which support the refolding of different proteins vary widely. This has lead to the adoption of refolding screens to rapidly identify the conditions that promote the refolding of the protein. Commercial screens have also been produced, such as the Novagen iFold screen. Commercial refolding additives have also been produced, with the aim of improving the refolding of most proteins from inclusion bodies, such as the novexin reagents. These refolding screens are, however, usually broad, including several different reductant, and also including redox couples, such as reduced and oxidized glutathione. Since the commercial screens include four or more different pHs, and both reducing conditions and redox couples, the number of additives that is examined is necessarily lower. The typical refolding volume is also low, usually 1 mL, leading to low amounts of refolded protein. Since the amount of refolded

protein is low the manufacturers of these screens usually recommend that a sensitive activity assay is used to assess the extent of refolding. This is problematic for kinases however, which typically require activation by an activating kinase, or other regulatory partners for full activity. The combination of the requirement for an activity assay which is difficult for kinases in this format, the low coverage of the different types of additives, and the low amounts of refolded protein available for readout from the screen make these commercial screens unsuited to the identification of conditions for the refolding of protein kinases.

### ***1.2.3 The Preparative Refolding of Proteins***

Methods for supporting the refolding of proteins from inclusion bodies are widely variable. Firstly the methods employed for the disaggregating of the inclusion bodies are different depending on the nature of the protein that is being refolded. Inclusion bodies can be disaggregated by the use of chaotropic denaturants such as guanidine or urea or detergents such as SDS. Physical means of disaggregating inclusion bodies such as temperature or pressure can also be used, although these methods are usually used with small quantities of protein and are difficult to scale to larger refolding experiments.

Once the inclusion bodies have been disaggregated, the denatured protein produced may be purified in the denatured state if a purification tag is present which functions in the denatured state. Typically poly-histidine tags are used. Physical means of disaggregating inclusion bodies do not allow for the purification of the protein of interest from the contaminant proteins found in disaggregated inclusion bodies. The use of high purity denatured protein for refolding experiments is expected to improve the recovery of correctly folded protein (Batas *et al.*, 1999).

To refold the protein from the disaggregated state the conditions which maintain the denatured, disaggregated state need to be removed to allow the protein to refold. When physical means have been used to disaggregate the inclusion bodies, the protein is refolded by the removal of the pressure or temperature change, in either a single step, in a series of steps or in a slow gradient. Where chemical denaturation has been used several strategies have been employed to allow and support the refolding of the disaggregated inclusion body proteins. Where a poly-histidine tag is present on the protein to be refolded, the denatured protein can be bound to an IMAC column and the chemical means of denaturation removed by

changing the buffer the column is equilibrated in. This method was used by Ryu *et al.* (2008) to obtain refolded extracellular superoxide dismutase. A similar strategy can be employed using size exclusion columns equilibrated with a gradient of the denaturant used. This technique has been employed to obtain  $\beta$ -lactamase and B lymphocyte stimulator from inclusion bodies (Harrowing and Chaudhuri, 2003; Cao *et al.*, 2005).

The most common means of refolding the denatured protein is the rapid dilution of denatured protein into buffer. The buffer conditions, and the presence and concentration of a number of refolding additives, both chemical, such as arginine (Tsumoto *et al.*, 2005) and poly-ethylene glycol and protein based, such as molecular chaperones (Wang *et al.*, 1999) are used to support the refolding of the target protein. Such additives can be used with the column approaches outlined above, but the larger amounts required usually rule out such an approach for cost reasons.

The final approach used for the refolding of inclusion body proteins is the removal of the disaggregating agent by dialysis. This strategy cannot be used when the disaggregating agent is a large molecule, such as a detergent, and so is usually used with chemical denaturants like guanidine and urea. The dialysis buffer used may contain refolding additives like those used in rapid dilution. This approach is typically attempted after the rapid dilution approach has failed, since the removal of denaturant is slower and therefore refolding is slower. The disadvantage of this approach to the refolding of proteins is that it is slower and when scaled to generate quantities of protein suitable for biophysical analysis may require several dialysis steps to completely remove the denaturant. This approach was used to generate adenylate kinase from inclusion bodies by Hibino *et al.*, (1994).

## **1.3 Protein Kinases**

### **1.3.1 The Human Kinome**

Protein kinases are a diverse group of proteins involved in many signalling pathways and cellular processes. Protein kinases are ATP-dependent phosphotransferase enzymes which transfer a single phosphate group from ATP to the side chains of serine, threonine or tyrosine residues. Other kinases transfer phosphate groups from ATP to histidine residues, but these form a separate group of enzymes and are outside the scope of this study. The protein kinases have been

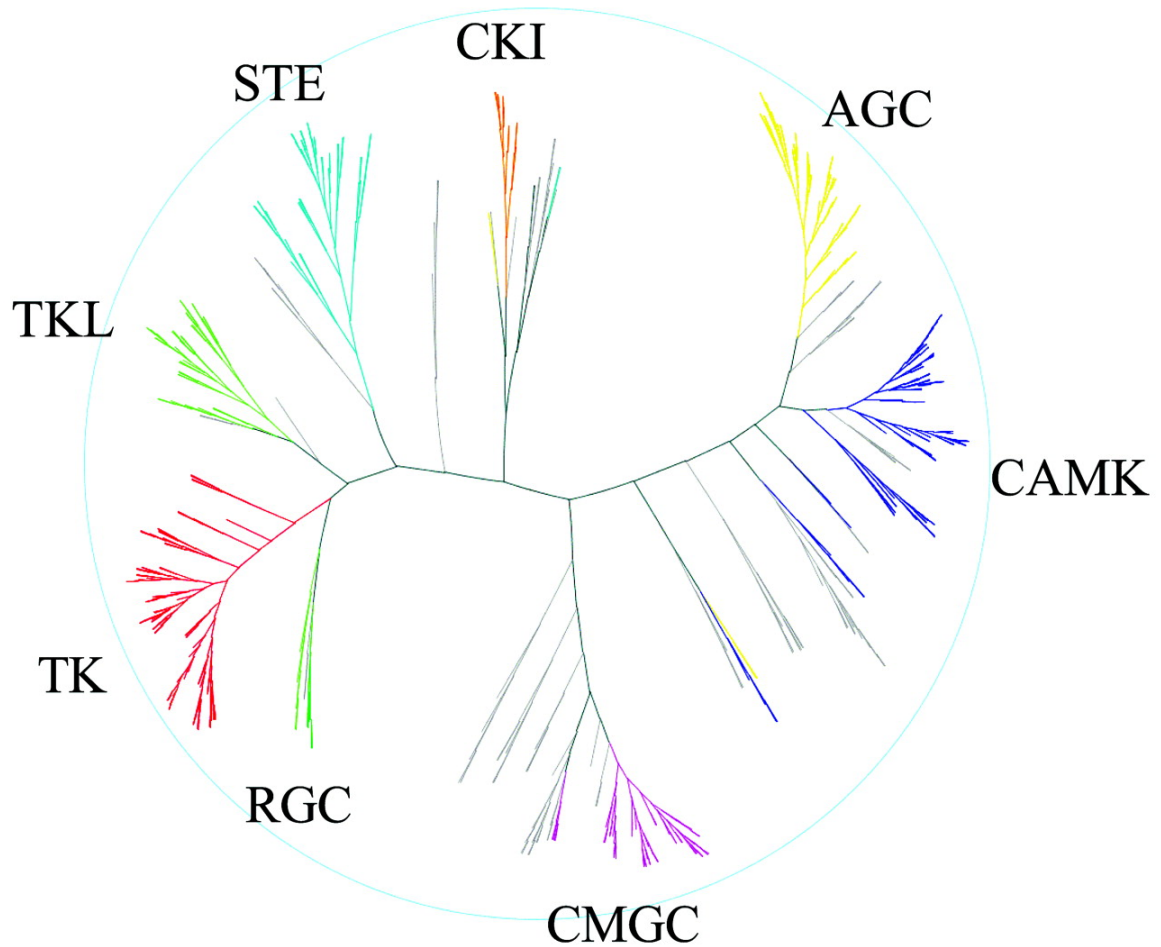
thought to require a divalent metal ion to allow the binding of ATP to the enzyme and the phosphoryltransfer activity (Adams, 2001), however this requirement has recently been questioned by studies on pseudokinases, which show kinase activity without a second  $Mg^{2+}$  ion thought to be essential (Kannan and Taylor, 2008).

Mammalian protein kinases are divided into three classes, depending on the residue which the kinase transfers phosphate groups to. Serine/threonine protein kinases transfer phosphate groups onto serine or threonine residues and tyrosine kinases transfer phosphates to tyrosine residues. A small number of kinases, termed dual-specificity kinases, is capable of transferring phosphate groups to both serine/threonine and tyrosine residues. Protein kinases only poorly phosphorylate free amino acids and instead use residues local to the phosphorylated residues to increase their affinity for their target (Adams, 2001).

Sequence comparisons of protein kinases has shown that these protein kinases shared a common region of 200-250 amino acids that conferred a kinase activity to these proteins. From this study a number of key regions and strictly conserved residues were identified (Adams, 2001). The protein kinase component of the human genome has subsequently been identified using a hidden Markov model of the eukaryotic protein kinase domain (Manning *et al.*, 2002). This consists of 478 protein kinases and 40 atypical protein kinases. Atypical protein kinases are defined as proteins which have kinase activity, but lack sequence similarity to the protein kinase domain. The human kinome contains 50 inactive protein kinase domains, termed pseudokinases, which lack one of the conserved catalytic residues of the kinase domain. It was expected, when found that these domains would be inactive and would function as scaffolds for other proteins (Manning *et al.*, 2002); however recently several pseudokinases have been shown to have catalytic activity and to bind ATP, albeit under slightly unusual conditions (Kannan and Taylor, 2008).

Sequence analysis of the human kinome has identified seven distinct groupings in the human kinome. These seven groups are the tyrosine kinase group; the tyrosine kinase-like group; the STE group (homologues of yeast sterile 7, 11 and 20 kinases); the CMGC group which contains the CDK, MAPK, GSK3 and CLK families; the CAMK group containing calcium and calmodulin dependent kinases; the AGC group containing the PKA, PKG and PKC families; and finally the CK1 group containing Casein Kinase 1 and closely related kinases. There are other

kinases in the kinome which do not fit into one of the defined groups (Manning *et al.*, 2002).



**Figure 1.4:** Dendrogram of 491 protein kinase domains from 478 genes identified by Manning *et al.* (2002). Conserved groups described above are shown in colour. Figure taken from Manning *et al.* (2002). The dendrogram was produced using a combination of techniques. The initial branching pattern was built from a neighbor-joining tree derived from a ClustalW protein sequence alignment of the human protein kinase domains. This was extensively modified by reference to other alignment and tree-building methods (hmmalign and parsimony trees) and by extensive pairwise sequence alignment of kinase domains (Manning *et al.*, 2002). The abbreviations of the kinase groupings are as follows: TK is the tyrosine kinase group; TKL the tyrosine kinase-like group; RGC is the receptor guanylate cyclase group; the STE group (homologues of yeast sterile 7, 11 and 20 kinases); the CMGC group which contains the CDK, MAPK, GSK3 and CLK families; the CAMK group containing calcium and calmodulin dependent kinases; the AGC group containing the PKA, PKG and PKC families; and the CK1 group containing Casein Kinase 1 and closely related kinases.

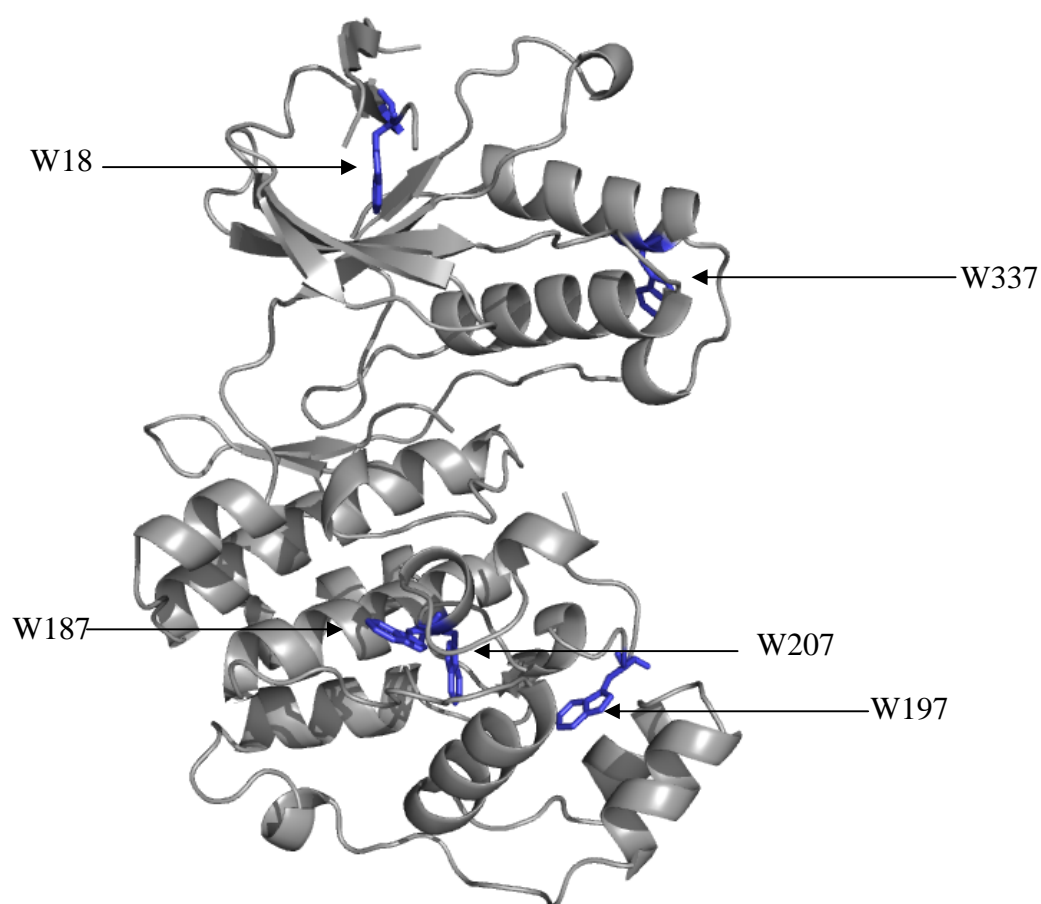
In this study of the refolding of protein kinases, five protein kinases in particular have been studied. These five kinases were p38 $\alpha$ , AKT2, KIS, PhK and TTK. A brief review of the biological role and any folding details available on these five kinases will now be given.



### 1.3.2 *p38 $\alpha$ , a Previously Studied Kinase*

p38 $\alpha$  is a member of the CAMK group of kinases in the human kinome. It is a 360 residue protein of 41 kDa which comprises a single kinase domain. p38 $\alpha$  is an effector kinase in the MAP signalling cascade, being activated by phosphorylation on residues Thr180 and Tyr182 (Raingeaud *et al.*, 1995) by upstream MAP kinase kinases, typically MKK6, MKK4 and MKK3 (Enslen *et al.*, 1998). p38 $\alpha$  phosphorylates downstream targets such as ATF2 and MAPKAP2 (Raingeaud *et al.*, 1996; Rouse *et al.*, 1994).

The equilibrium folding of p38 $\alpha$  has been studied by Davies (2004). The protein was found to unfold in the presence of GdnHCl and to adopt a native-like conformation when refolded from a high concentration of GdnHCl. The folding of p38 $\alpha$  was monitored using the probes of far-UV CD and tryptophan fluorescence. The protein contains five native tryptophans, two found in the N-terminal lobe of the kinase domain and the remaining three tryptophan residues found in the C-terminal lobe. To support the study of the folding of p38 $\alpha$  via tryptophan fluorescence a series of tryptophan to phenylalanine substitutions were performed to create single tryptophan mutants of p38 $\alpha$ . During this process, it was discovered that a single tryptophan residue in the core of the C-terminal domain was essential for the correct folding of the protein, and it was not possible to create a soluble, correctly folded mutant lacking this residue, despite attempts to replace the residue with several different residues, namely, tyrosine, histidine and lysine (Davies, 2004). A sequence analysis of several protein kinases indicated that the residue is absolutely conserved in kinases. The locations of the tryptophan residues of p38 $\alpha$  are highlighted in Figure 1.5.



**Figure 1.5:** Crystal structure of p38 $\alpha$ , with 5 native tryptophan residues highlighted and annotated. The essential tryptophan residue, residue 207 in the core of the C-terminal lobe is also indicated. Structure from pdb file 1WFC, rendered using ray tracing module of Pymol (DeLano, 2008).

The folding of the p38 $\alpha$  was observed initially by far-UV CD. The transition was too broad to be a two state transition, and was observed to appear to proceed via an intermediate, but was too broad to be well fitted to a three state folding model. The unfolding and refolding curves of the wild type protein and all of the soluble tryptophan to phenylalanine mutants created were super-imposable, indicating that the tryptophan to phenylalanine substitutions had not affected the folding of the protein. The unfolding and refolding transitions were also super-imposable, indicating that the folding was fully reversible (Davies, 2004).

The folding of p38 $\alpha$  was also observed by tryptophan fluorescence. The wild type protein unfolded via a three state transition, though the broadness of the transition meant that the data were not well fitted to three state folding models, with a large error in the parameters of the fit. A C-terminal tryptophan mutant was created, which replaced the tryptophans in the N-terminal lobe with phenylalanine.

This mutant also folded via a three state transition, although the parameters of the fit were better defined than was the case with the wild type protein. The folding of the single, core, essential tryptophan was also studied with a single tryptophan mutant containing only W207. This also folded via a three state transition, with the parameters of the fit being well defined (Davies, 2004). The refolding and unfolding transitions of the various tryptophan mutants were super imposable with themselves, indicating that the unfolding and refolding of the protein was fully reversible. The native spectra of the single tryptophan were examined and found to be too broad for a single tryptophan in a single conformation. This spectral broadening, which was observed to arise late in the refolding of the protein was attributed to a mixture of two conformational states being present under native conditions (Davies, 2004).

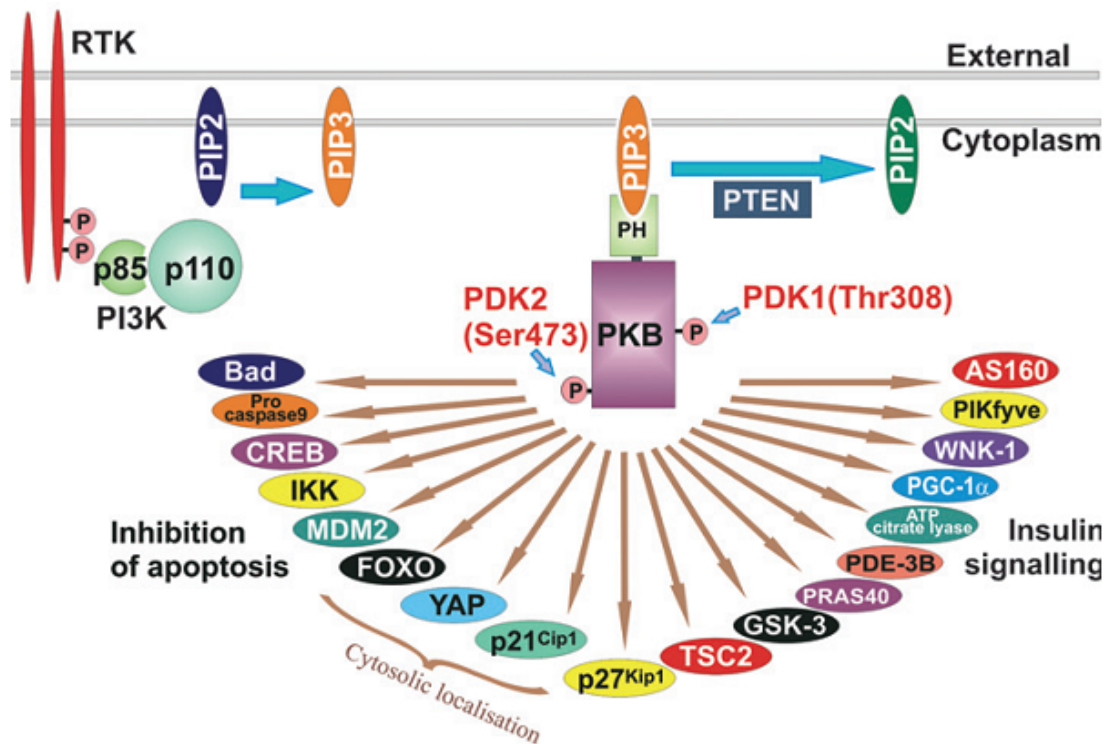
The folding of p38 $\alpha$  was observed to proceed via an equilibrium intermediate by both far-UV CD and tryptophan fluorescence. The two probes of folding used allows the nature of the intermediate to be speculated upon. At the denaturant concentration where the intermediate is formed the majority of the secondary structure is formed, as indicated by a high CD signal at 222 nm (Davies, 2004). The fluorescence spectra of the protein at this concentration indicate that there is still substantial exposure of the tryptophan residues to solvent, indicating that a molten globule intermediate is formed. From the behaviour of the different tryptophan mutants examined it can be determined that the C-terminal lobe is key to the folding of p38 $\alpha$  and that the intermediate forms in this region (Davies, 2004).

### **1.3.3 AKT2**

AKT2, also known as protein kinase B  $\beta$  (PKB $\beta$ ), is a serine/threonine protein kinase which is involved in growth factor and insulin signalling. Protein kinase B was originally discovered by homology cloning using specific probes for protein kinase A, and as the homologue of the v-Akt oncogene. The PKB family consists of three isoforms which are differentially expressed in human tissues. PKB $\alpha$  or AKT1 is expressed in all cell types, PKB $\beta$  or AKT2 is expressed in cells that are insulin targets, predominantly fat cells, liver and skeletal muscle. PKB $\gamma$  or AKT3 is not widely expressed, but is expressed in some cell types that are responsive to insulin. PKB isoforms are important in both cancer and in type 2 diabetes, with activated or overexpressed PKB found in several types of cancer. In type 2 diabetes the activation of the protein kinase activity is reduced (Sale and Sale, 2007).

The protein kinase B family members share a high sequence identity, with PKB $\beta$  and PKB $\gamma$  sharing 81% and 83% sequence identity with PKB $\alpha$ . The PKB protein is a cytosolic protein of 57 kDa, which contains three domains. These are a pleckstrin homology domain (PH domain) at the N-terminus of the protein, the kinase domain, which is central to the protein, and a final C-terminal domain. The PH domain binds two important cell signalling molecules with high affinity, phosphatidylinositol-3, 4, 5-trisphosphate and, phosphatidylinositol-3, 4-bisphosphate. The kinase domain consists of approximately 250 residues, and is very similar to the kinase domain of PKC and PKA. A conserved threonine in the domain, Thr309 in PKB $\beta$ , is required to be phosphorylated for full activity of the kinase domain. The C-terminal domain is ~40 residues in length and is most closely related to the PKC family. A conserved serine in this region is also required to be phosphorylated for full activity (Sale and Sale, 2007).

The activation of PKB occurs upon the binding of the PH domain to PIP3 in the membrane. PIP3 is generated by PI3K as a second messenger after activation of a receptor tyrosine kinase by ligand binding. The kinase domain of PKB is not directly activated by the binding to PIP3. Instead the binding is believed to cause a conformational change in the kinase domain and in the C-terminal domain which allows the key serine and threonine residues mentioned above to be phosphorylated by two upstream, activating kinases, PDK1 and PDK2. The phosphorylation of these residues then activates the kinase domain of PKB, which then phosphorylates downstream targets involved in insulin signaling (Hajduch *et al.*, 2001; Bae *et al.*, 2003), cell-cycle progression (Li *et al.*, 2002) and apoptosis (Kim *et al.*, 2001; Park *et al.*, 2002; Brunet *et al.*, 1999; Song *et al.*, 2005). The activation pathway and key substrates of PKB are shown in Figure 1.6.



**Figure 1.6:** Activation pathway and key substrates of PKB. Phosphorylated residue numbers shown are for the PKB $\alpha$  isoform. Figure adapted from Sale and Sale, 2007.

#### 1.3.4 KIS

KIS, or kinase interacting with stathmin, was initially discovered using a yeast two-hybrid system as interacting with stathmin, a small (19 kDa) cytoplasmic protein found enriched in neurons. (Maucuer *et al.*, 1995). The phosphorylation of stathmin, on its four phosphorylation sites, is correlated with the regulation of many cellular processes including neuronal differentiation, T-cell activation, stress response and cell cycle progression. (Maucuer *et al.*, 1997). Stathmin has been proposed to be a relay and integrator in cellular signaling networks (Sobel, 1991).

KIS is a 419 residue protein of 46.4 kDa. The full length protein has been predicted to contain two domains, a protein kinase domain at the N-terminus of the protein, and a C-terminal RNA binding domain. When KIS is overexpressed in HEK293 cells it is observed to be localized to the nucleus. This indicates that KIS may be important in regulating nuclear RNA processing through phosphorylation (Maucuer *et al.*, 1997).

KIS has also been shown to regulate a cyclin-dependent kinase inhibitor, p27<sup>Kip1</sup> through phosphorylation on serine 10. This phosphorylation of p27<sup>Kip1</sup> leads to the export of the regulator from the nucleus and its subsequent degradation. p27<sup>Kip1</sup> is involved in the regulation of the progression of the cell through the G1/S phases of

the cell cycle through the regulation of its inhibition of the cyclin dependent kinases (Sherr and Roberts, 2004). KIS expression is activated during G0/G1 phases of the cell cycle by mitogens, leading to a reduction in nuclear p27<sup>Kip1</sup> and cell proliferation. Inhibition of KIS activity presents a possible treatment for cell proliferation diseases such as cancers, through the maintenance of nuclear p27<sup>Kip1</sup> leading to cell cycle arrest.

As of the time of writing, there is no published structure of either the kinase domain of KIS, or of the full length protein. This makes the studying of the refolding of this protein kinase domain of particular interest.

### 1.3.5 *PhK*

Phosphorylase kinase is an enzyme which is involved in the regulation of glycogenolysis, which is activated through a cascade. The enzyme forms, in skeletal muscle, a hexadecameric complex, with four copies of each of four different subunits, which perform different roles. The subunits are termed the  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  subunits. (Brushia and Walsh, 1999). The  $\gamma$  subunit contains the active kinase domain, and the  $\alpha$ ,  $\beta$  and  $\delta$  subunits regulate its activity. The activity of the  $\gamma$  subunit is regulated by  $\text{Ca}^{2+}$  and this regulation is mediated by the  $\delta$  subunit, which is identical to calmodulin. The  $\alpha$  and  $\beta$  subunits inhibit the activity of the  $\gamma$  subunit, and the effect of the  $\delta$  subunit binding  $\text{Ca}^{2+}$  is to release this inhibition of activity (Brushia and Walsh, 1999).

The structure of the kinase domain of the  $\gamma$  subunit was solved in 1995 by Owen *et al.* (1995). They obtained inclusion bodies of the kinase domain of the  $\gamma$  subunit from rabbit skeletal muscle by expression in *E. coli*. They were able to obtain soluble protein by performing a rapid dilution of inclusion bodies solubilised in 8 M urea 50 mM Tris 25 mM DTT pH 8.5. The denatured and solubilised inclusion bodies were diluted 10 fold into a refolding buffer of 50 mM phosphate, 10% glycerol 2 mM EDTA, 1 mM EGTA 4 mM DTT pH 8.2. The protein was allowed to refold at 4 °C and the protein purified and crystallized. This marks the only example of the refolding of a protein kinase for the purpose of obtaining soluble protein for biophysical study that has been reported prior to the work presented here.

### 1.3.6 TTK, a Dual Specificity Kinase

Mps1 was first identified in yeast, during a mutagenesis screen which tested for mitotic spindle defective mutants. The original mutant causes a failure in the duplication of the yeast centrosome at a restrictive temperature. The mutant did not affect other aspects of cell division, as evidenced by the cells growing a bud and duplicating their DNA (Winey *et al.*, 1991). Subsequent to the identification of the mutant, the gene was cloned and fusion proteins created. These fusion proteins were purified from yeast extracts and shown to be phosphoproteins. The protein was also demonstrated to be a protein kinase with activity against both exogenous substrates and to be capable of autophosphorylation. A mutation was performed on the isolated gene, which abolished *in vitro* kinase activity. This mutant acted like a null mutant *in vivo* indicating that the *in vivo* action of the protein in the duplication of the yeast centrosome was dependent on the kinase activity of the protein (Lauzé *et al.*, 1995)

It was observed that the mutant cells produced continue to segregate their DNA despite the lack of a functional spindle. This suggested that the Mps1 gene product may be involved in more than spindle formation. The effect of several double mutants and induced Mps1 mutants showed that the Mps1 gene product was also involved in the checkpoint arrest of mitosis. The arrest of mitosis due to nocadazole depolymerisation of microtubules also did not occur in Mps1 mutants. It was shown that this failure of the checkpoint function in the mutants was independent of the failures in the spindle formation observed (Weiss and Winey, 1996).

The human homologue of yeast Mps1 or TTK has been shown to have similar, but divergent functions in human cell lines when compared to yeast Mps1. The expression of TTK varies according to the position of the cell cycle. TTK was present in the experiments performed at all stages of the cell cycle, but the expression levels increased as the cells approached mitosis. The kinase activity of the samples was assessed using fusion proteins, and was found to peak when the cells entered mitosis, which occurred 8 to 10 hours after the cells were released from a G1/S block. The expression levels increased ~3 fold whereas the kinase activity increased ~10 fold, indicating that there is an activation event occurring in addition to the increase in expression. Antibodies to TTK and siRNA inhibition of TTK both were capable of inactivating the spindle assembly checkpoint, similar to the role of Mps1 in yeast. However, similar experiments did not result in an effect on the duplication of the centrosome as was observed in yeast (Stucke *et al.*, 2002).

Human TTK is an 857 residue protein of a molecular mass of 96.9 kDa. It comprises three domains, a large N-terminal domain, followed by the kinase domain, and finally a short C-terminal domain. The roles of the N-terminal and C-terminal domains are not fully understood, however the N-terminal domain has been shown to have a role in the targeting of the protein to the kinetochore, the centre of the checkpoint signaling (Liu *et al.*, 2003).

The biological activity of the kinase domain is better understood. Evidence for cross-talk between the DNA damage checkpoint, and the spindle assembly checkpoint was gathered when it was observed that ablating the expression of human TTK reduced the phosphorylation of CHK2 and affected, therefore, downstream events from CHK2 phosphorylation, resulting in an impairment of the arrest in growth in response to DNA damage (Wei *et al.*, 2005). The presence of TTK is required for the localization of many of the other checkpoint proteins to the kinetochore including Bub1, BubR1, Mad1 and Mad2 (Abrieu *et al.*, 2001; Vigneron *et al.*, 2004). TTK is capable of autophosphorylation on a residue in the activation loop, T676, and a T676A mutant shows lower kinase activity *in vitro* but does not cause over-duplication of the centrosome when over expressed (Mattison *et al.*, 2007). This appears to be contradictory to the expected role of phosphorylation of T676 in the *in vivo* activity of TTK. However, studies performed using a construct of TTK fused to FK506 – binding protein allowed the dimerisation of TTK to be forced using a small molecule. This forced dimerisation resulted in autophosphorylation of TTK on T676. It was suggested that the localization of TTK to the kinetochore results in a high local concentration of TTK which could promote the trans-autophosphorylation of TTK on T676. The efficient trans-autophosphorylation of TTK in the kinetochore results in its activation, allowing it to perform its role in the spindle assembly checkpoint (Kang *et al.*, 2007).

## **1.4 Aims**

The existence of a limited number of protein folds in nature raises the question of whether similar protein folds possess similar folding properties, or whether the differences in the amino acid content of the proteins result in different folding properties for these proteins. Some evidence has been presented from studies performed on topologically complex proteins that does indicate that similar



structures can give rise to similar folding properties (Gloss, 2007; Mallam and Jackson, 2005).

Protein kinases are a sequence diverse family of protein domains which share a common, well defined fold. The protein kinase family contains proteins which have key roles in the control of cellular processes such as cell cycle progression and regulation (Maucuer *et al.*, 1997; Li *et al.*, 2002), and the regulation of glycogenolysis (Brushia and Walsh, 1999). Defects in the activation and regulation of protein kinase activity have been implicated in chronic diseases, such as diabetes and cancer. The development of small molecule inhibitors of protein kinases is an active area of pharmaceutical research, and the elucidation of the structure of protein kinases is a key part of this research, requiring the availability of milligram amounts of soluble, correctly folded protein. This has proved to be challenging to achieve with protein kinases, with many accumulating in inclusion bodies when the kinase domains are over expressed in bacteria.

Given the pharmaceutical importance of protein kinases, the difficulty in obtaining soluble, high purity recombinant protein kinase domains, and the existence of a previous study on the folding pathway of p38 $\alpha$ , a serine/threonine protein kinase (Davies, 2004), a further study into the folding of protein kinases would provide important insights into the possibility of achieving soluble protein through a common route, create mechanisms for testing for the refolding of protein kinases, and examine the possibilities of common folding properties in the conserved kinase fold, despite sequence divergence between protein kinases.

The specific aims of the work undertaken were:

- To create, and test, using a model protein kinase, a protein refolding screen, suitable for use with protein kinases, with associated means of determining the extent of the refolding of the protein used.
- To test this screen with additional protein kinases, to allow the comparison of the refolding of several protein kinases, to elucidate possible commonalities in the refolding of protein kinases.
- To develop a screen more suited to protein kinases than the initial screen based on the results achieved.
- To mutagenise, express, purify and characterize a soluble protein kinase used in the screen for more detailed examination of the folding of a protein kinase.

- To characterize the folding of a protein kinase used in the screen.
- To use the results of the refolding screens performed, the existing study on p38 $\alpha$  folding, and the folding study performed on an additional kinase to draw conclusions about the existence of a common protein kinase folding pathway.

## **Chapter 2: Design, Testing and Evaluation of a Protein Refolding Screen using p38 $\alpha$ as a Model Protein Kinase**

### ***2.1 Introduction***

In order to study the refolding of protein kinases, it is necessary to be able to refold the inclusion bodies produced upon expression of many protein kinases in *Escherichia coli*. The refolding of proteins is a challenging procedure, the success of which cannot be predicted accurately. To find conditions which lead to the refolding of protein kinases in quantities which allow further study it may be necessary to survey many conditions and chemical additives which may aid refolding. A screen is therefore desirable to rapidly test these conditions. This chapter describes the design and testing of such a screen for the refolding of protein kinases, with its associated methods of following the success of refolding in each case, using the test protein, human protein kinase p38 $\alpha$ .

To aid in the design of the screen, and to validate the methods chosen as part of the screening system it is necessary to select a test protein to use with the screen. p38 $\alpha$  was selected as a test protein since it had several desirable features for aiding in the design and testing of a screen. Firstly, p38 $\alpha$  is able to be obtained from *E. coli* in both a soluble, correctly folded form and in an insoluble form, depending on the expression conditions used. Secondly, as shown by Davies (2004), p38 $\alpha$  is capable of refolding from the denatured state under equilibrium conditions. Although the refolding technique used within the screen does not use equilibrium conditions, the results of Davies mean that should p38 $\alpha$  fail to refold within the screen, this should be due to the conditions used, and not a generic property of p38 $\alpha$ .

The study of the refolding of protein kinases via a refolding screen reveals details of the folding that are not easily accessible via equilibrium methods. Firstly, the screen allows a rapid comparison of the effects of pH upon the folding of protein kinases and secondly commonalities in the response of proteins to refolding additives may suggest similar folding pathways.

Refolding screens have been made available commercially recently, as the desire to obtain more challenging proteins for structural studies has increased. However, for application to protein kinases, these screens have several

disadvantages. Firstly, the volumes that are used in these screens tend to be low, limiting the techniques that can be used to quantify the refolded protein. As such, most screens recommend some form of activity assay for assessing the extent of refolding. Activity assays for protein kinases, are however problematic for a refolding screen application. The high concentrations of chemical additives used in refolding would interfere with the activity assay, either through effects on the protein itself, e.g. partial unfolding of the kinase, or the target protein by low concentrations of denaturant, or through effects on the assay readout. In addition, many protein kinases require activation by another protein kinase, e.g. p38 $\alpha$  requires activation by MKK6 dependent phosphorylation. The requirement for this phosphorylation by another protein kinase, and the purification of the activated protein kinase of interest from unactivated kinase and activating kinase makes activity assays unsuitable for a screen.

To enable the study of the refolding of protein kinases therefore, a refolding screen is needed, that uses more generic readout methods than commonly suggested. It is necessary to show that this screen is able to identify conditions leading to refolding of the target protein kinase on a scale that would yield amounts of protein suitable for further study, and that this protein adopts the correct fold. This chapter describes also the scaling of the refolding of p38 $\alpha$  to multi mg scale, gives evidence that the correct fold is adopted by the refolded protein, and examines possible explanations for the differences observed in the refolding of p38 $\alpha$  at different pHs.

## ***2.2 Materials and methods***

### ***2.2.1 Materials***

NV-10 was purchased from Novexin Ltd (Cambridge, UK). 3-(1-pyridino)-1-propanesulfonate and  $\beta$ -cyclodextrin were purchased from Fluka (Buchs SG, CH). Tris was purchased from Acros Organics (Geel, BE); P20 surfactant was supplied by Biacore (Chalfont St. Giles, UK) and dimethyl sulphoxide by Fisons (Ipswich, UK). All other chemicals were supplied by Sigma-Aldrich (Poole, UK).

### **2.2.2 Transformation of *E. coli* with Plasmid DNA**

In order to transform *E. coli* with plasmid DNA, a 2 µl aliquot of plasmid DNA, at a concentration of 100-150 ng/µL was added to 50 µL of competent cells of various strains. The mixture was mixed by gently flicking, and incubated on ice for 30 minutes. The sample was heat-shocked at 42 °C for 45 s and subsequently incubated on ice for a further 5 minutes. A 200 µL aliquot of pre-warmed SOC Broth (2 % w/v tryptone, 0.5 % w/v yeast extract, 8.56 mM NaCl, 2.5 mM KCl, 10 mM MgCl<sub>2</sub>, 20 mM Glucose, pH 7.0; Hanahan, 1983) was added, and the sample incubated at 37 °C for 1 hour at 300 rpm in a heated shaking block. 150 µL of the incubated sample was spread on LB Agar plates containing an appropriate antibiotic for the strain and the plasmid used.

### **2.2.3 Preparation of Plasmid DNA**

Plasmid DNA for Gateway cloning reactions, restriction digests, site directed mutagenesis and *E. coli* transformations was prepared using a QIAprep Spin Miniprep Kit (Qiagen), consisting of buffers P1, P2 N3 and PE and pre-packed silica columns, using either the centrifuge method, or the vacuum manifold method according to the manufacturer's instructions.

Briefly, a 10 mL overnight culture of *E. coli* cells of various strains containing the plasmid of interest was centrifuged and the supernatant discarded. The cell pellet was gently resuspended in 250 µL buffer P1 and lysed in alkaline conditions by the addition of 250 µL buffer P2. 350 µL of buffer N3 was used to precipitate the protein present in the sample. The sample was centrifuged and the supernatant was applied to a QIAprep spin column. The column was washed by the addition of 750 µL of buffer PE to the column, followed by centrifugation for 1 minute at 13000 rpm, or the application of vacuum to the column. 5-10 µg of DNA was then eluted in 50 µL of 10 mM TrisHCl, pH 8.5. The concentration of the eluted DNA was assessed by UV spectroscopy using equation 2.1.

$$\text{Concentration } (\mu\text{g/mL}) = (A_{260} - A_{320}) \times 50 \quad (2.1)$$

### **2.2.4 Expression of p38α as Inclusion Bodies**

p38α was prepared from insoluble aggregates using the following procedure. An expression construct for p38α was used to transform *E. coli* strain BL21\* (DE3)

cells as previously described (Section 2.2.2). The agar plates were incubated overnight at 37 °C. For expression, a single colony was picked from the plate and a 75 mL starter culture of Terrific broth containing 100 µg/mL ampicillin was inoculated with this colony. The starter culture was incubated overnight at 37 °C. Terrific broth was made as follows, 12 g tryptone, 24 g yeast extract, 4 mL glycerol were added to 900 mL distilled water. The media was autoclaved and 100 mL of 0.17 M  $\text{KH}_2\text{PO}_4$  0.72 M  $\text{K}_2\text{HPO}_4$  added once cool.

5 mL of starter culture was used to inoculate expression cultures of 600 mL terrific broth containing 100 µg/mL ampicillin and incubated in a shaking incubator at 37 °C and 180 rpm. When  $\text{OD}_{600}$  reached ~0.6 the expression of the protein kinase was induced by the addition of IPTG to a final concentration of 0.4 mM. The incubation was allowed to proceed for 20 hours before the biomass was harvested.

### ***2.2.5 Isolation and Solubisation of Inclusion Bodies***

Biomass was harvested from the expression cultures by centrifugation at 6000 g. Isolated cell paste was weighed and stored at -80 °C until used.

Inclusion bodies were isolated from the cell paste using a modified version of the method found in Georgio and Valax (1999). Frozen cell paste was resuspended in 50 mM Tris 200 mM NaCl 5 mM  $\beta$ -mercaptoethanol pH 8.0. The volume of buffer used was 50 mL per gram of cell paste. The cells were resuspended using a homogeniser and lysozyme added to a final concentration of 0.5 mg/mL. The resuspended cells were incubated at 4 °C for 3 hours before the lysis of the cells was completed by sonication. This was performed by 4 rounds of 30 s sonication followed by 2 minutes relaxation time. Samples were kept on ice between periods of sonication.

Cell debris and inclusion bodies were separated from the soluble material by centrifugation at 35,000 g. The supernatant was retained for further analysis, and the pellet washed with lysis buffer to remove residual lysozyme. The pellet that was recovered after this washing was also washed with lysis buffer containing 2% Triton X-100. Subsequent to this wash, the recovered pellet was washed with lysis buffer containing 2 M urea.

The washed pellet was solubilised using a chemical denaturant. The buffer 8 M urea 50 mM Tris 200 mM NaCl 10 mM  $\beta$ -mercaptoethanol pH 8.0 was added to

the washed pellet and incubated at 30 °C in a water bath for 1 hour. After the solubilisation, the inclusion body preparation was centrifuged at 35,000 g for 1 hour at 6 °C to remove insoluble cell debris still remaining in the preparation. The protein concentration of the inclusion body preparation was estimated by measuring the  $A_{280}$  of the inclusion body preparation, using  $\epsilon_{280}$  of 48130 M<sup>-1</sup>cm<sup>-1</sup>, and the purity was assessed by reducing SDS-PAGE (Section 2.2.7).

#### **2.2.6 Production of Native p38 $\alpha$**

Recombinant human p38 $\alpha$  was produced according to Davies (2004). Native, soluble p38 $\alpha$  for comparing refolding with inclusion body protein was denatured with 8 M urea, 10 mM DTT for 1 hr at 30 °C and stored at 4 °C until use. Protein concentration was calculated from absorbance at 280 nm, using  $\epsilon_{280}$  of 48130 M<sup>-1</sup>cm<sup>-1</sup>. The molecular weight of the protein and the absence of covalent modifications were confirmed by ESI-MS (Section 2.2.8).

#### **2.2.7 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)**

Protein samples were analysed by SDS-PAGE using a modification of the method of Laemmli (1970). 10% polyacrylamide Bis-Tris NuPAGE gels were purchased from Invitrogen (UK). Samples for analysis were prepared by 1:1 dilution in 2x SDS sample buffer (Invitrogen). Reducing conditions were created by the addition of  $\beta$ -mercaptoethanol to a final concentration of 10%. Samples were heated to 95 °C for 5 minutes prior to analysis.

Samples were separated by electrophoresis in 50 mM MES 50 mM Tris 1 mM EDTA 0.1% SDS pH 7.3 running buffer using an Invitrogen mini-cell gel tank at a separation voltage of 230 V. Proteins were visualized by staining using InstantBlue protein stain (Expedeon) or via silver staining (Section 2.2.12), depending on the expected amount of protein.

#### **2.2.8 Mass Spectrometry**

Mass spectrometry analysis of expressed proteins was used for quality control and identification of proteins. It was carried out using a LC ESI-MS. An aliquot of protein containing ~10 $\mu$ g of protein was separated by reverse phase HPLC on a Agilent 1100 series HPLC system using a Jupiter 5uC5 300Å reverse phase

column, with a linear gradient of 6 % to 95 % Acetonitrile + 0.25 % Formic Acid over 16 minutes. Eluate from the column was injected into a LCT ESI-TOF (Micromass) mass spectrometer. Spectra were constantly acquired for the duration of the elution. Data were collected using MassLynx (Micromass) and the protein spectra deconvoluted using the built in “MaxEnt1” module. The instrument was calibrated using CsCl and the calibration confirmed by analysis of a solution of myoglobin.

### **2.2.9 Design of a Refolding Screen for Kinases**

A series of additives known to be effective in facilitating the refolding of proteins, were selected to be included in a screen for the refolding of p38 $\alpha$ . In total 31 additives were chosen and grouped according to their chemical features. Arginine, glycine, L-proline, sarcosine and an arginine/glutamate mix were grouped together into the “amino acids” group (Willis *et al.*, 2005; Arakawa *et al.*, 2007; Baybes *et al.*, 2005; Arakawa *et al.*, 1985). Glucose, betaine, sorbitol and trimethylamine N-oxide were grouped into the “osmolytes” group (Blackwell and Horgan, 1991; Arakawa and Timasheff, 1985; Arakawa and Timasheff, 1982; Kadi *et al.*, 2006; Mello and Barrick, 2003). Sodium chloride, sodium sulphate and ammonium sulphate were grouped into the “simple salts” group (Willis *et al.*, 2005; Nishimura *et al.*, 2001). Guanidine (2, 1 and 0.5 M) and urea (2, 1, 0.5 M) were grouped as “denaturants” (Willis *et al.*, 2005; Yasuda *et al.*, 1998). Lauryl maltoside, CHAPS and Triton X-100 are grouped as “detergents” (Yasuda *et al.*, 1998; Ruan *et al.*, 2003). Cyclohexanol, 1-pentanol, ethanol, glycerol,  $\beta$ -cyclodextrin, ethylene glycol and PEG 3440 were grouped as “alcohols and polyols” (Willis *et al.*, 2005; Ruan *et al.*, 2003; Machida *et al.*, 2000; Silow and Oliveberg, 2003; Chong and Cheng, 2000; Goldberg *et al.*, 1996). 3-(1-pyridino)-1-propanesulphonate, formamide and NV-10 were grouped as “other additives” (Yasuda *et al.*, 1998; Zardeneta and Horowitz, 1992). These 31 additives and a control lacking a specific refolding additive were formatted into a 96-well refolding screen, utilising three different buffers for refolding (Table 2.1). The three buffers used were 100 mM MES (pH 5.8), 50 mM Tris (pH 8.0) and 100 mM CAPSO (pH 9.5).



### ***2.2.10 Refolding of Human p38 $\alpha$***

Refolding of p38 $\alpha$  was initiated by rapid dilution of denatured protein into various renaturation buffers in a 96-well screen, formatted in four 24 deep well plates. A volume of 5 mL of each renaturation buffer was aliquoted into each well. Under rapid agitation (~120 rpm) at 4 °C, 100  $\mu$ L of denatured protein solution, at 5 mg/mL in 8 M urea, was added in a single step to each renaturation buffer, for a final protein concentration of 0.1 mg/mL, and a final urea concentration of 160 mM using a repeater pipette with a 10 mL tip. Addition of denatured protein to an entire screen was complete within one minute of starting the addition of protein. Refolding was allowed to occur, under gentle agitation (~30 rpm) overnight at 4 °C. After refolding, samples identified as containing soluble protein by SDS-PAGE were concentrated 10-fold using Centricon concentrators, and then dialysed against 10 mM HEPES, 150 mM NaCl, pH 7.4.

### ***2.2.11 ePAGE of Refolded Protein***

Refolded protein was taken directly from the screen and added to ePAGE sample buffer (Invitrogen) and 10 %  $\beta$ -mercaptoethanol. Samples were heated to 70 °C for 10 minutes in a PCR thermal cycler block, according to the manufacturers instructions. Samples were run using 104 lane ePAGE gels (Invitrogen) and silver stained to visualise bands. Samples for which no band was visible were not further analysed.

### ***2.2.12 Silver Staining of SDS-PAGE and E-PAGE Gels.***

SDS-PAGE gels were stained using a modified version of the method described in Heukeshoven and Dernick (1988) when the protein level was too low for staining by the Coomassie based InstantBlue protein stain.

Subsequent to electrophoresis the gel was immersed in fixing solution, consisting of 40% (v/v) methanol, 10% (v/v) glacial acetic acid for 30 minutes to precipitate the protein present and remove SDS. Fixed gels were incubated with an oxidizing solution consisting of 30% (v/v) ethanol, 0.5 M sodium acetate, 0.5% (w/v) glutaraldehyde, 80 mM sodium thiosulphate for 30 minutes. Gels were washed with MilliQ deionised water for 10 minutes. This wash was performed a total of three times. After washing the gels were incubated with a silver solution, 0.1% (w/v)

silver nitrate, 0.02% (v/v) formaldehyde for 40 minutes. The staining was developed by exposing the gels, after washing, to 2.5% (w/v) sodium carbonate, 0.01% (v/v) formaldehyde until the bands were visualized. Once bands were visible the reaction was stopped using a solution of 40 mM EDTA, before washing with MilliQ deionised water.

When staining ePAGE gels, the procedure was adapted by doubling the time taken at each step in the procedure outlined above to account for the greater thickness of the ePAGE gels.

### ***2.2.13 Denaturing Capillary Electrophoresis***

Denaturing capillary electrophoresis was performed on refolded protein samples subsequent to concentration and dialysis using an Agilent ALP-5100 instrument. 4  $\mu$ L of concentrated, dialysed refolded protein solution was added to 2  $\mu$ L of reducing sample buffer and heated to 95 °C for 5 minutes. The heated solution was diluted with 24  $\mu$ L of water and analysed using the protein analysis program on the ALP-5100 instrument on a HT-2 pre-prepared chip. Calibration of the elution time of various molecular weight proteins was carried out automatically using the supplied protein ladder.

### ***2.2.14 Analytical Size Exclusion Chromatography***

Analytical size exclusion chromatography was performed using an Ettan LC system (GE Healthcare). Proteins were eluted from a pre-packed analytical scale Superdex 75 column of 2.413 mL bed volume (GE Healthcare) in 50 mM Tris, 150 mM NaCl, pH 9.0 at a flow rate of 50  $\mu$ L/min. Samples were diluted 1:1 with 0.2 mg/mL myoglobin in 10 mM HEPES before analysis and 25  $\mu$ L loaded onto the column. Absorbance was monitored at 280 nm for p38 $\alpha$  and myoglobin and at 410 nm to identify the haem group within myoglobin, to allow the elution time of myoglobin to be determined.  $A_{280}$  peak areas were compared to a standard curve to calculate the p38 $\alpha$  protein concentration.

Analysis of the multimeric state of denatured inclusion body protein and denatured soluble protein was carried out as above, with the exception that a Superdex 200 column was used in place of the Superdex 75 and the column was equilibrated in 8 M urea, 50 mM Tris, 150 mM NaCl, pH 9.0 and proteins were eluted in this buffer.

### **2.2.15 Detection of Binding Activity by Surface Plasmon Resonance**

The binding activity of refolded p38 $\alpha$  samples was assessed using a Biacore 3000 instrument (Biacore AS). An ureidoquinazoline target definition compound (UTDC) (Figure 2.5, Sullivan *et al.*, 2005) was immobilised in a single lane of a CM5 chip (Biacore) using standard amine coupling kit (Biacore). The surface of the chip was activated by a solution of EDC (1-Ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride) and NHS (*N*-hydroxysulphosuccinimide). This resulted in the creation of semi-stable amine reactive NHS-ester on the chip surface. UTDC at 400  $\mu$ M was then flowed over the surface, with the free amine group reacting with the ester, resulting in the release of NHS and the formation of a stable amide bond between the UTDC and the chip surface. Unreacted NHS-esters were removed by treatment of the surface with ethanolamine (1 M, pH 8.5). Control lanes were prepared in a similar manner, except that UTDC was not added to the surface subsequent to activation. p38 $\alpha$  was flowed over the prepared surface in 10 mM HEPES, 150 mM NaCl, 0.05 % P20 surfactant, 0.5 % DMSO. Response units due to protein binding to immobilised UTDC were compared to response units due to binding to the control flow lane. To eliminate non-specific binding protein was injected in the presence of 10  $\mu$ M UTDC to abolish specific binding. Samples were maintained at 8 °C until analysis. Response units due to specific p38 $\alpha$  binding were compared to a standard curve prepared by measurement of binding of soluble, correctly folded p38 $\alpha$ . Each condition was analysed in triplicate.

### **2.2.16 Large Scale Refolding**

Larger scale refolding of p38 was carried out by drop wise addition of solubilised inclusion bodies, at a protein concentration of 5 mg/mL, to renaturation buffer until the protein concentration reached 0.1 mg/mL and the urea concentration 160 mM, at a rate of ~ 1 mL/min. A total of 20 mg of protein was refolded using 1 L of refolding buffer in a 1 L beaker. The buffer was vigorously stirred using a magnetic stirring bar at a speed of ~120rpm. Refolding was allowed to occur, overnight at 4 °C under gentle stirring. The refolding solution was then concentrated 10-fold using an Amicon stirred cell with a 10 kDa cut-off ultra filtration membrane (Millipore) and subsequently dialysed against 10 mM HEPES, 150 mM NaCl, pH 7.4. A portion of the resulting solution was used for analysis via analytical size exclusion chromatography and assayed for binding activity by a surface plasmon

resonance (SPR) method. The remaining protein was further concentrated to approximately 1 mL using an Amicon cell and Millipore membrane as previously. The sample was then applied to a Superdex 75 12/60 column equilibrated in 50 mM Tris, 150 mM NaCl, pH 8.0 and was eluted from the column in the same buffer at a flow rate of 1 mL/min. Fractions identified as containing monomeric p38 $\alpha$ , by comparison of peak position with analytical size exclusion chromatography, were pooled and concentrated using a Centricon spin concentrator with a 10 kDa cut-off.

### ***2.2.17 Circular Dichroism***

Natively folded p38 $\alpha$  and refolded p38 $\alpha$  to be analysed by circular dichroism were extensively dialysed against 10 mM sodium phosphate buffer, pH 7.4, before analysis. Circular dichroism analysis of p38 $\alpha$  samples was performed on a JASCO J-810 spectropolarimeter, using a 1 mm path length quartz cuvette. Spectra were measured at a temperature of 20 °C, with a resolution of 0.5 nm. Spectra were collected between 185 and 260 nm. A scanning rate of 100 nm/min, time constant of 0.5 s and a bandwidth of 1 nm were used, and 8 scans were averaged per spectrum. Buffer blanks were subtracted from all spectra. The concentration of samples for CD analysis was confirmed by A<sub>280</sub> measurement before analysis was performed.

### ***2.2.18 iCycler Thermal Stability Measurements***

Thermal stability measurements of proteins were made using a PCR based thermal cycler (iCycler, Biorad) fitted with a iQ5 real-time PCR detection system (Biorad). Measurements were performed in iQ real-time PCR plates sealed with optical tape (Biorad).

Protein samples for analysis were prepared for analysis in a total volume of 25  $\mu$ L in iQ real time PCR plates. 12  $\mu$ L of protein solution at a concentration of 0.4 mg/mL was mixed with 12  $\mu$ L of a 2X buffer solution to regulate the pH in each well. Sypro-Orange (Molecular Probes) was diluted 1:40 and 1  $\mu$ L of this dilution was added to each well for a final dilution of 1:1000. The sealed plate was centrifuged at 1000 g for 1 minute to remove air bubbles.

The temperature was raised from 12 to 90 °C in steps of 0.2 °C. The temperature was maintained for 12 s at each step, and the fluorescence of the Sypro-Orange dye was recorded at each step. The resulting unfolding curve was fitted to

Equation 2.2 in order to obtain the thermodynamic parameters for the unfolding curve.

$$Y_{obs} = \frac{e^{\left( \frac{-\Delta H_m \left( 1 - \frac{T}{T_m} \right)}{RT} \right)} \times (a_u + b_u T) + (a_n + b_n T)}{1 + e^{\left( \frac{-\Delta H_m \left( 1 - \frac{T}{T_m} \right)}{RT} \right)}} \quad (2.2)$$

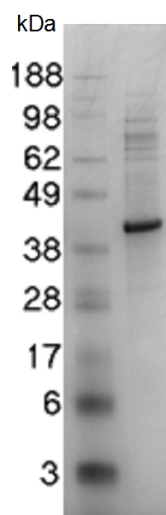
Where,  $T$  is the temperature in degrees Kelvin,  $T_m$  is the mid-point of the melting transition,  $\Delta H_m$  is the change in enthalpy at  $T_m$ ,  $R$  is the molar gas constant,  $a_n$  and  $b_n$  define the pre transition region of the melting curve and  $a_u$  and  $b_u$  define the post transition region.

## 2.3 Results

### 2.3.1 Expression of p38 $\alpha$ as Inclusion Bodies

p38 $\alpha$  was expressed as inclusion bodies, and those inclusion bodies purified as described in sections 2.2.4 and 2.2.5. The expression of wild type p38 $\alpha$  was induced with 0.4 mM IPTG at a temperature of 37 °C, since lower temperatures and IPTG concentrations are known to result in the expression of wild type p38 $\alpha$  in a soluble form (Davies, 2004).

Isolated inclusion bodies were solubilised by denaturation with 8 M urea 50 mM Tris 200 mM NaCl 10 mM  $\beta$ -mercaptoethanol at 30 °C for 1 hour. This procedure produced an inclusion body preparation of p38 $\alpha$  with high purity (Figure 2.1). The molecular mass of the p38 $\alpha$  inclusion body preparation was determined by ESI-MS (Section 2.2.8, data not shown) and was determined to be in good agreement with the expected mass of 41,311.4 Da, calculated from the sequence of the expressed protein. The yield of p38 $\alpha$  achieved was 375 mg protein per litre expression culture.



**Figure 2.1:** Purity of p38 $\alpha$  inclusion body preparation assessed by reducing SDS-PAGE. 10% Bis-Tris gel run with NuPAGE MES-SDS running buffer. Lane 1, molecular weight markers (SeeBlue Plus 2, Invitrogen); lane 2, p38 $\alpha$  inclusion body preparation in 8 M urea.

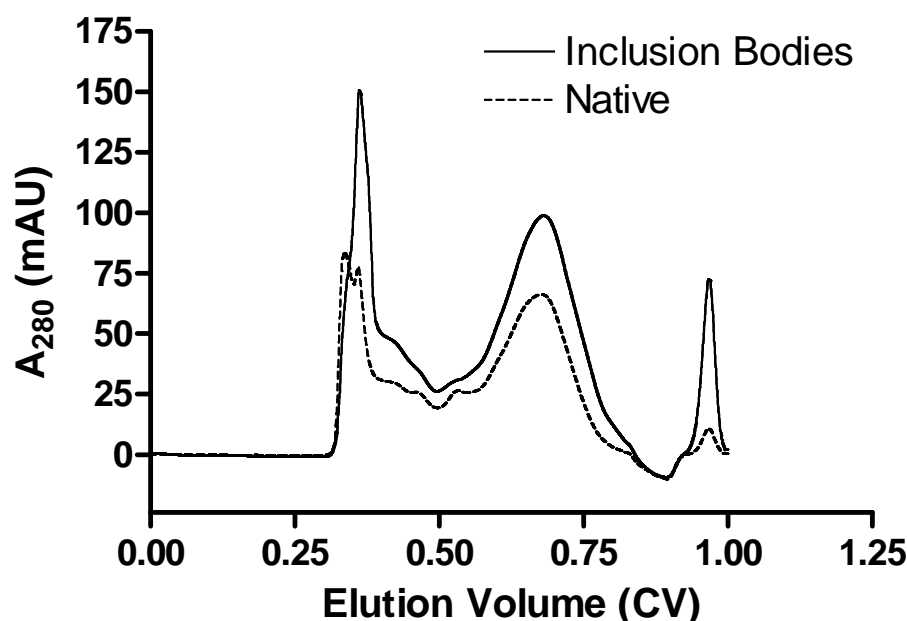
### 2.3.2 Expression of p38 $\alpha$ in a Soluble Form

Wild type p38 $\alpha$  was also expressed in a soluble form, as described in Davies (2004), with the exception that the final ion exchange step, using a monoQ column, was omitted, since the expressed protein was found to be of sufficient purity following purification by ion exchange and size exclusion chromatography. Soluble p38 $\alpha$  was either stored at 4 °C until use or denatured with 8 M urea as described in section 2.2.6. Soluble p38 $\alpha$  was used as a control protein for the analytical methods that were used to test the refolding system. The yield of soluble protein achieved was 50mg per litre, consistent with the yields achieved by Davies (2004). The molecular mass of the purified p38 $\alpha$  was determined by ESI-MS (Section 2.2.8, data not shown) and was determined to be in good agreement with the expected mass of 41,462.20 Da, calculated from the sequence of the expressed protein and was also in agreement with the results of Davies (2004).

### 2.3.3 Comparison of the Monomeric State of the Two Forms of p38 $\alpha$

As a quality control, to show that the denaturation of inclusion body preparations and soluble protein were producing fully denatured protein, and to compare the denatured state of the two preparations, it was necessary to examine denatured p38 $\alpha$  by size exclusion chromatography. Both preparations produced a

single peak corresponding to a retention time consistent with monomeric p38 $\alpha$  (Figure 2.2).



**Figure 2.2:** Elution profile of native folded protein and inclusion body protein denatured with 8 M urea from analytical scale Superdex 200 column (volume 2.4 ml). Elution buffer 8 M Urea, 50 mM Tris, 150 mM NaCl pH 8.0. Column volume calculated as fraction of manufacture's stated column bed volume.

#### **2.3.4 Design of a Protein Refolding Screen**

The refolding screen was formatted to allow the screen to be adapted to automation if the screen was to be adopted for use within AstraZeneca on a regular basis. For this reason the screen was designed to utilise standard plate formats. The refolding reaction occurred in 24 deep well plates. The refolding method of rapid dilution was decided upon, since it was the easiest to perform in the formats selected, as opposed to methods such as dialysis, or column based refolding which have low throughput using standard methods, and have problems with cross contamination of additives when adapted for higher throughput. The volume of each refolding reaction was set at 5 mL, since this volume allowed a high shaking speed of the 24 well plates used, giving better mixing when liquid was added to a well, whilst minimising cross contamination. The volume would also allow for a quantity of protein to be recovered which would allow analysis by several methods, which would improve confidence in the results.

Since protein kinase domains are found in the intracellular environment, and are therefore expected to be in a reducing environment, and to fold in such an environment, it was decided to maintain the refolding screen in a reducing environment, as opposed to an oxidising environment, or in the presence of a redox couple system, such as reduced/oxidised glutathione. On the basis of the work of Willis *et al.* (2005), DTT was chosen as the reductant of choice, since it provided the best balance between effectiveness and cost.

The refolding of proteins is known to be influenced by the pH at which refolding is attempted. The ionisation of key residues in proteins, particularly histidine residues, affects the formation of hydrogen bonds which are important for the structure of the protein. To allow the effect of changes in pH to be examined, and to allow any effect of the changes in ionisation to be visible, three pHs were selected for use in the screen, pH 5.8, pH 8.0, pH 9.5. These pHs were selected since they are spaced away from a neutral pH and so should increase the magnitude of any effect. In addition, the calculated pI of p38 $\alpha$  is 5.48 using the method of Bjellqvist *et al.*, (1993). The lowest of the screen pHs, pH 5.8 is close to the pI and the remaining two pHs are significantly greater. The different pHs will allow the comparison of the effect of the relationship between the pH at which refolding occurs, and the pI to be identified. Given the low pI of p38 $\alpha$ , it is not necessary to include refolding conditions at pH 7.4, since these would be expected show little difference from pH 8.0, as these are both significantly higher than the pI. Although it is common in refolding to use three component buffer systems that allow a wide pH range to be used whilst keeping the chemical components of the buffer the same, it was felt that this was not necessary and lower concentration buffers were used instead. 100 mM MES was used at pH 5.8, 50 mM Tris at pH 8.0, and 100 mM CAPSO at pH 9.5.

The refolding screen included a large number of chemical additives that were included to test for their ability to assist the refolding of protein kinases. These additives were selected on the basis of a review of the literature on protein refolding. The mechanisms of action of some of these additives are known, and varied between the additives. For example, polyethylene glycol acts as both a crowding agent, increasing the effective protein concentration, promoting folding, and presents a hydrophobic surface, inhibiting the formation of hydrophobic aggregates of protein. Several detergents were also included to exploit this effect.



Chemical additives were included that were not expected to allow the protein kinases tested in the screen to refold. These additives were, instead, expected to allow an intermediate fold of the protein to remain in solution, which, as the additive was removed by dialysis, would promote a slower, and therefore more successful, folding to the native state to occur. The denaturants that were included were both urea and guanidine hydrochloride. These denaturants were included at three different concentrations, 0.5 M, 1 M and 2 M, since the concentration at which such an intermediate would accumulate is not known for all protein kinases. Both denaturants were included since their different strengths allowed more possible points on any unfolding / refolding transition to be covered. Protein based additives, such as purified chaperones e.g. GroEL/GroES complexes or human HSPs e.g. HSP90, or antibody fragments or scaffold proteins were not included in the screen,

The concentration at which refolding is performed is critical for the outcome of refolding. Refolding at high concentrations is known to result in low yields of refolding, since the high concentration of exposed hydrophobic sites promotes aggregation. However, if the concentration was too low then this would not allow enough protein to be recovered from the screen for subsequent analysis. Since p38 $\alpha$  was already known to successfully refold under equilibrium conditions at a concentration of 0.1 mg/mL, this concentration was selected for the screen.

The 31 refolding conditions and a control condition at each pH, giving a total of 96 conditions, were formatted into a single screen, which was laid out as shown in table 2.1.

**Table 2.1** Layout of the initial protein kinase refolding screen.

0.9M Arginine	0.5M Glycine	1M Proline	2M Sarcosine	50mM Arginine/50mM Glutamate	1M Glucose	10% Betaine	20% Sorbitol	1M TMAO	0.5M NaCl	0.5M Sodium Sulphate	1M Ammonium Sulphate
2M Guanidine	1M Guanidine	0.5M Guanidine	2M Urea	1M Urea	0.5M Urea	0.006% Lauryl Maltoside	20mM CHAPS	20mM Triton X-100	1mM Cyclohexanol	5mM Pentanol	10% Ethanol
50% Glycerol	60mM Cyclodextrin	10% Ethylene Glycol	0.04% PEG 3440	1M NDSB	1M Formamide	NV-10	No Additive	0.9M Arginine	0.5M Glycine	1M Proline	2M Sarcosine
50mM Arginine/50mM Glutamate	1M Glucose	10% Betaine	20% Sorbitol	1M TMAO	0.5M NaCl	0.5M Sodium Sulphate	1M Ammonium Sulphate	2M Guanidine	1M Guanidine	0.5M Guanidine	2M Urea
1M Urea	0.5M Urea	0.006% Lauryl Maltoside	20mM CHAPS	20mM Triton X-100	1mM Cyclohexanol	5mM Pentanol	10% Ethanol	50% Glycerol	60mM Cyclodextrin	10% Ethylene Glycol	0.04% PEG 3440
1M NDSB	1M Formamide	NV-10	No Additive	0.9M Arginine	0.5M Glycine	1M Proline	2M Sarcosine	50mM Arginine/50mM Glutamate	1M Glucose	10% Betaine	20% Sorbitol
1M TMAO	0.5M NaCl	0.5M Sodium Sulphate	1M Ammonium Sulphate	2M Guanidine	1M Guanidine	0.5M Guanidine	2M Urea	1M Urea	0.5M Urea	0.006% Lauryl Maltoside	20mM CHAPS
20mM Triton X-100	1mM Cyclohexanol	5mM Pentanol	10% Ethanol	50% Glycerol	60mM Cyclodextrin	10% Ethylene Glycol	0.04% PEG 3440	1M NDSB	1M Formamide	NV-10	No Additive

0.1 M MES pH 5.8
50 mM Tris pH 8.0
0.1 M CAPSO pH 9.5

Following the addition of a solution of denatured protein to the screen, refolding was allowed to occur, under gentle agitation overnight at 4 °C. Low temperature was utilised, since this has been shown to lead to higher yields of refolded protein. The extended time allowed folding reaction, possibly slowed by the low temperature, to fully complete before further analysis was performed on the refolded protein.

The analytical methods that are available to identify the extent of refolding that had occurred in the screen are often sensitive to the refolding additives that were used, and would give unreliable readouts if they were present. Some potential readouts would also require higher protein concentrations than would be present in the screen. Therefore, subsequent to refolding, the screen conditions were concentrated and dialysed against 10 mM HEPES 150 mM NaCl pH 7.4 as described in section 2.2.10. This also increased the stringency of the screen, by driving the equilibrium of the protein in solution towards aggregation by removing solubilising additives and increasing protein concentration.

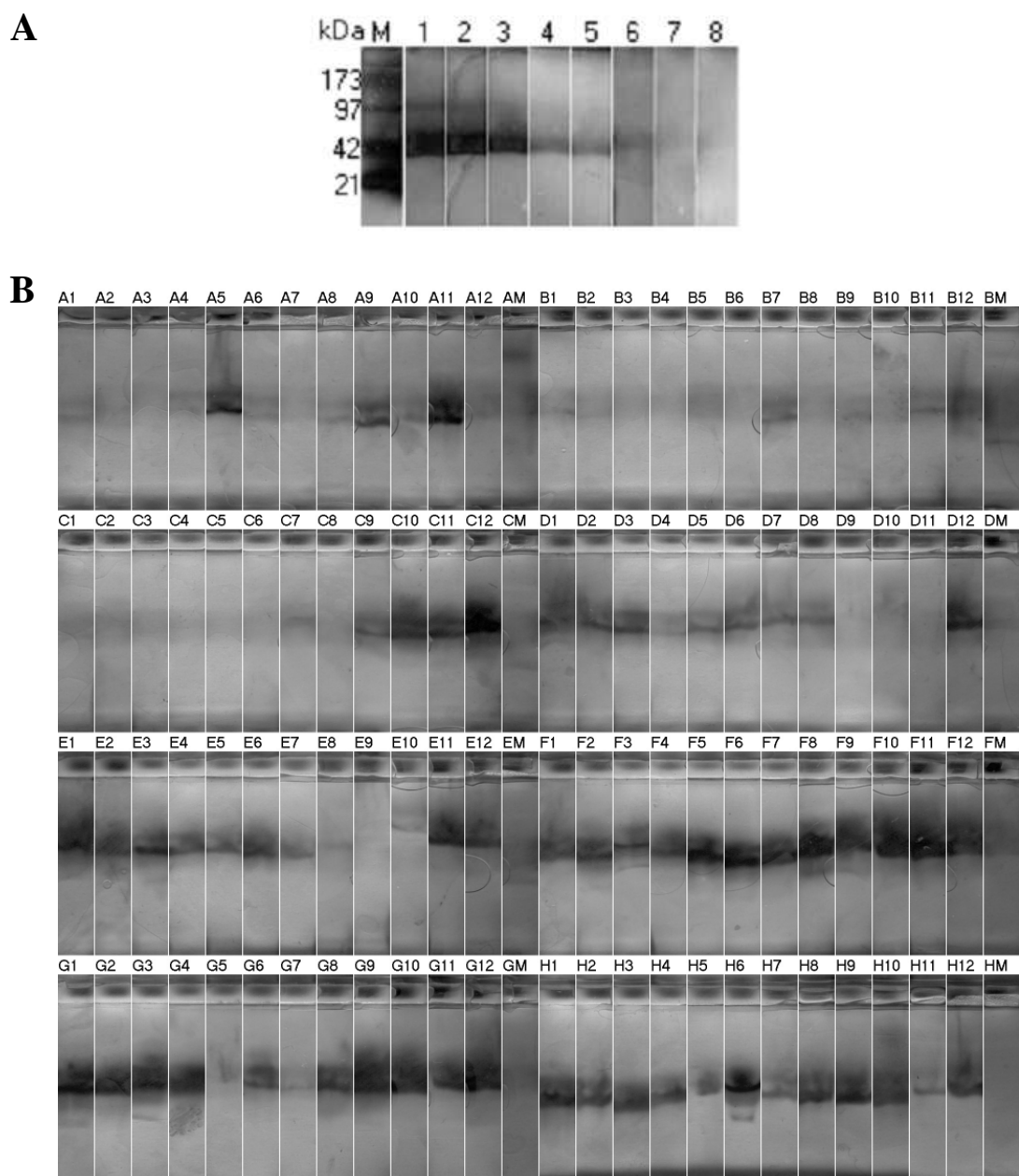
### ***2.3.5 Analytical Readouts for Refolding Screens***

A number of analytical methods were applied to provide an assessment of the yield and folded state of the refolded protein. Purified folded p38 $\alpha$  was used to identify the limits of detection for each method and to demonstrate the repeatability of the methods. The analytical methods were also selected to allow the number of conditions analysed, and therefore the time taken to be lowered at each stage, as the information content of the readouts increased.

An analytical method was required to identify the effectiveness of the screen additives at maintaining p38 $\alpha$  in a soluble form after dilution of the denatured p38 $\alpha$  into the various conditions. SDS-PAGE, using a 96 well ePAGE gel was chosen because of its relative insensitivity to refolding additives, compared to other analytical methods, and the low limit of detection which was critical for detecting soluble protein at the low concentration present in the screen (maximum of 0.1 mg/mL). The ePAGE gels were used to assess, in a qualitative manner, whether refolded protein was present in the various screen conditions that were used before more time consuming analytical methods were applied to the screen results. A series of dilutions of p38 $\alpha$  was run in quadruplicate on a single gel and stained according to the methods in sections 2.2.11 and 2.2.12. Using

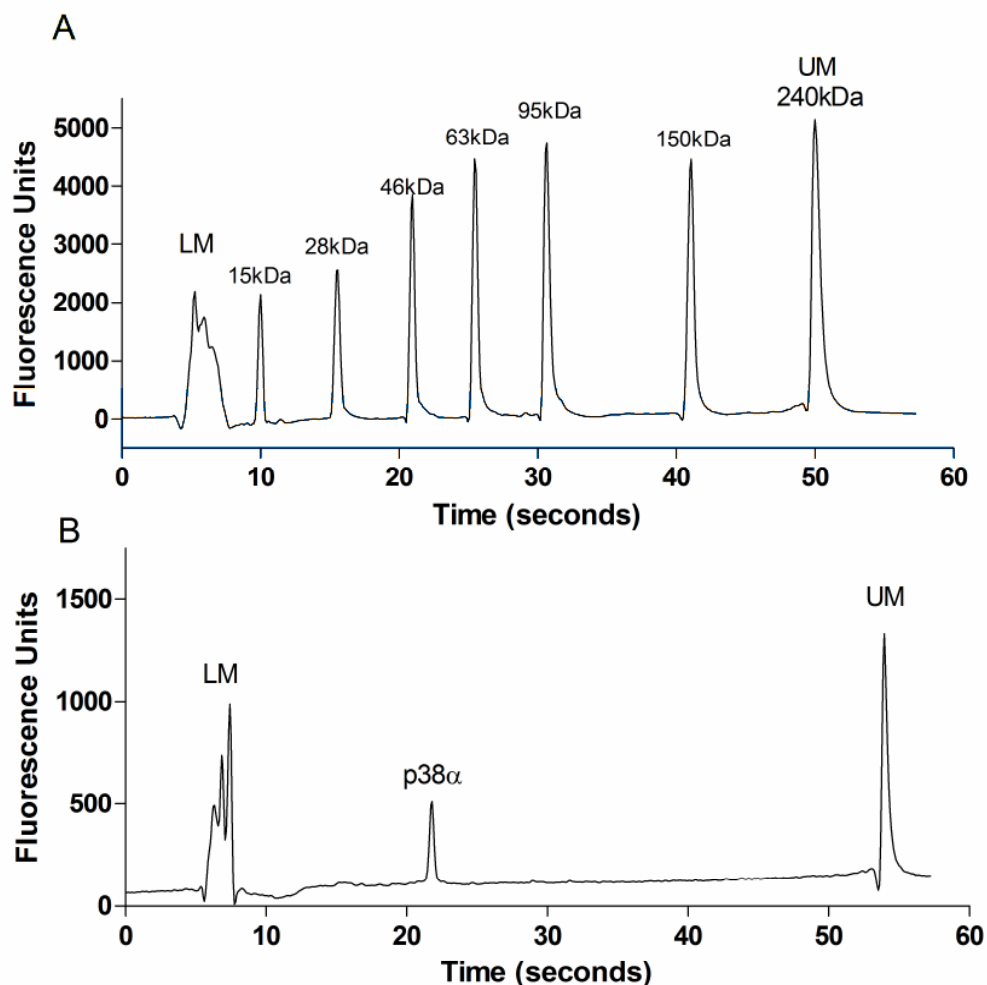
this dilution series, the lowest detected band contained ~8 µg/mL protein in a 10 µL load volume, which corresponded to a screen recovery of 1.25 % with a 20µL load volume (Figure 2.3).

Samples for use with this analytical method were taken directly from the refolding buffers after overnight incubation under gentle agitation at 4 °C (section 2.2.10). The incubation allowed aggregates of misfolded protein to precipitate under 1 g. This analytical method was used to restrict the number of conditions that were further analysed. If a particular condition did not give a detectable band on the gel, i.e. the recovery of soluble protein was less than 1.25 %, then the condition was not further analysed. Control p38α was used to test the ability of this analytical method to identify soluble p38α in the presence of the various refolding additives (data not shown). Refolding conditions containing guanidine were found to precipitate the SDS in the sample buffer, preventing the analysis of these samples. All samples containing guanidine were therefore analysed by further analytical methods, despite the lack of a visible band.



**Figure 2.3:** (A) Dilution series of native folded p38a run on 6% E-PAGE gel under reducing conditions and stained using silver staining. Lane M, molecular weight markers; lane 1, 520  $\mu\text{g/mL}$ ; lane 2, 260  $\mu\text{g/mL}$ ; lane 3, 130  $\mu\text{g/mL}$ ; lane 4, 65  $\mu\text{g/mL}$ ; lane 5, 32.5  $\mu\text{g/mL}$ ; lane 6, 16  $\mu\text{g/mL}$ ; lane 7, 8  $\mu\text{g/mL}$ ; lane 8, 4  $\mu\text{g/mL}$ . (B) Example of E-PAGE gel from screen; numbered lanes correspond to single conditions from screen (Table 1). Gel images were processed using E-Editor (Invitrogen) to align lanes.

Capillary electrophoresis was chosen as a second analytical technique to be applied to refolded protein produced by the screen to quantify the concentration of soluble protein after concentration and dialysis. This analytical method was chosen because it offers a high throughput, quantitative method for analysing the total protein remaining in solution. The earlier analytical method, SDS-PAGE, only allowed a semi-quantitative assessment of protein concentration and was used in a qualitative manner. The mean recovery was 237  $\mu\text{g/mL}$  with a standard error of 2 %, demonstrating that the assay is reproducible. Sample data from this analytical method are shown in Figure 2.4. This technique was used as a high throughput method to screen for the recovery of soluble protein after concentration and dialysis. The recovery of soluble protein identified was used as an additional restriction on the number of conditions analysed by further techniques. Conditions that did not show a recovery of at least 5 % were not analysed further.

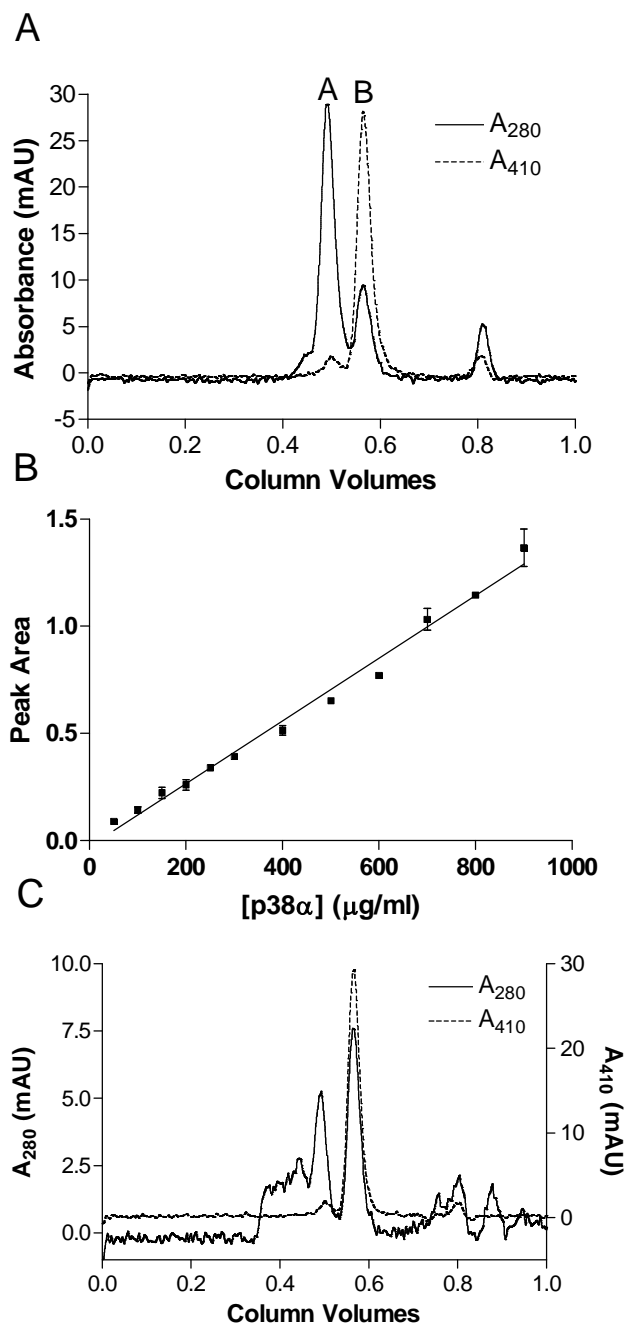


**Figure 2.4:** Capillary electrophoresis analysis of refolded p38 $\alpha$ . . Panel A – protein ladder, LM – lower marker, UM – upper marker, molecular weight of markers indicated above peaks. Panel B – example screen readout electropherogram. p38 $\alpha$  concentration identified as 20  $\mu\text{g/mL}$ , corresponding to a screen recovery of 2%.

*In-vitro* refolding of a protein may result in intermediate species which are prone to intermolecular interactions with other p38 molecules resulting in the formation of soluble oligomers, which are not present in correctly folded preparations of native protein. Therefore, in order to identify if the soluble protein is in the correct monomeric form, an analytical method, which separates and quantifies the monomeric protein recovered from a refolding condition was required. For these purposes, analytical size

exclusion chromatography was used, as described in section 2.2.14. To account for injection-to-injection variability due to the auto sampler used on the system, it was necessary to include an internal standard. Myoglobin was selected due to the haem group, which would allow the elution peak to be identified by measurement of absorbance at 410nm, and because myoglobin has a small hydrodynamic radius, meaning it would be well separated from the sample protein in the elution. A sample elution profile, from a standard curve is shown in Figure 2.5A. A standard curve of p38 $\alpha$  was analysed via this method in triplicate to allow the concentration of monomeric protein to be calculated, and to demonstrate the consistency of the assay. The assay was reproducible and proved to have a limit of detection of 50  $\mu\text{g/mL}$  (Figure 2.5B). A typical chromatogram obtained from a screen condition is shown in Figure 2.5C.



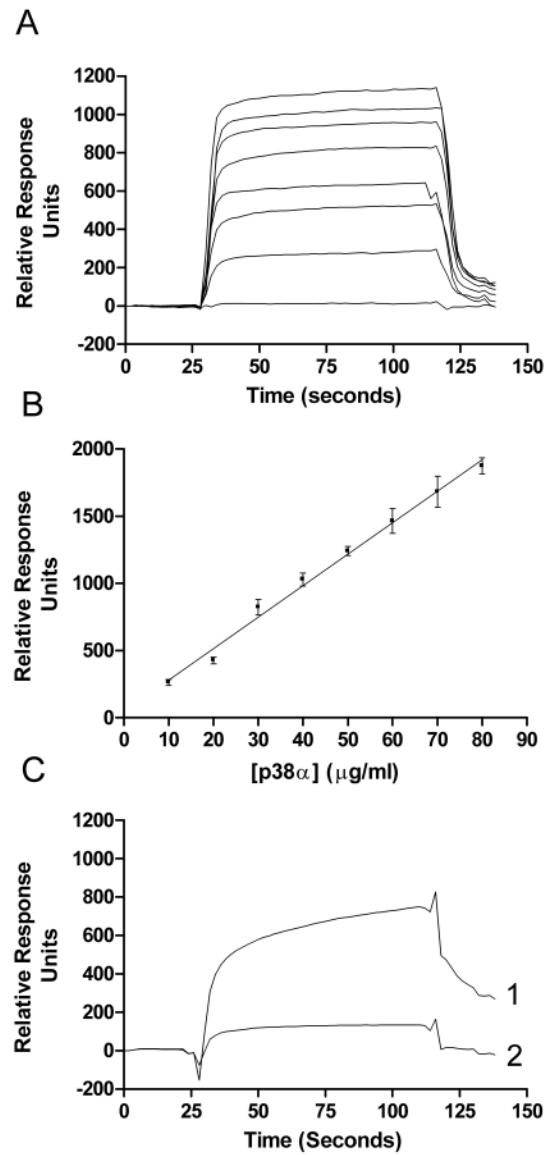


**Figure 2.5:** Analytical size exclusion chromatography analysis of native folded and refolded p38α. Panel A - Elution profile of 0.5 mg/mL p38α and 0.1 mg/mL Myoglobin. Peak A – p38α, MW 41.3 kDa, Peak B – Myoglobin, MW 16.2 kDa. Elution time recorded in column volumes. Panel B – standard curve of p38α analysed by analytical size exclusion chromatography. Panel C – sample elution profile from screen condition. p38α concentration identified as 165 μg/mL, corresponding to a screen recovery of 16.5 %.

The refolding of a protein may also result in a soluble form of the protein which is monomeric but which is not correctly folded. To identify the recovery of correctly folded p38 $\alpha$ , the binding of refolded protein to a UTDC (Figure 2.6) known to bind to unphosphorylated correctly folded p38 $\alpha$  was quantified using SPR as described in section 2.2.15. The structure of the compound used is shown in figure 2.6. Conditions in the refolding screen are likely to lead to the formation of non-native p38 species, which may interact non-specifically with the compound at the surface of the chip. To control for this non-specific binding, the binding of p38 $\alpha$  to the prepared surface was analysed in the absence and presence of an excess of the UTDC in solution.

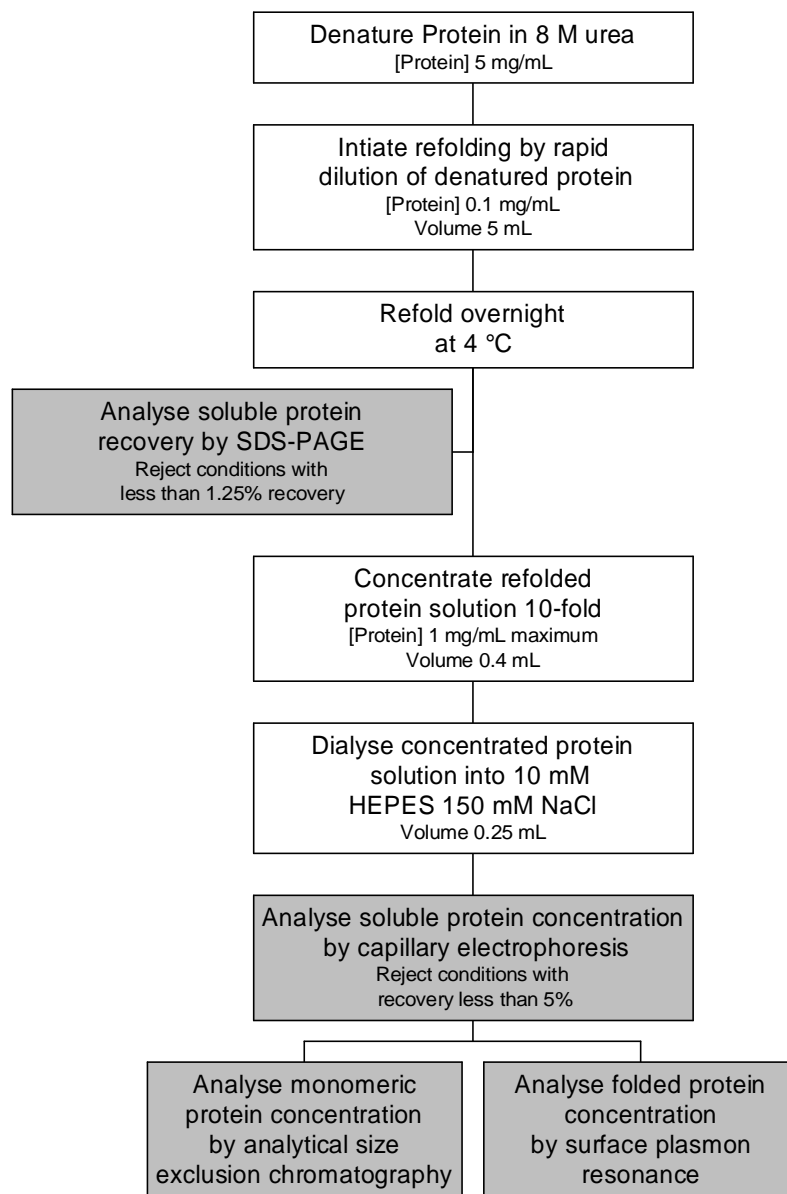
**Figure 2.6:** Structure of UTDC used to detect concentration of correctly folded protein present in refolded protein samples. Figure adapted from supporting information of Sullivan *et al.*, 2005.

The excess of free UTDC at 1700-fold above the affinity in solution prevented the specific binding of p38 $\alpha$  to the prepared surface, and so allowed the amount of non-specific binding to be identified. The relative response units used to quantify the binding of p38 $\alpha$  to the prepared surface was the difference in response units in the presence and absence of free UTDC in solution at 10  $\mu$ M. A wide range of protein concentrations was tested using solubly expressed, purified p38 (Figure 2.7A), and identified the linear range of response as 10 to 80  $\mu$ g/mL protein. A standard curve of relative response units was calculated from a triplicate analysis of control p38 $\alpha$ , showing that the analytical method is reproducible (Figure 2.7B). Typical response curves obtained when analysing a screen condition are illustrated in Figure 2.7C. This analytical readout was performed in parallel with analytical size exclusion chromatography.



**Figure 2.7:** Surface plasmon resonance analysis of refolded and native p38 $\alpha$  binding to immobilised TDC. Panel A – series of injections of p38 $\alpha$  onto prepared chip showing increasing response with increasing p38 $\alpha$  concentration. Panel B - Standard curve for analysis of control soluble p38 $\alpha$  binding by surface plasmon resonance. Each point mean of three replicate analyses. Error bars show standard deviation. Panel C – sample analysis of screen condition post dialysis. Curve 1 – refolded p38 $\alpha$  sample, curve 2 – refolded p38 $\alpha$  sample + 10  $\mu$ M TDC in solution. p38 $\alpha$  concentration identified as 32  $\mu$ g/mL, corresponding to a screen recovery of 14.2% once dilution of analysed samples corrected.

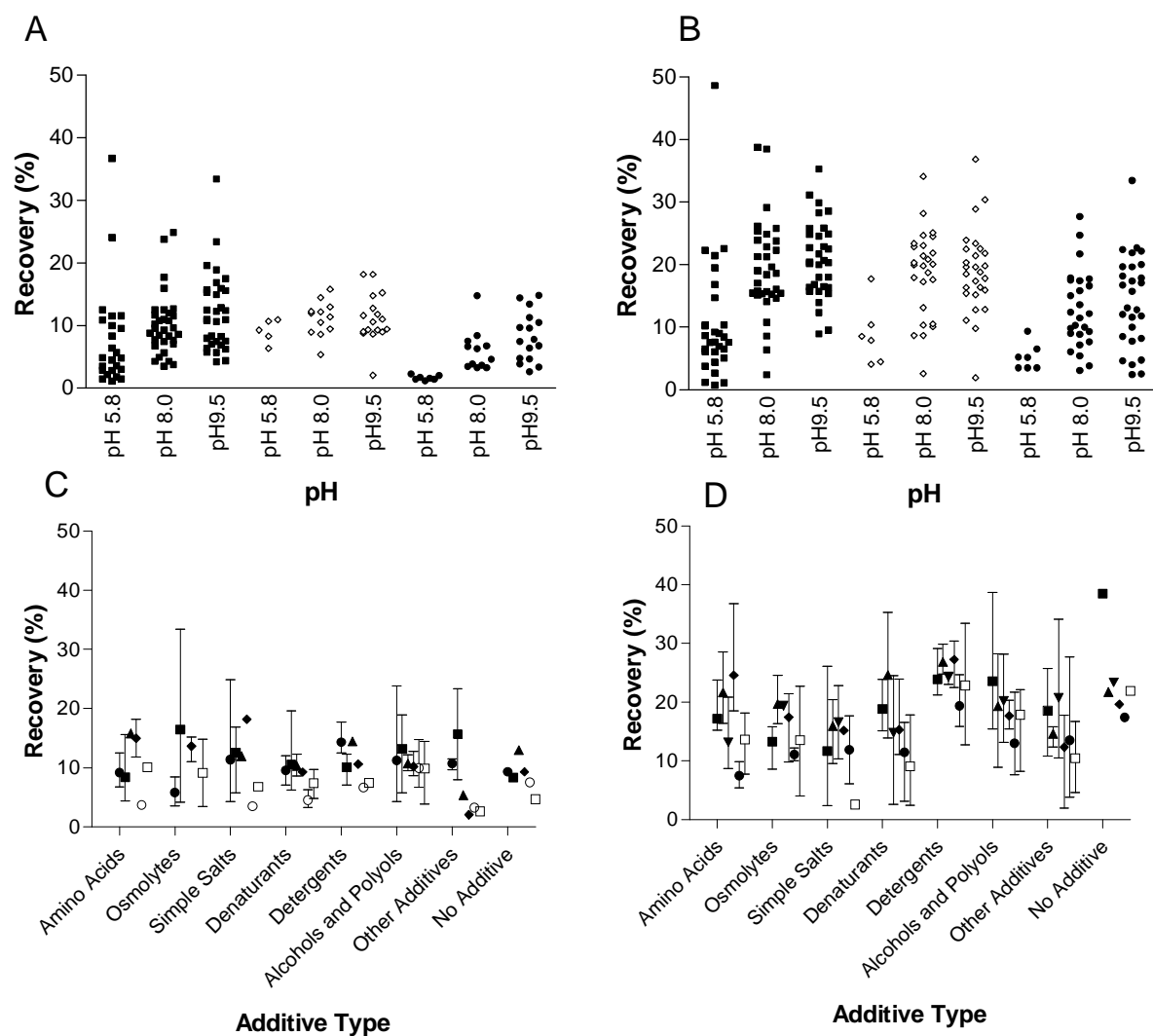
The described analytical readout methods and the screening design described in section 2.3.4 were combined to form a screening system. This system is summarised in the following diagram.



**Figure 2.8:** Overview of the process of screening for conditions that enhance the refolding of p38 $\alpha$ . The process was designed to rapidly identify the yield of refolding in a large number of conditions. Analytical methods are shown in shaded boxes. Volume per condition at each stage and maximum protein concentration are shown for each screen stage also.

### ***2.3.6 Refolding Screen for p38 $\alpha$***

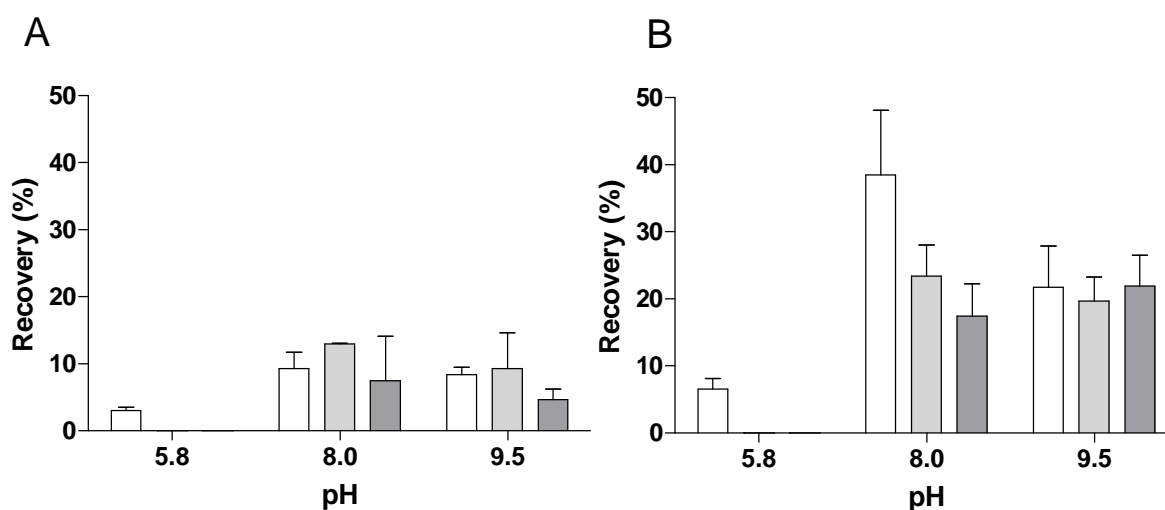
Inclusion bodies and soluble p38 $\alpha$  were denatured with 8 M urea and tested in the screen, in order that the refolding of both sources of protein could be compared. The refolding screen was performed in triplicate to provide reliable data on the refolding of p38 $\alpha$ . The recovery of soluble protein measured by capillary electrophoresis was generally higher than the recovery of monomeric protein measured by analytical size exclusion chromatography and the recovery of folded protein measured by SPR was generally similar to the recovery of monomeric protein (Figure 2.9 A, B). The refolding yields obtained for each additive in the screen proved to be reproducible across the three replicates, as indicated by the standard error of the mean shown in Figure 2.8. The refolding yields identified by the analytical methods used were compared for p38 $\alpha$  derived from inclusion body and for denatured native p38 $\alpha$ .



**Figure 2.9:** Recoveries of soluble, monomeric and folded protein obtained when refolding inclusion bodies (A, C) and soluble, denatured protein (B, D). For A and B, each point represents the mean of three experiments for a single refolding condition, ■ – soluble protein recovery; ○ – monomeric protein recovery; ● – folded protein recovery. For C and D, each point represents the mean of 3 analyses of all the additives in the given additive type, ■ – pH 8.0 soluble protein recovery; ▲ – pH 9.5 soluble protein recovery; ▼ – pH 8.0 monomeric protein recovery; ◆ - pH 9.5 monomeric recovery; ○ – pH 8.0 folded protein recovery; □ – pH 9.5 folded protein recovery. Error bars indicate range of individual additive means.

Strong differences were found in the effect of the pH on the efficiency of the refolding. With both sources of refolded protein (inclusion body protein and denatured soluble protein), there was a large drop in the recoveries of soluble protein, of monomeric

protein and of correctly folded protein at pH 5.8 when compared to the same recoveries at pH 8.0 and pH 9.5. This drop in the recovery of refolded protein was generally consistent across the 31 refolding additives used in the screen (Figure 2.9A, B). However, there were some exceptions. Notably the additives proline and NDSB gave recoveries of soluble protein as high as 50 %, but did not give similarly high recoveries of monomeric and folded protein.



**Figure 2.10:** Recovery of soluble (white), monomeric (light grey) and folded (dark grey) protein when refolding inclusion body protein (A) and denatured soluble protein (B) in buffer alone. Mean of 3 replicates, bars show standard error of the mean.

High levels of recovery of refolded protein were observed in conditions that included no chemical refolding additive (Figure 2.10). Recoveries of approximately 10 % were observed with inclusion body protein and 20 % with native, denatured protein. These recoveries indicate that p38 $\alpha$  refolds fairly easily, as evidenced by the high number of conditions under which high recovery of refolded protein was achieved (Figure 2.9 A, B). Although p38 $\alpha$  was found to refold relatively easily at high pH, many of the refolding additives that were used enhanced the refolding of p38 $\alpha$ . Figure 2.9 C & D illustrate the effectiveness of the groups of additives at high pH. From these plots, it is clear that the detergents group is efficient in enhancing the refolding of p38 $\alpha$ , when compared to conditions lacking a specific refolding additive. In addition, there are several additives

that gave an increased recovery of refolded protein in the screen. Betaine, PEG and ethylene glycol were particular examples of these types of additives.

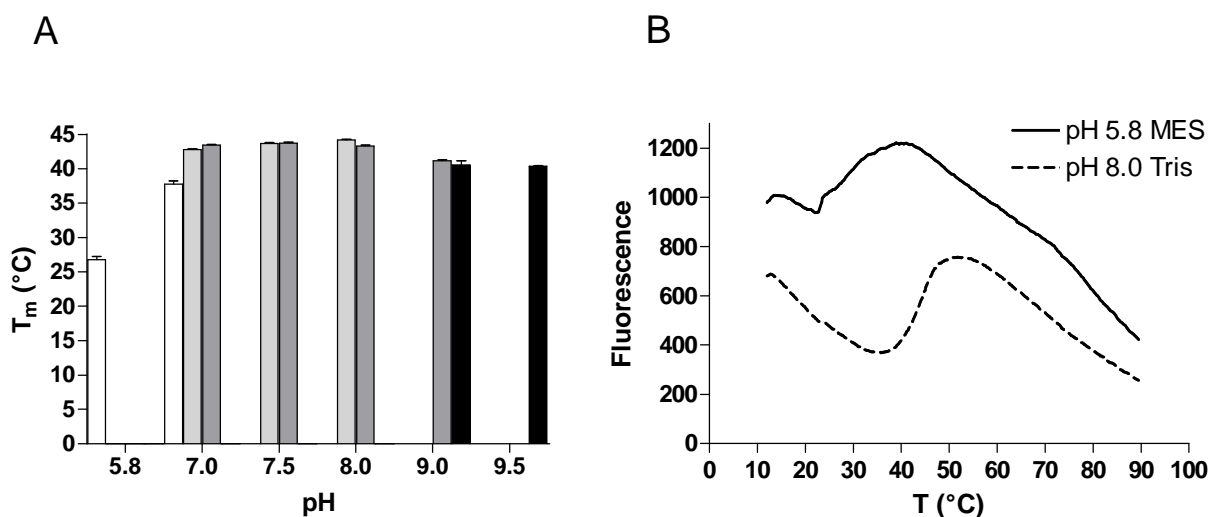
A clear difference was noted in the recoveries of soluble, folded and monomeric protein obtained when refolding inclusion body protein and native denatured p38 $\alpha$ . The recovery of refolded protein obtained with denatured soluble protein was generally approximately two fold greater than the recovery of refolded protein from inclusion body protein at high pH (Figure 2.9 A, B). This was consistent across the full range of additives. There was no difference in the recoveries obtained at low pH. Conditions of 50 mM Tris pH 8.0 resulted in particularly high recovery of soluble protein when refolding protein sourced from soluble expression in *E. coli*. This high recovery of soluble protein was not carried through to the recoveries of monomeric and folded protein (Figure 2.10B). The recoveries of monomeric and folded protein were similar to those achieved with 100 mM CAPSO pH 9.5. These discrepancies between the different readouts are probable due to misfolded soluble aggregates of protein. This high soluble recovery not seen with monomeric and folded protein underlines the need for multiple readout methods to support the results obtained. Despite lower yields (Figure 2.9), the recoveries of refolded protein obtained with inclusion body derived p38 $\alpha$  were still consistent with the production of significant amounts of protein for structural studies.

### ***2.3.7 Thermal Melting Analysis of p38 $\alpha$ at Various pHs***

As noted previously, p38 $\alpha$  refolded to a considerably higher yield under alkaline conditions than in acidic conditions. However, the change in pH was not the only change between these conditions, the buffer used had also changed. To examine whether the effect on the refolding of p38 $\alpha$  was due to the buffer or the pH alone, or a combination of both, the thermal stability of natively folded p38 $\alpha$  was examined in a series of buffers (Figure 2.11) using a thermal stability measurement, according to section 2.2.18. p38 $\alpha$  was tested in 100 mM MES at pH 5.8 and pH 7.0; 10 mM HEPES at pH 7, pH 7.5 and pH 8.0; 50 mM TrisHCl at pH 7.0, pH 7.5, pH 8.0 and pH 9.0; and 100 mM CAPSO at pH 9.0 and pH 9.5. This range of pH measurements and buffers would allow the effect of both to be examined.



The results showed that at pH 5.8 (in MES buffer) the protein has a large drop in  $T_m$ , of  $\sim 15^\circ\text{C}$ , when compared to higher pH conditions (Figure 2.11 A). p38 $\alpha$  in 100 mM MES at pH 7.0 shows a lower  $T_m$  than other buffers at the same pH,  $37.8^\circ\text{C} \pm 0.4$  as opposed to  $42.8^\circ\text{C} \pm 0.1$  and  $43.5^\circ\text{C} \pm 0.1$  (Table 2.2). This indicates that the lowering of the pH to create acidic conditions causes a major destabilisation of the tertiary structure of p38 $\alpha$ , reflected in the low  $T_m$ . The pH 7.0 conditions indicate that the 100 mM MES buffer does have a negative effect on the stability of the native state of p38 $\alpha$ , but that this effect only accounts for  $\sim 1/3$  of the observed effect at pH 5.8.



**Figure 2.11:** Thermal melting analysis of native p38 $\alpha$  in a series of buffers and pHs. (A)  $T_m$  of p38 $\alpha$  in various buffers at a range of pHs. Mean of 6 replicates, error bars show standard error of mean. Concentrations of buffers, MES – 100mM.; HEPES – 10mM; Tris – 50mM; CAPSO – 100mM. (B) Thermal melting profiles of p38 $\alpha$  in 100 mM MES, pH 5.8 and in 50 mM Tris, pH 8.0.

It can readily be seen that p38 $\alpha$  is most stable in the region of pH 7.0 to pH 8.0. Under more alkaline conditions, pH 9.0 and pH 9.5 there is a slight destabilisation of the protein, with the  $T_m$  falling from a peak of  $44.1^\circ\text{C} \pm 0.1$  at pH8.0 to  $40.4^\circ\text{C} \pm 0.1$  (Table 2.2).

**Table 2.2** T<sub>m</sub> values for p38 $\alpha$  under various pH and buffer conditions.

Buffer pH	100 mM MES	10 mM HEPES	50 mM TrisHCl	100 mM CAPSO
5.8	26.8°C $\pm$ 0.4			
7.0	37.8°C $\pm$ 0.4	42.8°C $\pm$ 0.1	43.5°C $\pm$ 0.1	
7.5		43.6°C $\pm$ 0.2	43.7°C $\pm$ 0.1	
8.0		44.1°C $\pm$ 0.1	43.3°C $\pm$ 0.1	
9.0			41.1°C $\pm$ 0.1	40.6°C $\pm$ 0.6
9.5				40.4°C $\pm$ 0.1

This indicates that the optimum pH for p38 $\alpha$  refolding, according to the thermal stability measurements, is pH 7.0-8.0. This corresponds to the results of the refolding screen, which saw that a higher yield of refolded p38 $\alpha$  was obtained at pH 8.0 compared to pH 9.5 in the absence of refolding additives.

### 2.3.8 Larger Scale Refolding

The refolding of p38 $\alpha$  was performed at a larger scale, and the refolded protein purified to allow the secondary structure content of the refolded material to be examined and the extent of refolding to an active conformation identified. The examination of the secondary structure content of the refolded protein provides evidence that the screen is identifying conditions producing correctly folded p38 $\alpha$ , and the identification of the fraction of protein which is active supports this. This refolding was carried out at a 20 mg scale by drop wise addition of p38 $\alpha$  to refolding buffer under the same conditions as in the refolding screen (Section 2.2.16). Six conditions were selected from the screen that had shown the highest recoveries of soluble, monomeric and folded protein. The conditions from the screen that were tested were 100 mM CAPSO, pH 9.5; 100 mM CAPSO, 10 % betaine, pH 9.5; 100 mM CAPSO, 10 % ethylene glycol (v/v), pH 9.5; 100 mM CAPSO, 0.04 % PEG 3440 (w/v), pH 9.5; 50 mM TrisHCl, pH 8.0; and 50 mM TrisHCl, 1 mM cyclohexanol, pH 8.0. The recoveries of protein from the scaled-up refolding were found to be similar to those obtained in the refolding screen (Table 2.3), showing that scaling up did not affect the yields obtained.

**Table 2.3:** Comparison of refolding recoveries of p38 $\alpha$  from screen runs and scaled-up experiments.

Refolding experiment	Renaturation buffer	Soluble recovery <sup>a</sup>	Monomeric recovery <sup>b</sup>	Folded recovery <sup>c</sup>
Screen	100 mM CAPSO	9 $\pm$ 2	13.0 $\pm$ 0.1	8 $\pm$ 2
Screen	100 mM CAPSO + 10 % Betaine	15.3 $\pm$ 0.7	11.0 $\pm$ 0.7	15 $\pm$ 1
Scaled-up	100 mM CAPSO	13	10	9
Scaled-up	100 mM CAPSO + 10 % Betaine	15	13	13

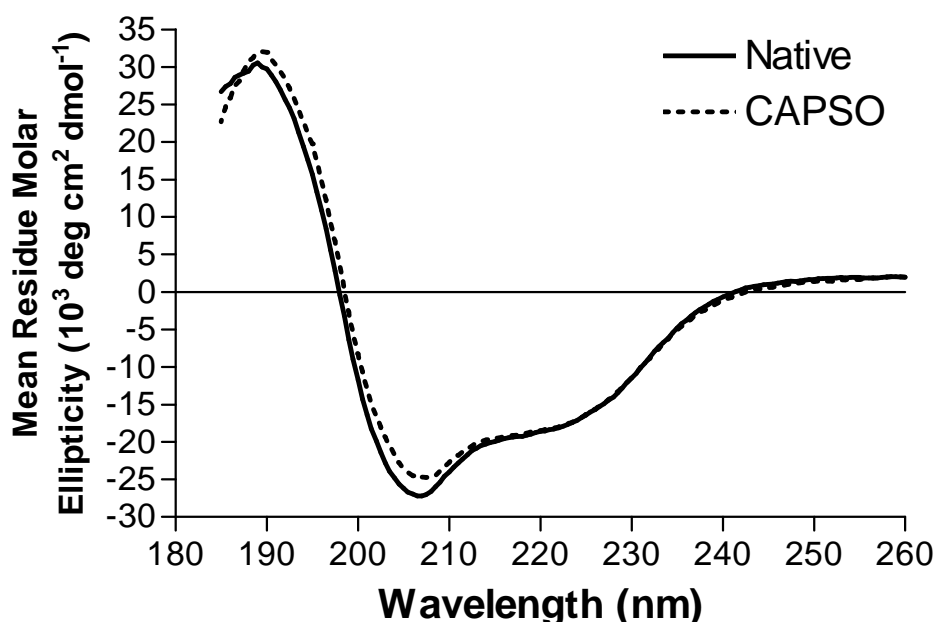
<sup>a</sup>Soluble protein recovery was assessed by capillary electrophoresis; <sup>b</sup>Monomeric protein recovery was measured by analytical size exclusion chromatography; <sup>c</sup>Folded protein was evaluated by binding activity via a surface plasmon resonance method. Protein recoveries are expressed as a % of total initial protein. Refolding screen recoveries are mean of 3 experiments  $\pm$  SE of mean.

The refolded material was further purified after refolding by size exclusion chromatography to remove soluble aggregates. The resultant, purified, refolded protein was analysed to determine the secondary structure content, and binding activity. The refolding of p38 $\alpha$  from inclusion bodies at a 20 mg scale was also used to calculate the yield of soluble, correctly folded protein that could be achieved by this method. An average yield of 13 % monomeric, folded protein was obtained from the method used. This corresponded to a recovery of monomeric, folded protein of ~ 50 mg protein per litre of expression culture. This final yield was comparable with the yield of expression achieved when p38 $\alpha$  was expressed in a soluble form.

### 2.3.9 Far-UV Circular Dichroism of Native and Refolded p38 $\alpha$

The far-UV CD spectra of natively folded p38 $\alpha$  and p38 $\alpha$  refolded from inclusion bodies in the presence of 100 mM CAPSO, 10 % betaine (w/v), pH 9.5 (after dialysis, see section 2.2.16) were identical (Figure 2.12), indicating that the secondary structure content of the natively folded and the refolded protein was similar. The other conditions that were scaled up also showed similar far-UV CD spectra (data not shown), indicating that the secondary structure content of these refolded protein samples was also similar to

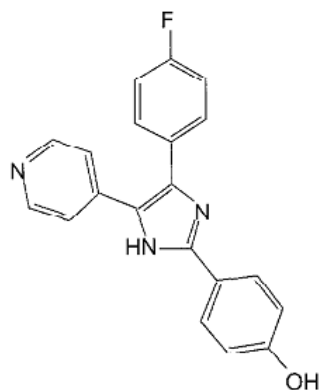
the native state. This shows that the screen is able to identify conditions producing correctly folded protein.



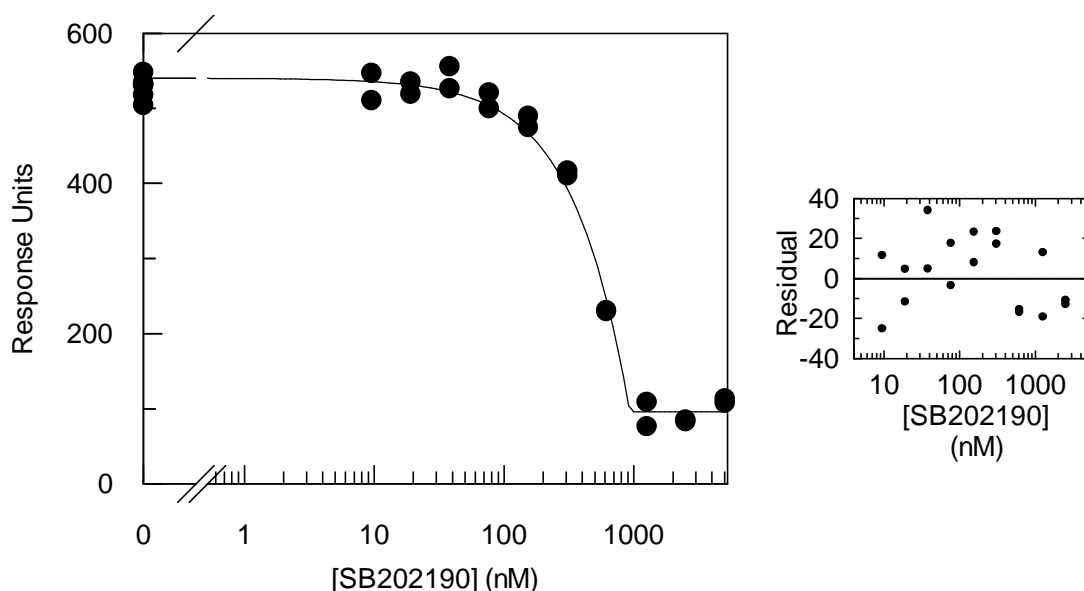
**Figure 2.12:** Circular dichroism spectra of natively folded p38 $\alpha$  and inclusion body derived p38 $\alpha$  refolded in 100 mM CAPSO, 10 % Betaine, pH 9.5. Spectra were collected at 20 °C, and are the average of eight scans.

#### 2.3.10 Binding of Refolded p38 $\alpha$ to SB202190

The binding of p38 $\alpha$  to a known p38 $\alpha$  inhibitor, SB202190, was used to test the activity of the refolded protein. A modification of the surface plasmon resonance method (section 2.2.15). Refolded p38 $\alpha$  was applied to a CM5 chip (Biacore) which had been functionalised by binding the UTDC used previously to the chip via an amine coupling reaction. P38 $\alpha$  was injected in the presence of increasing concentrations of SB202190 (Figure 2.13), a commercial p38 $\alpha$  inhibitor that is known to bind to the ATP binding site in the unphosphorylated form of p38 $\alpha$  with a  $K_d$  of 38 nM (Frantz *et al.*, 1998). The response units at each concentration were measured and the data fitted to a standard stoichiometric binding model (Figure 2.14). The protein concentration was 1  $\mu$ M.



**Figure 2.13:** Structure of SB202190 used to assess the activity of p38 $\alpha$  refolded on a 20 mg scale. Figure adapted from Wilson *et al.* (1997).



**Figure 2.14:** Inhibition of the binding of refolded p38 $\alpha$  to immobilized UTDC by SB202190 detected with SPR. Data fitted to a stoichiometric binding model, with response unit background at 96 and residuals shown on the right.

The results showed that a high proportion of the binding sites present were active. 92.7%  $\pm$  4 of the theoretically present active sites were found to be active. This indicates the refolding procedure used has produced p38 $\alpha$  in which the ATP binding site is correctly formed in the majority of the refolded sample. This result complements the earlier data on the secondary structure content, indicating that the refolded p38 $\alpha$  is correctly folded.

## ***2.4 Discussion***

The refolding of protein kinases is a challenging process with little reported success in the production of protein suitable for structural studies. Overcoming this challenge is likely to require the screening of a large number of conditions, in order to identify those which allow refolding of the protein. It is also critical that the analytical methods used to quantify refolded protein are robust and sensitive.

The refolding of proteins generally requires a low protein concentration to reduce aggregation; this means that the analytical techniques used must be highly sensitive. Methods were chosen which had low limits of detection, corresponding to minimum useful yields for refolding protein for structural studies. For example, the limit of detection for the staining of SDS-PAGE gels corresponds to a refolding yield of 1.25 %.

The refolding screen makes use of a number of refolding additives, some of which are present at high concentrations. These additives may interfere with particular analytical methods. For example, additives such as guanidine, which cause large changes in the refractive index of the sample, interfere with SPR measurements. Detergents also cause large baseline deflections for denaturing capillary electrophoresis and analytical size exclusion chromatography. The analytical methods selected were chosen to address these concerns. An initial SDS-PAGE screening method was selected that is relatively insensitive to the additives used, and the refolding additives were also removed by dialysis into a neutral buffer.

A hierarchical analysis design was used to eliminate unproductive conditions early in the process (Figure 2.8) by examining total soluble protein content following refolding and dialysis. Refolding conditions with a low recovery of refolded protein were first eliminated by SDS-PAGE analysis of the refolded protein. Capillary electrophoresis was also used to eliminate conditions with low recoveries of refolded protein subsequent to concentration and dialysis, with conditions with recoveries of less than 5 % being eliminated to allow manageable numbers of samples to be progressed to the final two analytical methods.

A series of analytical methods for analysing the refolded protein obtained from the screen were identified, and these methods were shown to have low limits of detection.

We showed that ePAGE analysis of the refolded protein was able to identify yields of soluble protein in the presence of refolding additives of as least as low as 1.25 % (Figure 2.3). Analytical size exclusion chromatography was shown to be able to quantify the recovery of monomeric protein to a limit of 50 µg/mL (Figure 2.5). Binding analysis by SPR of refolded p38α was shown to be able to identify the yield of folded p38α to a limit of 1 % (Figure 2.7). These limits of detection are expected to be compatible with screening of kinases with lower yields of refolding than those shown by p38α.

Strong differences in the yields of refolding of inclusion body protein and denatured soluble protein were identified at higher pH values (Figure 2.9 A, B). The reasons for this difference are unclear. There is no discernible difference in the monomeric state of inclusion body protein and native denatured protein in 8 M urea that can be identified by analytical size exclusion chromatography (Figure 2.2). This demonstrates that there is no difference in the aggregation states of the denatured protein from both sources. Some amount of secondary structure is known to be transiently present in the unfolded state of proteins in high concentrations of chaotropic denaturants, such as guanidine and urea (Dill and Shortle, 1991). However, this is primarily sequence dependent and would be expected to be the same in both protein preparations (Dill and Shortle, 1991). It is noteworthy, that no additional contaminant proteins were identified by SDS-PAGE of inclusion body protein and soluble protein (Figure 2.1).

The refolding screens carried out identified that there is a strong dependence of the refolding ability of p38α on pH. At the low pH used in the screen, the refolding of p38α was inefficient, with low yields of soluble protein identified by capillary electrophoresis (Figure 2.9 A, B). The thermal melting results showed that the inefficiency of refolding at lower pH values is due to a destabilisation of the protein's native structure in low pH buffers (Figure 2.11, Table 2.2). This destabilisation is likely to be due to changes in the ionization of residues in the protein, since the pI of p38α is around pH 5.5. It is not known if this effect is common to serine/threonine protein kinases or is a specific effect for p38α. However, a low pI is not a common feature of all protein kinases, with several related protein kinases having calculated pIs of at least 8.5.

Conditions identified by the screen leading to comparatively high recoveries of refolded protein were used to refold protein at a higher scale than was performed in the

screen to provide sufficient protein for further analysis and to identify if there were significant differences in the refolding yields obtained at a larger scale. The refolded protein from this higher scale refolding experiment was used for the structural analysis of the refolded protein. It was shown that the far-UV CD spectra of control natively folded p38 $\alpha$  and refolded p38 $\alpha$  were similar (Figure 2.12), with a calculated secondary structure content analogous to the published structure of p38 $\alpha$  (Shortle and Ackerman, 2001). This indicates that the protein obtained from refolding inclusion bodies is correctly folded. The similarity of the refolding yields obtained from the refolding screen and from the larger scale refolding performed demonstrates that the refolding screen is capable of identifying refolding conditions that are transferable to high scale refolding for the production of refolded protein for structural studies. This transferability is key for the screen to be useful for identifying refolding conditions for other, more challenging protein kinases that are produced in an insoluble form in *E. coli*.

A screening system for the refolding of a model protein kinase, has been established. This screen has identified and described a series of specific analytical methods that quantify the yields of refolded protein, identifying the oligomeric state of the protein and whether the refolded protein has adopted the correct fold. The yield of refolded protein depends strongly on the pH at which the protein is refolded, and the source of the protein to be refolded also strongly affects the yield of refolded protein. In addition, we have shown that refolding can be performed at larger scale, resulting in correctly folded protein without reducing the recovery yields.



## **Chapter 3. Application of the Refolding Screen to Additional Protein Kinases and Refinement of the Screening Procedure**

### ***3.1 Introduction***

The screen described in the previous chapter has been shown to be effective in detecting and improving the yield of refolding of the model protein kinase, p38 $\alpha$ . However, examining the refolding of only a single protein kinase in this fashion does not give much insight into the refolding of protein kinases in general, nor does it allow the screen to be properly described as being a generic screen for the refolding of protein kinases. In order to address these shortcomings it is necessary to test the refolding screen with several other protein kinases with varying degrees of ease of soluble expression, and whose kinase domains have been isolated from the context of a larger, multi domain protein.

The human kinome contains, at the current count, 527 protein kinases, which can be split into several distinct classes based on shared sequence features. The analysis of Manning *et al.* (2002) identifies seven distinct groupings of the human kinome. These seven groups are the tyrosine kinase group; the tyrosine kinase-like group; the STE group (homologues of yeast sterile 7, 11 and 20 kinases); the CMGC group which contains the CDK, MAPK, GSK3, and CLK families; the CAMK group containing calcium and calmodulin dependent kinases; the AGC group containing the PKA, PKG and PKC families; and finally the CK1 group containing Casein Kinase 1 and closely related kinases. There are other kinases in the kinome which do not fit into one of the defined groups. For the purposes of the kinase panel that was to be constructed to test the refolding screen with a wider array of proteins, the Tyrosine Kinase and Tyrosine Kinase like group were not considered.

The initial screen developed in chapter 2 examines a large number of refolding additives, and several pHs, but it does not examine the interaction of the effect of multiple additives on the refolding of the protein. This was by design, since the interaction of selected additives could be examined at a later stage. However, the screening of p38 $\alpha$  revealed that a number of the additives used in the screen had no

effect, or had a detrimental effect on the folding of the model kinase. The factors that gave rise to no effect or a detrimental effect on the refolding of the model kinase, and the expanded kinase panel, could be eliminated from the screen, and the remaining additives formatted into a factorial screen to examine the interactions of the additives selected.

A powerful feature of factorial screens is that they are readily analysed using statistical techniques, leading to the rapid and robust identification of the key factors that affect the result of the experiments being performed. The conditions which are tested for their effect on the outcome are termed factors. A powerful feature of factorial screening is that factors can have a number of discrete values in the screen, although simple screens will adopt only two values for each factor. The number of experiments required for a factorial screen can be very large however. If four factors are used, each having four levels, for example four different folding additives at four different concentrations, then the number of experiments required would be the number of levels for each factor, to the power of the number of factors, or  $4^4$  or 256 experiments. For a more simple screen in which factors are only present or absent, for any number of factors the number of experiments required is  $2^n$ , where  $n$  is the number of factors, so 4 factors, 16 experiments, 5 factors, 32 experiments and 6 factors, 64 experiments.

A feature of the statistical mathematics used in interpreting the factorial screen results is that it can be used to estimate the effects of missing combinations of factors from other experiments performed. This approach is called a fractional factorial screen. In such a screen the number of experiments is reduced at the expense of the resolution of the screen, and the statistical tests performed on the results are used to “fill-in” the missing data. This approach allows the size of the screen to be kept reasonable with higher numbers of factors. However, care must be taken to not reduce the number of experiments required too much, or the resolution of the screen will suffer, and the results will be less reliable. A balance must therefore be sought between the resolution of the screen and the numbers of experiments required.

This chapter describes the selection of four additional kinases from the kinome to form a kinase panel for testing the refolding screen. The kinases selected

were AKT2, KIS, phosphorylase kinase and TTK. These kinases were refolded through the previously described screen and the results compared between the kinases. Effective additives were selected and combined into a fraction factorial screen which was applied to the kinase panel and the model kinase p38 $\alpha$ . The results were analysed using statistical methods and the key factors and combinations of factors analysed to identify the commonalities between the refolding of the tested kinases. The differences between the kinases were also elucidated.

## **3.2 Methods and Materials**

### **3.2.1 Materials**

NV-10 was purchased from Novexin Ltd (Cambridge, UK). 3-(1-pyridino)-1-propanesulfonate and  $\beta$ -cyclodextrin were purchased from Fluka (Buchs SG, CH). Tris was purchased from Acros Organics (Geel, BE); P20 surfactant was supplied by Biacore (Chalfont St. Giles, UK) and dimethyl sulphoxide by Fisons (Ipswich, UK). All other chemicals were supplied by Sigma-Aldrich (Poole, UK). Plasmids for the expression of AKT2, KIS(1-313) and TTK (514-804) were supplied by AstraZeneca. The creation of a synthetic PhK construct was performed by Geneart (Regensburg, DE).

### **3.2.2 Gateway Cloning**

The Gateway cloning system was used to clone the synthetic PhK construct into a plasmid vector suitable for expression in *E. coli*. Two separate recombination reactions, the BP reaction and the LR reaction are used to recombine the DNA of interest into vectors suitable for either site directed mutagenesis, or expression of the protein. The recombination reaction depends on specific sequence elements present in the donor and the acceptor sequences. The recognition elements are also specific to the 5' and 3' end of the element to be recombined. There are four types of sites, B, P, L and R sites. B sites are recombined specifically with P sites, and L sites are specifically recombined with R sites. The recombination between B and P sites generates L sites, and the recombination between L and R sites generates B sites. The recombination between the sites is mediated by specific recombinases, termed

Clonase enzymes. The Clonase enzymes catalyse the recombination between either B and P sites, or between L and R sites, being termed BR Clonase or LR Clonase depending on the reaction catalysed.

#### *3.2.2.1 Gateway Cloning – BP Reaction*

Inserts containing *attB* sites were recombined into vectors containing *attP* sites, either for the purposes of cloning, or in order to perform site directed mutagenesis. Two units of BP clonase (Invitrogen) were added to 75 ng of source vector containing insert flanked by *attB* sites and 75 ng of circular pDONOR221. The reaction mixture was diluted to a final volume of 5 µL with water. The reaction was incubated for 1 hour at 25 °C at 300 rpm in a heated shaking block. 2 µL of the incubated mixture was used to transform *E. coli* strain DH5α (Section 2.2.2) competent cells (Invitrogen) which were plated on LB Agar plates containing Kanamycin. Overnight cultures were processed using a Spin Miniprep Kit (Qiagen), and the isolated plasmid DNA checked for incorporation of the insert by a BSRG1 digest

#### *3.2.2.2 Gateway Cloning – LR Reaction*

To produce expression plasmids, entry vectors produced by the BP reaction, and confirmed as containing the insert of interest by BSRG1 digestion, were recombined with expression vectors modified to contain *attR* sites. Two units of LR Clonase (Invitrogen) were added to 75 ng of entry vector and 75 ng of circular expression vector, usually pT7#3.3. The reaction mixture was diluted to a final volume of 5 µL with water. The reaction was incubated for 1 hour at 25 °C at 300 rpm in a heated shaking block. 2 µL of the incubated mixture was used to transform *E. coli* strain DH5α (Section 2.2.2) competent cells (Invitrogen) which were plated on LB Agar plates containing an appropriate antibiotic. Overnight cultures were processed using a Spin Miniprep Kit (Qiagen), and the isolated plasmid DNA checked for incorporation of the insert by a BSRG1 digest.

### ***3.2.3 Production of Kinase Panel Inclusion Bodies***

Protein kinase domains were prepared from inclusion bodies using the following procedure. Expression constructs for the protein kinase of interest were used to transform *E. coli* strain BL21\* (DE3) cells according to section 2.2.2. The agar plates were incubated overnight at 37 °C. For expression, a single colony was picked from the plate and a 75 mL starter culture of Terrific Broth (Section 2.2.4) was inoculated with this colony. The starter culture was incubated overnight at 37 °C in a shaking incubator at 180 rpm.

5 mL of starter culture was used to inoculate expression cultures of 600 mL terrific broth and incubated in a shaking incubator at 37 °C and 180 rpm. The OD<sub>600</sub> was monitored, and at OD<sub>600</sub>  $\approx$  0.5 the temperature was reduced to 20 °C. When OD<sub>600</sub> reached  $\sim$ 0.6 the expression of the protein kinase was induced by the addition of IPTG to a final concentration of 0.1 mM. The incubation was allowed to proceed for 20 hours before the biomass was harvested. The biomass was harvested by centrifugation at 6,000 g for 20 minutes at 4 °C. The inclusion bodies were isolated according to the procedure laid out in section 2.2.5.

### ***3.2.4 Refolding of the Kinase Panel using a Refolding Screen***

The refolding of the kinase panel proteins was carried out as described previously for p38 $\alpha$  in chapter 2. Briefly, 100  $\mu$ L of concentrated, denatured inclusion body protein at 5 mg/mL was aliquoted into 5 mL of various renaturation buffers and allowed to refold overnight at 4 °C (Section 2.2.10). The solutions were then tested for soluble protein (Section 2.2.11), concentrated 10 fold and dialysed into a neutral buffer. The extent of refolding was then determined by three procedures; capillary electrophoresis (Section 2.2.13), analytical size exclusion chromatography (Section 2.2.14) and surface plasmon resonance based binding experiments (Section 2.2.15). These procedures are fully laid out in chapter 2.

### **3.2.5 Fractional Factorial Screens for the Refolding of Protein Kinases**

Selected conditions from the broad refolding screen were used to construct a fractional factorial screen for the refolding of protein kinases. The screen was constructed using DesignExpert 7 software and was carried out as laid out above for the initial refolding screen. The collected data were analysed using the data analysis functions of the DesignExpert 7 software.

### **3.2.6 Cumulative Probability of Studentised Residuals**

The cumulative probability of studentized residuals was calculated by DesignExpert 7 according to the equation:

$$P(i^{th} residual) = \frac{i}{(N + 1)} \quad (3.1)$$

Where P is the cumulative probability of a point,  $i$  is the order of the value in an ordered list of the residuals and  $N$  is the number of residuals in the list.

## **3.3 Results**

### **3.3.1 Selection of the Kinase Panel**

The selection of the kinase domains to be used in the kinase panel to test the screen is an important part of the testing of the screen and to the application of the results to identify similarities in the refolding of protein kinases. If the protein kinases selected were too closely related then this could lead to false positives when considering the similarities observed in the refolding. To avoid closely related kinases being selected, the expanded dendrogram poster of Manning *et al.* (2002) was used (expanded dendrogram available online at <http://kinase.com>). The kinases were chosen from different areas on the dendrogram to avoid close sequence relationships. Four kinases were chosen, to form a total kinase panel of five kinases with the inclusion of the previous data collected on p38 $\alpha$ .

The first consideration of the kinase panel was to include a protein kinase which is known to refold from inclusion bodies. This would provide extra support for the results obtained with p38 $\alpha$  showing that the refolding screen is able to identify refolded protein. For this purpose the kinase domain of Phosphorylase kinase was selected. Owen *et al.* (1995) had refolded this protein from inclusion bodies in order to produce soluble protein which was crystallised and the structure solved from those crystals. Phosphorylase kinase (PhK) is a multi domain protein, and the functional protein is composed of several subunits. The  $\gamma$ -subunit contains the catalytic kinase domain, with the other subunits being involved in substrate recognition and control of the catalytic activity. The isolated kinase domain is constitutively active. Phosphorylase kinase is a member of the CAMK family and is not closely related to p38 $\alpha$ . The expression construct for this protein was created by gene synthesis, and cloned into an expression vector as described in section 3.3.2.

Although tyrosine kinase and tyrosine kinase like (TK and TKL) kinase domains had been excluded from the kinase panel, the isolated kinase domain of TTK was included as it is a dual specificity kinase, and it was thought that changes in the active site required to accommodate the tyrosine side chain might result in differences in the folding of the kinase domain. The TTK kinase domain has been isolated from the context of a much larger multi-domain protein which regulates the activity of the kinase domain as well as performing other functions. The isolated TTK kinase domain is constitutively active, like the kinase domain of PhK. TTK does not fit into one of the families of protein kinases in the analysis of Manning *et al.* (2002), but is found on the arm of the dendrogram which terminates in the CMGC group, which contains p38 $\alpha$  (Figure 1.4). An expression construct for TTK (514-804) was supplied by AstraZeneca (Section 3.2.1). This expression construct consisted of a codon optimised expressed sequence contained in a pT7#3.3 vector.

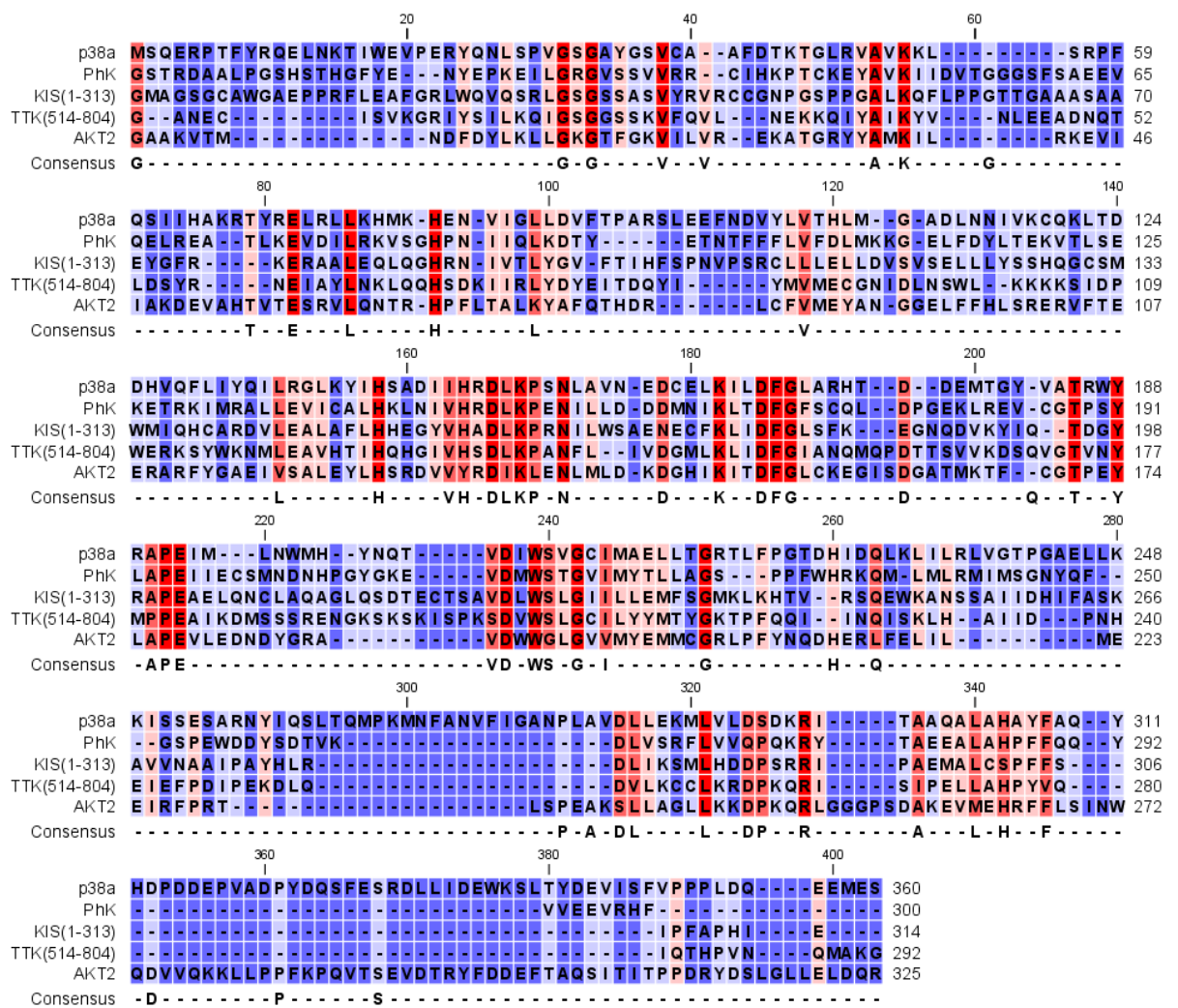
The isolated kinase domain AKT2 or PKB $\beta$  was included as a representative of the AGC group which comprises the protein kinase A, G and C families. These kinases have a interesting sequence feature. As described, previous work on the equilibrium folding of p38 $\alpha$  had identified an absolutely required, conserved core tryptophan, W207. In the protein kinase B family and other AGC family members there is a double tryptophan motif in the core of these kinases. This raised the possibility of different folding pathway driven by the double tryptophan motif in the

core of the protein. The AGC group also contains many important kinases, so the inclusion of a member of this group was important for representing the kinase domain. The expression construct for AKT2 was also supplied by AstraZeneca (Section 3.2.1).

The final kinase selected to form this kinase panel was included as a kinase regarded to be a challenging kinase to produce in a soluble form. KIS (kinase interacting with stathmin) is a two domain protein which has yet to have its structure solved and limited information on the *in vivo* function of the kinase is available. The two domains comprise a kinase domain at the N-termini of the protein and a second C-terminal domain whose function is currently unknown, but based on sequence details is suspected to be an RNA binding domain. The production of the kinase domain in KIS is difficult and has yet to be achieved in significant amounts. Soluble protein can be generated as a GST fusion; however upon removal of the GST domain, the protein aggregates indicating that the protein was mis-folded, or partially folded (AstraZeneca unpublished data). If the protein were able to be produced in a soluble, folded form from a refolding screen this would be especially advantageous as it could lead to the possibility of a crystal structure of the protein. The Manning *et al.*,(2002) analysis indicates that KIS is closely related to TTK. Despite this feature, and the possibility of similarities between the kinases being due to their close relationship, the very different behaviours of the two proteins, and the lack of structural information for KIS, it was included as an interesting target for refolding. The KIS kinase domain expression construct, representing residues 1-313 of the full length protein, was supplied by Astrazeneca in a pT7#3.3 vector (Section 3.2.1).

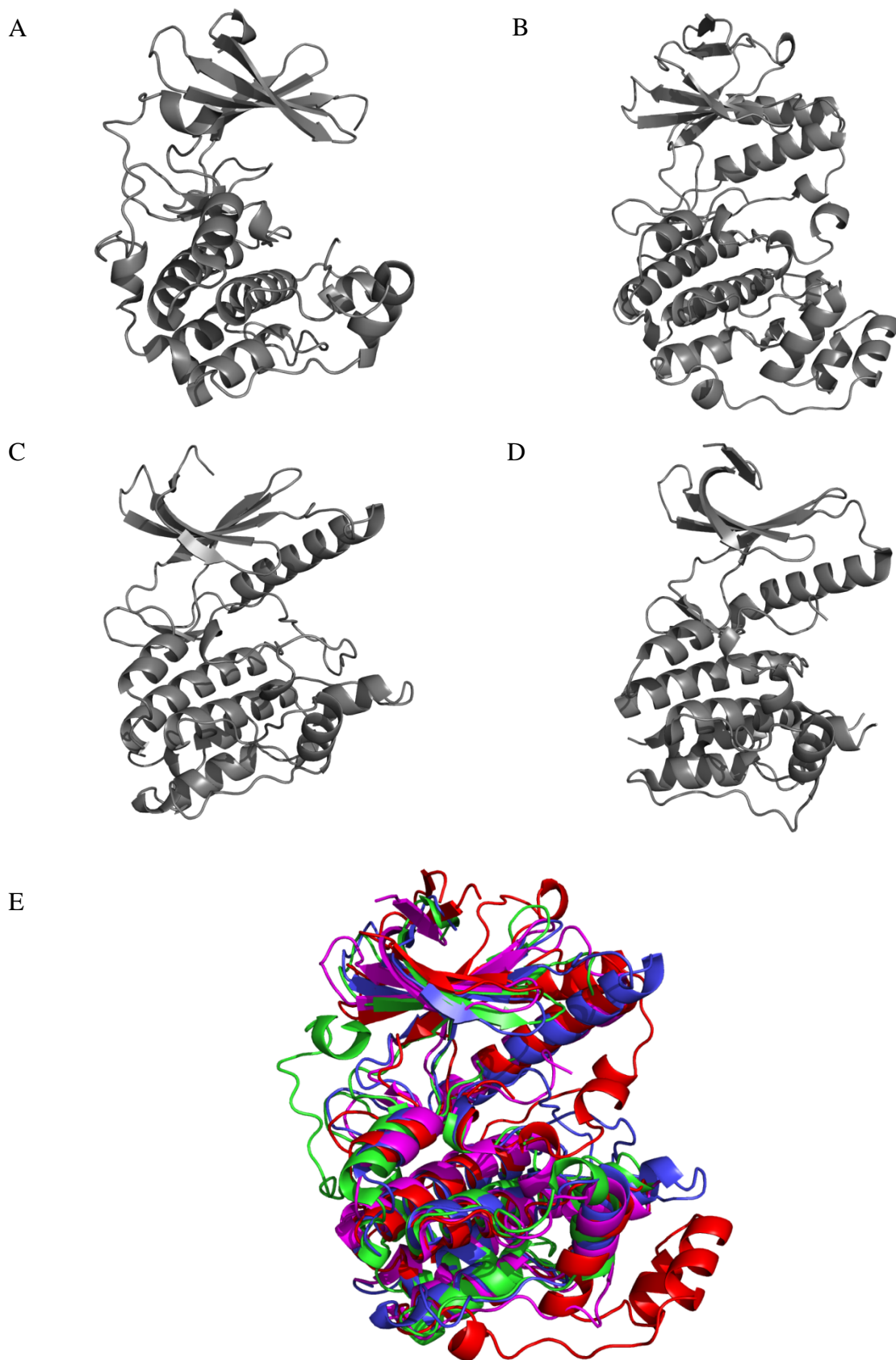
The proteins which comprised the kinase panel share certain sequence features which are common to the kinase fold. However, overall sequence identity is low. A sequence alignment of the five members of the kinase panel identifies little common sequence (Figure 3.1). The sequence features identified as common, such as the DFG motif are mostly related to the catalytic function of the proteins. The conserved tryptophan can however be seen.





**Figure 3.1:** Sequence alignment of the kinase domains of the members of the kinase panel. Consensus sequence indicates residues at least 80% conserved. Key kinase domain motifs such as the DFG motif and the conserved core tryptophan can be noted. Alignment performed using CLC Bio Sequence Viewer, using proprietary pairwise alignment function, utilising default parameters. Background colour indicates degree of residue conservation between kinases from blue to red; unique residues to completely conserved residues.

Although the members of the kinase panel show low sequence homology the crystal structures of the kinase domains which have been solved for all of the members of the panel apart from KIS, show a common kinase fold (Figure 3.2).



**Figure 3.2:** Crystal Structures of the members of the kinase panel. (A) – crystal structure of AKT2. (B) – crystal structure of p38 $\alpha$ . (C) – Crystal structure of PhK. (D) – Crystal Structure of TTK. (E) – Alignment of the members of the kinase panel to p38 $\alpha$ . p38 $\alpha$  shown in red, AKT2 in green, PhK in blue and TTK in magenta. All images produced using ray tracing module of Pymol (Delano, 2008). Alignment performed in the same software.

The alignment of the kinase domains indicates that the members of the kinase panel have very similar folds. p38 $\alpha$  shows several additional  $\alpha$ -helices in both the N-terminal lobe and the C-terminal lobe. There is more variability in the arrangement of the  $\beta$ -sheets of the N-terminal lobe than in the  $\alpha$ -helices of the C-terminal lobe. From the sequence alignment and the structural alignment of the proteins selected for the kinase panel it is clear that the selected proteins are sequence diverse but structurally similar. The proteins of the kinase panel have a diverse set of physical properties. A selected set of these properties are summarised in table 3.1

**Table 3.1:** Selected physical properties of the members of the kinase panel constructs. pI calculated according to Bjellqvist *et al.*, (1993).

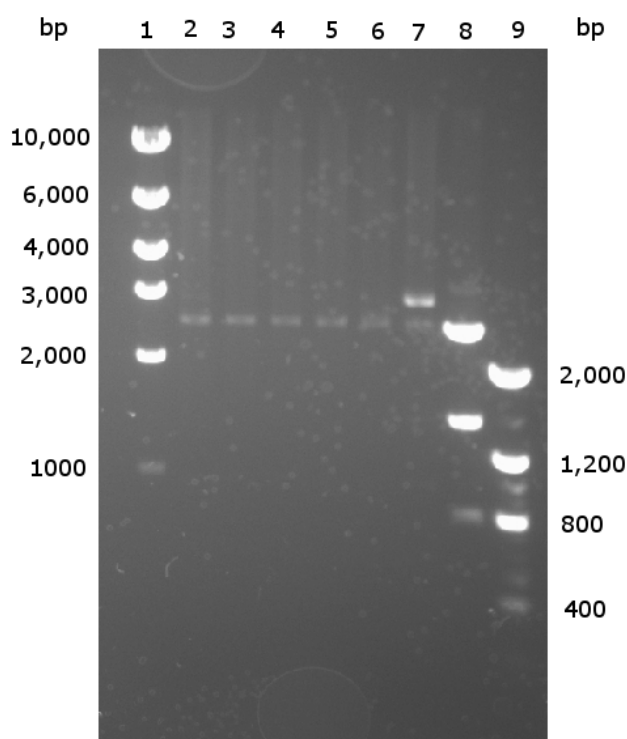
Kinase	# of Residues in Kinase Domain	# of Residues in Construct	Calculated Mass of Construct	Calculated pI of Construct	N-Terminal 6His tag
AKT2	340	354	41.3 kDa	5.65	Y
KIS	313	339	37.5 kDa	6.65	Y
p38 $\alpha$	360	360	41.3 kDa	5.48	N
PhK	299	326	37.5 kDa	6.06	Y
TTK	290	315	37.7 kDa	8.42	Y

### 3.3.2 Production of an Expression Construct for PhK Kinase Domain

Based on the studies of Owen *et al.* (1995) which determined the crystal structure of the rabbit Phosphorylase Kinase kinase domain, the equivalent residues from the human PhK $\gamma$ 1 subunit were identified. This sequence was made into a synthetic gene optimised for expression in *E. coli* by Geneart (Section 3.2.1). The synthetic gene was supplied in a vector not suitable for expression and was cloned into a suitable expression vector using the Gateway cloning system (Section 3.2.2)

The cloning of the coding unit for PhK was a two step process. First, the insert supplied by Geneart was cloned into a subcloning vector, from which it was cloned into the expression vector of choice. The Gateway cloning system does not utilise restriction enzymes, but instead depends on site specific recombinases to exchange the inserts between their recognition sites, altering the recognition site in

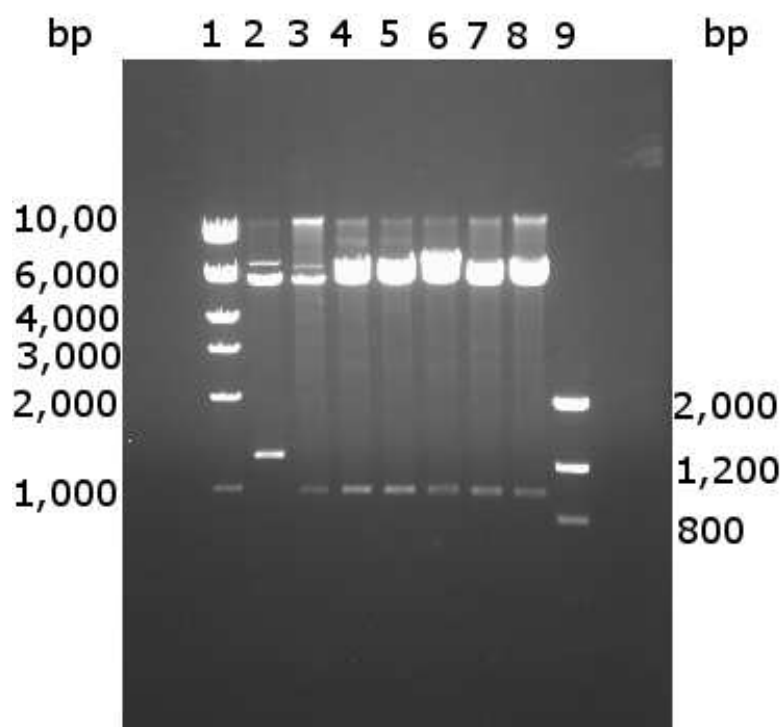
the process, preventing the back reaction. The initial reaction is termed the BP reaction and moves the insert from the supplied vector into an entry vector, in this case pDONOR221. The reaction was carried out using the synthesized gene according to the procedure in section 3.2.2.1. The BSRG1 digestion of the new plasmids generated from the BP reaction (Section 3.2.2.1) which had been produced by incubation in *E. coli* strain DH5 $\alpha$  was used to determine if the insert had been moved into the entry vector. Due to the nature of the Gateway system it is not possible for the insert to be inserted in the wrong orientation. The incorporation of the insert abolishes a BSRG1 restriction site present in the unreacted vector, resulting in a shift in the band pattern and the loss of a difficult to visualise, low mass band. The digestion of the plasmids generated from the BP reaction is shown in Figure 3.3.



**Figure 3.3:** BSRG1 digest of BP reaction produced plasmid putatively containing the PhK coding insert. Lane 1 – High range markers; Lane 2-7 – BSRG1 digests of BP reaction minipreps; Lane 8 – BSRG1 digest of unreacted pDONOR221, Lane 9 – low range markers. Expected fragment sizes for unreacted pDONOR221 : 2514 bp, 1458 bp and 790 bp. Expected fragment sizes for pDONOR221 with PhK coding region incorporated: 2514 bp and 965 bp.

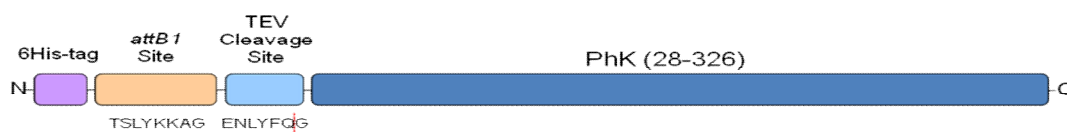
The digestion of the minipreps products of the BP reaction do not show the band patterns expected. However, the concentration of DNA used was low, and the highest size band is very weak. Despite the lack of visible bands allowing the confirmation of the success of the BP reaction, the plasmids produced were carried forward to the LR reaction. This is because the incorporation of the insert into the pDONOR221 plasmid changes the sequence of the recognition site, allowing its recognition by the LR clonase and preventing its recognition by the BP clonase. Therefore, any unreacted pDONOR221 which is carried through would not be recognised by the LR clonase enzyme, and so recombination would not occur, and this would be detected by a cleavage pattern which would be the same as the unreacted vector in the LR reaction minipreps.

The LR reaction is used to recombine the insert into a final expression vector. The system is modular, allowing the incorporation of the insert into many different vectors for expression with different purification tags and for expression in different expression systems, e.g. *E. coli*, insect cells and mammalian cell culture. The DNA from the minipreps performed on the LR reaction product were digested with BSRG1 to check for the incorporation of the insert into the selected pt7#3.3 N6his vector which had been selected as the expression vector (Figure 3.4).



**Figure 3.4:** BSRG1 digest of PhK minipreps following LR reaction. Lane 1 – Markers, Lane 2 – Unreacted pT7#3.3. N6His, Lane 3-8 – BSRG1 digests of LR reaction minipreps; Lane 9 – low weight markers. Expected fragment sizes for unreacted pt7#3.3 : 5647 bp, 1283 bp and 402 bp. Expected fragment sizes for pT7#3.3 with PhK coding region incorporated: 5648 bp and 965 bp.

The BSRG1 digest indicates the PhK coding region was successfully incorporated into the expression vector, despite the incorporation of the insert into pDONOR221 after the BP reaction not being confirmed. The expression construct that was created contained an N-terminal 6 histidine tag, the translation of the *attB* recombinase recognition site, and TEV protease cleavage site and the sequence of the PhK kinase domain (Figure 3.5).



**Figure 3.5:** Diagrammatic representation of the PhK gamma subunit kinase domain expression construct used. Major elements are shown, namely the N-terminal 6 histidine tag, *attB1* site, TEV protease recognition site, and the PhK gamma subunit kinase domain, residues 28-326 from the full length Human PhK $\gamma$  sequence. Amino acid sequences of the *attB1* site and TEV protease recognition site are shown in single letter code. Cleavage point by TEV protease indicated by vertical dotted red line.

### 3.3.3 Expression of Kinase Panel Proteins as Inclusion Bodies

The plasmids encoding the proteins of the kinase panel were transformed in *E. coli* strain BL21\* (DE3) according to the procedure outlined in section 2.2.2. The proteins were then expressed as inclusion bodies, and purified according to sections 3.2.3 and 2.2.5. In initial experiments the inclusion bodies were produced by the induction of expression with a higher concentration of IPTG, and the expression was performed at a temperature of 37 °C, similar to the procedure used for the production of p38 $\alpha$  as inclusion bodies (Section 2.2.4). This procedure, however, resulted in inclusion body preparations which had low purity with several contaminant proteins present in the inclusion bodies at a high abundance. All of the kinase panel proteins were produced with poly histidine tags, and so it would have been possible to purify the proteins from the crude inclusion body preparations by denaturing immobilised metal affinity chromatography; however, an alternative approach was taken which was simpler and reduced the time taken for the expression and isolation of the inclusion bodies. The expression of the kinase panel proteins was performed at a lower temperature (20 °C). The kinase domains still accumulated as inclusion bodies; however the quality and purity of the inclusion body preparations was dramatically improved (data not shown).

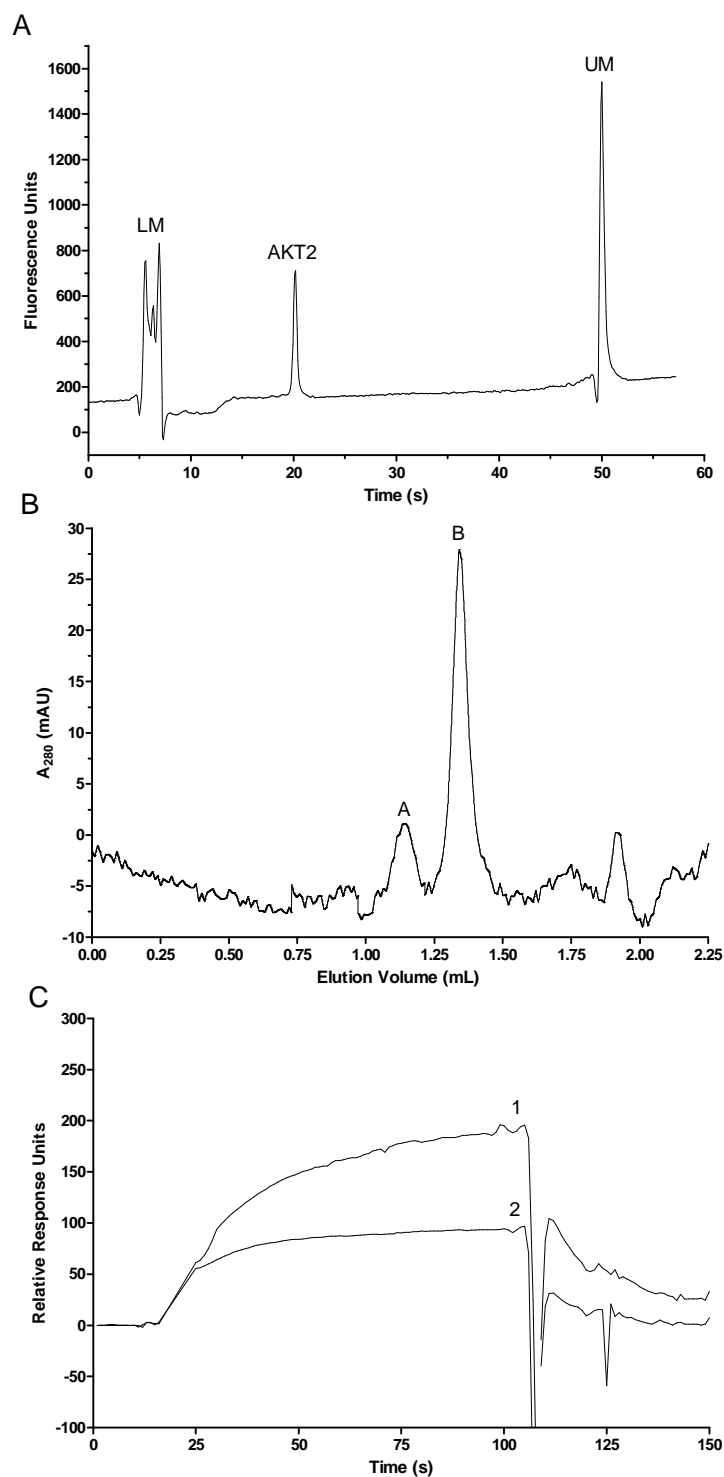
The identity of the major component of the inclusion body preparations performed was confirmed by ESI-MS analysis of the inclusion body preparations, according to the procedure outlined in section 2.2.8. The measured masses of the inclusion body proteins obtained were in good agreement with the expected

molecular masses calculated from the sequence of the proteins derived from *in silico* translations of the plasmid DNA sequences (data not shown).

#### ***3.3.4 The Refolding of AKT2***

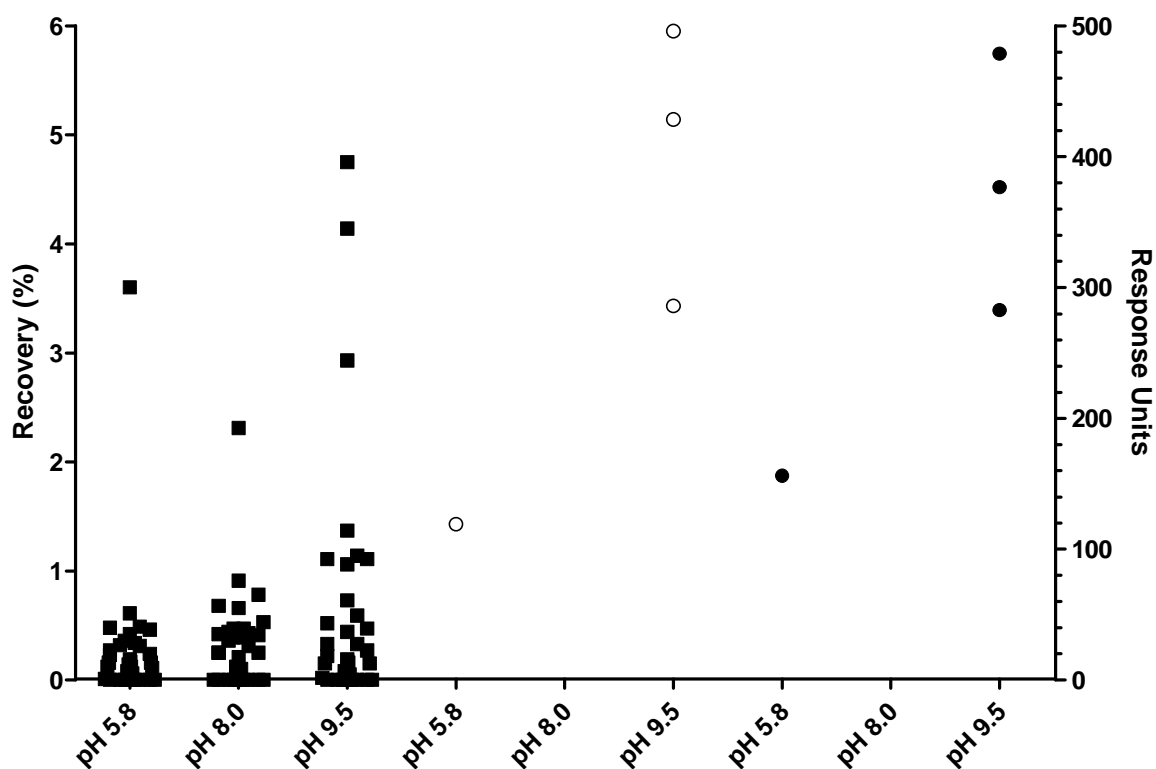
The refolding conditions for AKT2 were tested using the screen developed for p38 $\alpha$  (Chapter 2, Cowan *et al.*, 2008). Soluble protein was not available for AKT2, so it was not possible to create a standard curve for the SPR binding activity assay (folded protein recovery) as was possible for p38 $\alpha$ . Instead the data for this analytical readout are presented as response units rather than as a concentration of protein. Examples of the analytical readouts used to assay the refolding recovery of AKT2 are shown in Figure 3.6.



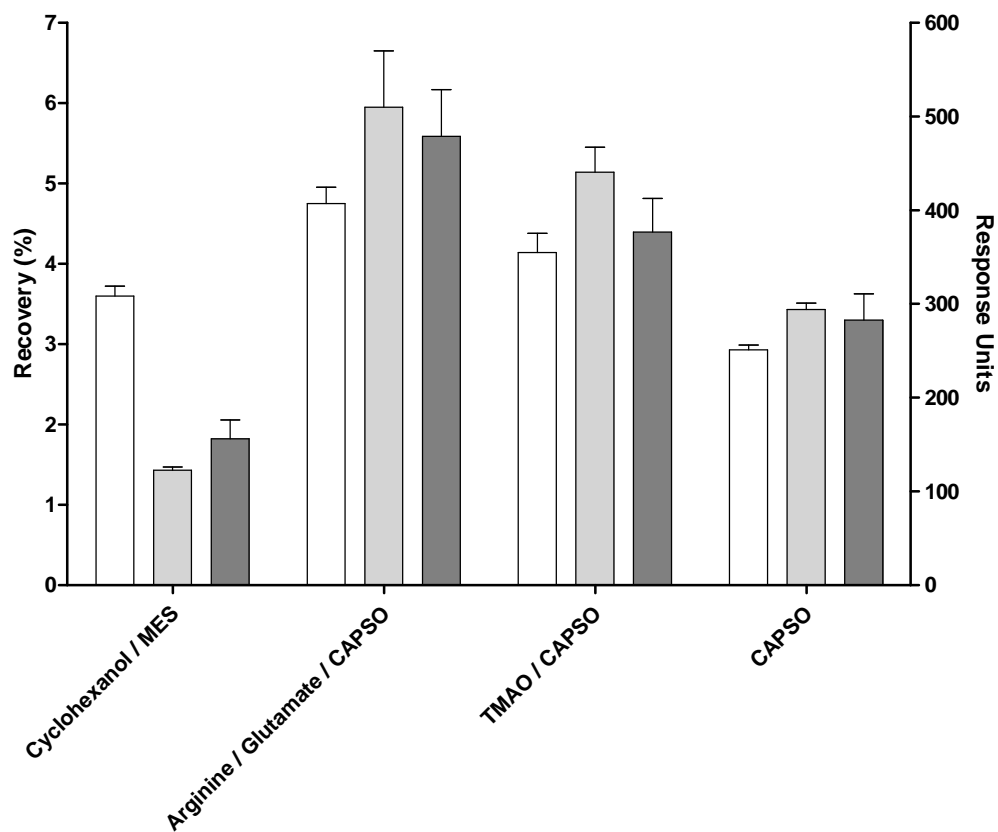


**Figure 3.6:** Example screen readout data for AKT2 refolding screens. (A) Capillary electrophoresis of AKT2. LM – lower marker, UM – upper marker. (B) Analytical size exclusion chromatography of refolded AKT2. Peak A – monomeric AKT2. Peak B – Myoglobin. (C) Surface plasmon resonance analysis of refolded AKT2. Curve 1 – refolded AKT2 sample. Curve 2 – refolded AKT2 sample + 10  $\mu$ M UTDC in solution.

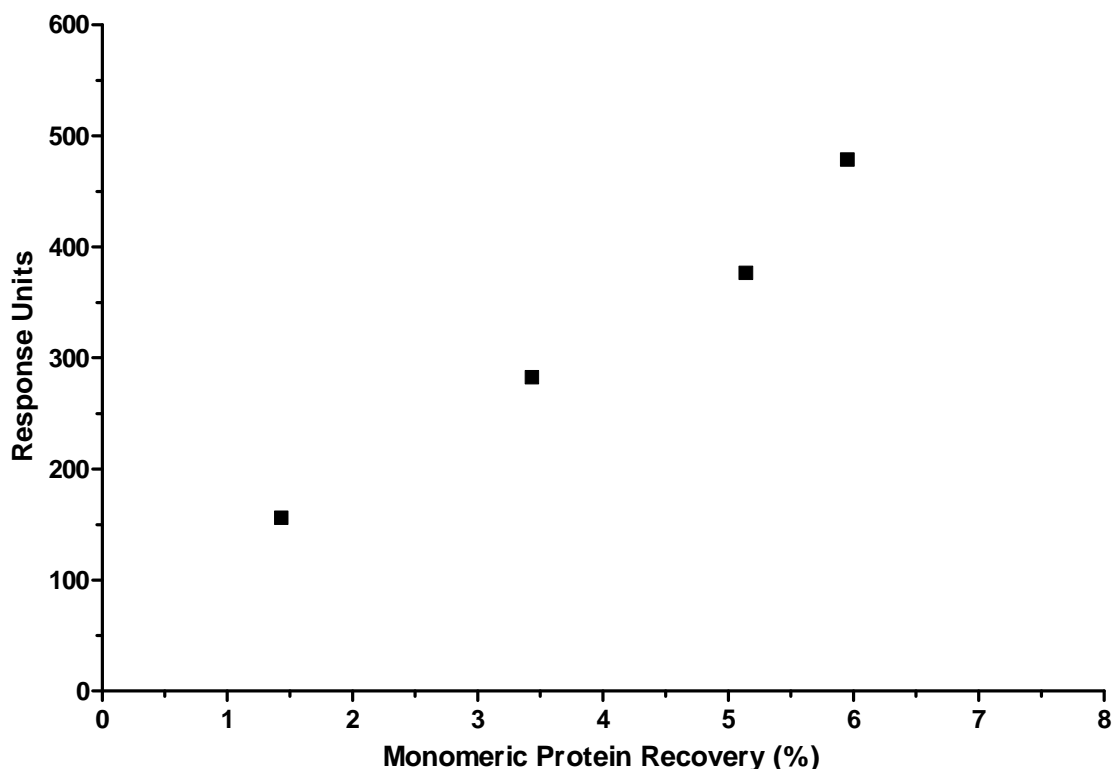
The refolding of AKT2 is of a much lower efficiency than p38 $\alpha$ . The average soluble protein recovery at pH 5.8 was 0.30 % as opposed to an average recovery of 0.36 % at pH 8.0 and 0.70 % at pH 9.5. The highest yield of refolding achieved with AKT2 was 6% (Figure 3.7) which was achieved with 50 mM arginine 50 mM glutamate in 100 mM pH 9.5 CAPSO (Figure 3.8). This compares to yields of 13% achieved with p38 $\alpha$  in pH 9.5 100 mM CAPSO. The number of conditions which gave detectable monomeric and folded protein was also minimal. Only four conditions gave measurable refolding in these analytical readouts. The conditions which gave measurable recoveries in the analytical size exclusion assay (monomeric protein recovery) and in the SPR based binding activity assay (folded protein recovery) were 1 mM Cyclohexanol 50 mM MES pH 5.8; 50 mM Arginine 50 mM Glutamate 100 mM CAPSO pH 9.5; 1 M TMAO 100 mM CAPSO pH 9.5 and 100 mM CAPSO pH 9.5 (Figure 3.8).



**Figure 3.7:** Recovery of refolded protein and response units obtained for the refolding of AKT2 in the kinase refolding screen. Results are grouped by pH of the renaturation buffer. ■ – Soluble Protein recovery, ○ – Monomeric protein recovery; ● – response units from SPR binding assay (folded protein recovery). Results are mean of 2 screens.



**Figure 3.8:** Recovery of refolded protein obtained for the refolding of AKT2 for conditions resulting in measurable recovery of monomeric protein. Soluble Protein recovery and Monomeric protein recovery plotted as percentage recovery, SPR binding activity assay (folded protein recovery) results plotted as response units. Soluble protein recovery – white, monomeric protein recovery – light grey, SPR binding activity assay – dark grey. Error bars show SE of mean.



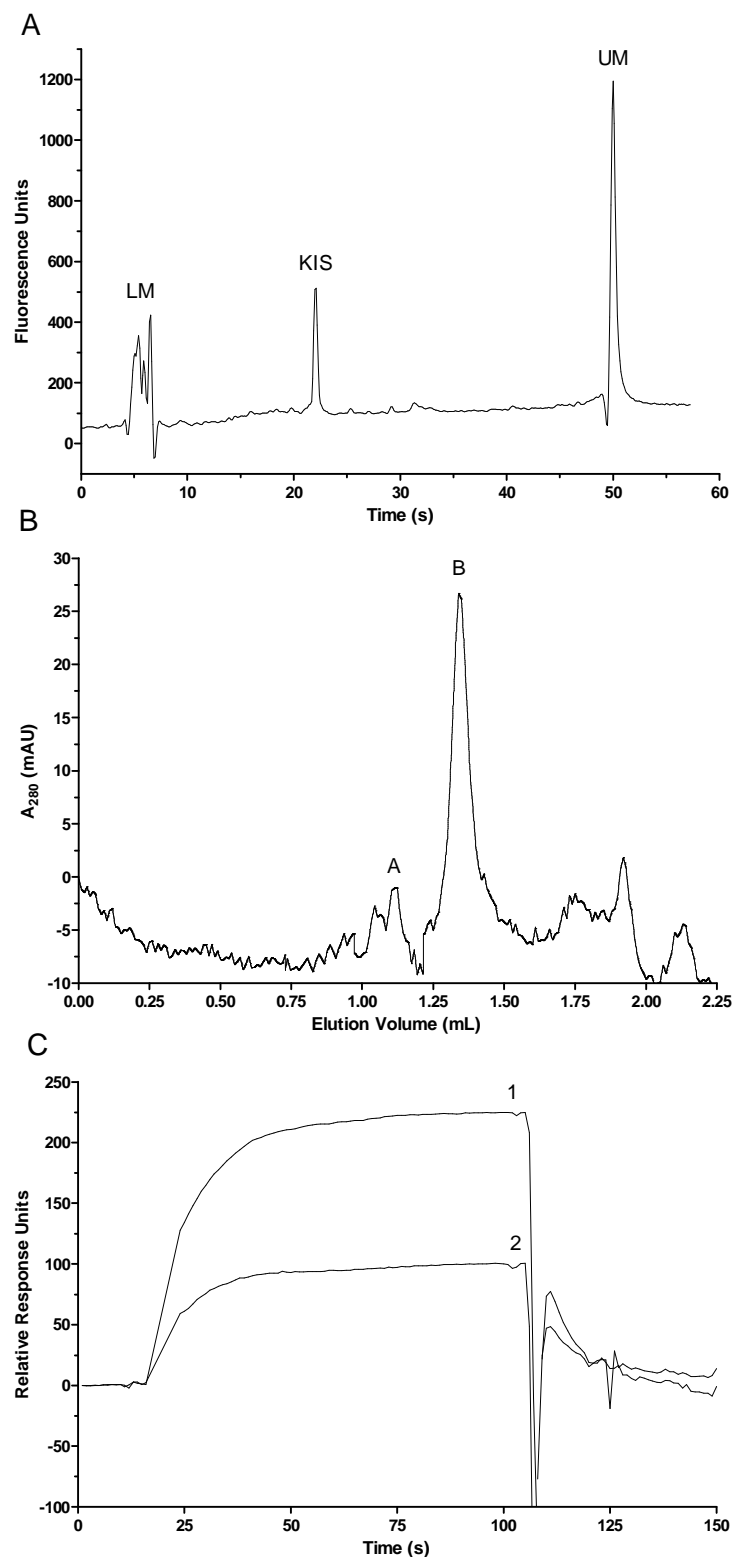
**Figure 3.9:** Correlation between monomeric protein recovery and response units from the SPR based folded protein assay for the refolding of AKT2. Results are mean of two experiments.

There does not appear to be a significant difference between the recoveries of soluble protein achieved at different pHs. The significance of any difference between the distributions of soluble protein recoveries was tested using a T-test. None of the differences between the average recoveries of soluble protein between the different pH groups were found to be statistically significant. This compares to the observation of a significant difference in the average recovery of soluble protein when refolding p38 $\alpha$  at different pHs. The comparison of the recoveries of soluble protein is particularly useful for AKT2, since a low number of conditions resulted in measureable recoveries of monomeric protein. Given the low yield of refolding AKT2 no particular additive or additive group stands out as being effective in assisting the refolding of AKT2. A high pH appears to be beneficial, since it appears to result in high recoveries of refolded protein, but it does not appear to be essential as a high recovery is recorded at pH 5.8. The repeatability of the refolding of AKT2 is good, with low variability between the experiments despite the low level of

recovery of refolded protein. The correlation between the level of monomeric protein recovery and the SPR based binding assay is good (Figure 3.9).

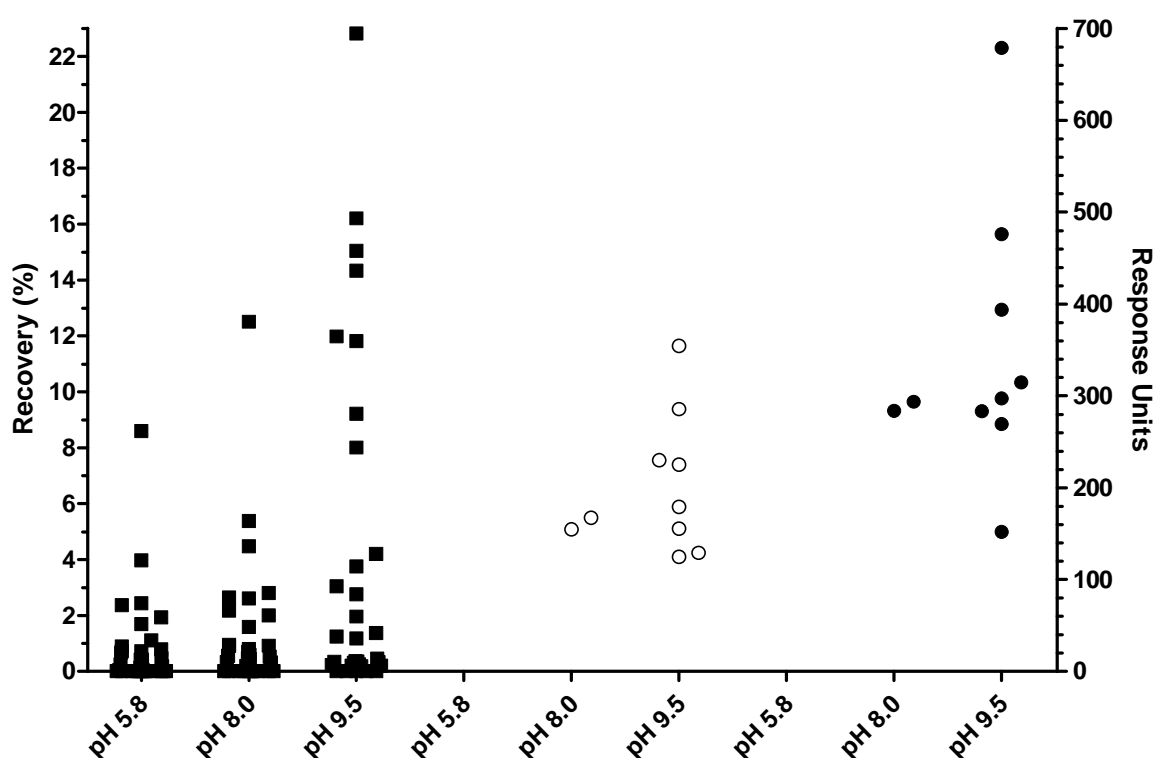
### ***3.3.5 The Refolding of KIS***

The refolding of the kinase domain of KIS, residues 1-313, was performed using the screen described previously (Chapter 2, Cowan *et al.*, 2008). Soluble protein was not available for KIS, so it was not possible to create a standard curve for the SPR binding activity assay (folded protein recovery) as was possible for p38 $\alpha$ . Instead, the data for this analytical readout are presented as response units rather than as a recovery of protein. Examples of the analytical readouts used to assay the refolding recovery of KIS are shown in Figure 3.10. Figure 3.10B shows a condition with a poor recovery of monomeric protein of less than 2 % recovery.



**Figure 3.10:** Example screen readout data for KIS refolding screens. (A) Capillary electrophoresis of KIS. LM – lower marker, UM – upper marker. (B) Analytical size exclusion chromatography of refolded KIS. Peak A – monomeric KIS. Peak B – Myoglobin. (C) Surface plasmon resonance analysis of refolded KIS. Curve 1 – refolded KIS sample. Curve 2 – refolded KIS sample + 10  $\mu$ M UTDC in solution.

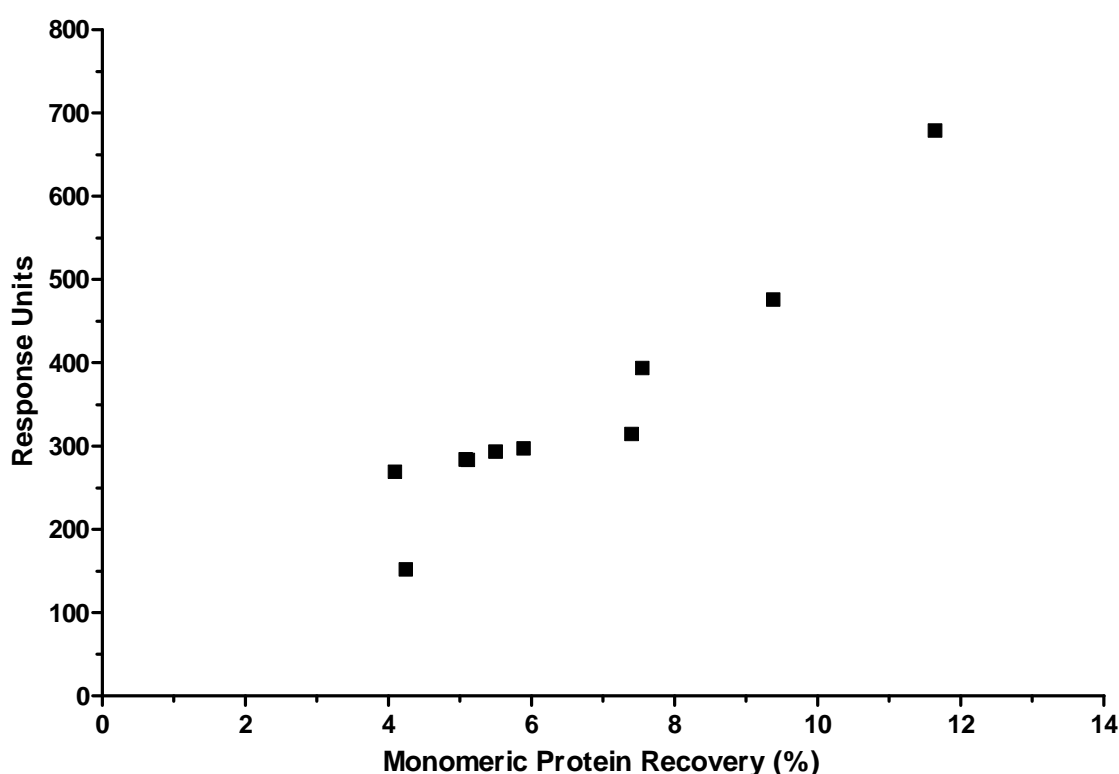
The refolding of KIS is of similar efficiency to that of p38 $\alpha$ . The recoveries of refolded protein, for the best conditions, are in the range of 10% to 15%. A much wider number and range of conditions lead to high recoveries than do so with the refolding of AKT2. Despite this, the number and variety of conditions leading to high levels of recovery of refolded protein is much lower than the number seen for p38 $\alpha$ . Ten conditions supported the refolding of KIS to a level which was detectable with all three screen readouts used (Figure 3.11). The highest recoveries of monomeric protein was found at pH 9.5, but pH 5.8 did not lead to the identified recovery of monomeric protein (Figure 3.11)



**Figure 3.11:** Recovery of refolded protein and response units obtained for the refolding of KIS in the kinase refolding screen. Results are grouped by pH of the renaturation buffer. ■ – Soluble Protein recovery, ○ – Monomeric protein recovery; ● – response units from SPR binding assay. Results are mean of 2 screens.

The average soluble protein recovery at pH 5.8 was 0.87 % as opposed to an average recovery of 1.37 % at pH 8.0 and 4.13 % at pH 9.5. The average recoveries of refolded protein, as measured by the recovery of soluble protein, are significantly different between pH 5.8 and pH 9.5, but not between pH 5.8 and pH 8.0, as determined by the students T-test, at a significance level of 5 %. This is due to the

higher number of high recovery results at pH 9.5 that are not present at pH 8.0. There is also a significant difference between the average recovery of soluble protein at pH 8.0 and pH 9.5. Furthermore, it can be seen that there are more conditions that have resulted in high recoveries of monomeric protein from pH 9.5 conditions than from pH 8.0 conditions. The amount of data points present does not allow for the testing of the significance of the differences in the average monomeric recoveries from these conditions.

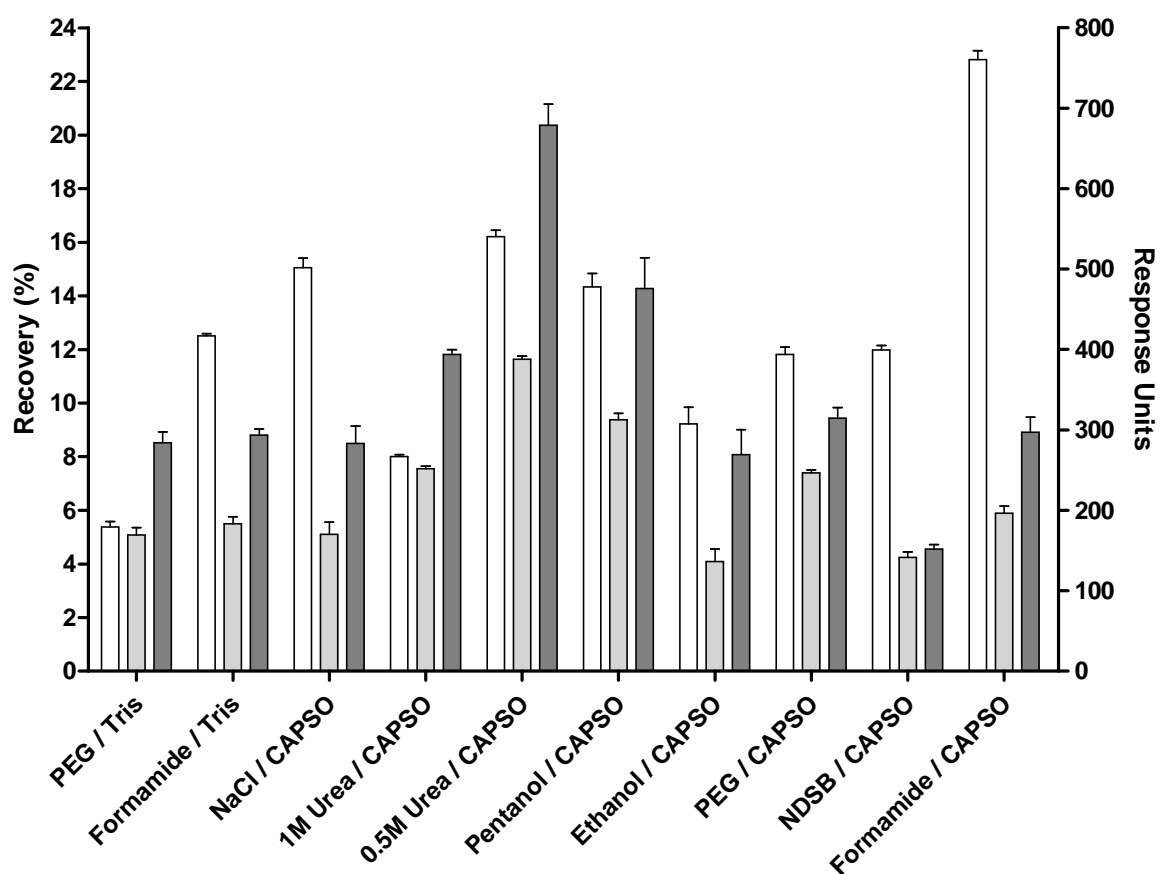


**Figure 3.12:** Correlation between monomeric protein recovery and response units from the SPR based folded protein assay for the refolding of KIS. Results are mean of two experiments.

The correlation between the monomeric protein recovery, and the response units from the SPR based binding activity assay is good (Figure 3.12). The Pearson correlation co-efficient for the two readout results is 0.951. This value indicates that the correlation between the two readouts is significant. The variation between the screens performed was low, as indicated by low standard errors of means shown in Figure 3.13.



Ten conditions gave identifiable recoveries from the analytical size exclusion assay of the refolded protein and the SPR based binding activity assay. Of these 10 conditions, two were at pH 8.0 in 50 mM Tris buffer. The additives that showed effectiveness in promoting the refolding of KIS at pH 8.0 were 0.04 % PEG3440 and 1 M formamide. The remaining eight conditions were various additives in pH 9.5 100 mM CAPSO. The additives effective in this buffer were; 0.5 M NaCl; 1 M and 0.5 M urea; 5 mM pentanol; 10 % ethanol; 0.04 % PEG3440; 1 M NDSB and 1 M formamide. Two additives have been identified as being effective in facilitating refolding at both pH 8.0 and pH 9.5. These additives were 0.04 % PEG3440 and 1 M formamide. The recoveries of soluble and monomeric protein and the response units from the SPR based binding activity assay are shown in Figure 3.13.



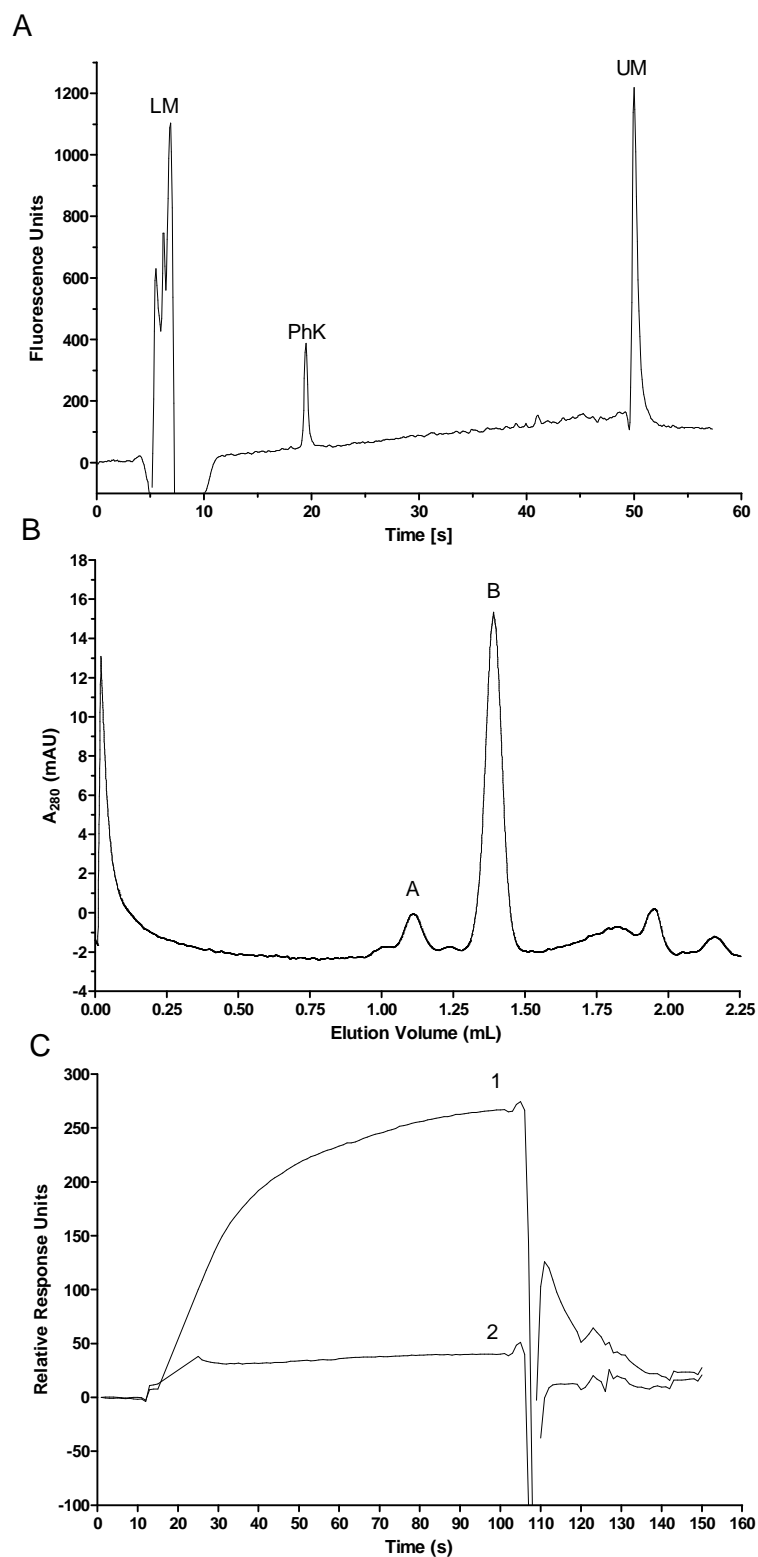
**Figure 3.13:** Recovery and response units obtained for the refolding of KIS for successful conditions. Soluble Protein recovery and Monomeric protein recovery plotted as percentage recovery, SPR binding activity assay results plotted as response units. Soluble protein recovery – white, monomeric protein recovery – light grey, SPR binding activity assay – dark grey. Error bars show SE of mean.

Several of the effective refolding conditions display a substantial drop in recovery between the soluble protein recovery and the monomeric protein recovery. It appears that for these additives approximately 50 % of the soluble protein is in the form of soluble oligomers. This effect is most dramatic with a renaturation buffer of 1 M formamide 100 mM CAPSO pH 9.5 where the monomeric protein yield is nearly 25 % of the soluble protein yield. Approximately 75 % of the soluble protein is in the form of soluble oligomers. If the correlation of the soluble protein recovery and the monomeric protein recovery is examined it can be seen that there is a poor correlation between the two, with a correlation co-efficient of 0.267 (Figure 3.13). This is indicative of the formation of differing proportions of monomers and soluble oligomers with differing renaturation conditions. Some conditions, however, do not show such a discrepancy between the soluble protein recovery and the monomeric protein recovery. The conditions, 0.04 % PEG3440 in 50 mM Tris pH 8.0 and 1 M urea in 100 mM CAPSO pH 9.5 display a close association between these two analytical readouts, suggesting the absence of soluble oligomers. In addition, 0.5 M urea in 100 mM CAPSO pH 9.5 displays a smaller difference between the soluble protein recovery and the monomeric protein recovery. The monomeric protein recovery is ~70% of the soluble protein recovery. Similarly 5 mM pentanol in 100 mM CAPSO pH 9.5 and 0.04 % PEG3440 in 100 mM CAPSO pH 9.5 give greater than 60% recovery of monomeric protein compared to soluble protein recovery.

The highest recovery of monomeric KIS was achieved with a renaturation buffer of 0.5 M urea in 100 mM CAPSO pH 9.5. A scaling up of the refolding reaction to produce milligram amounts of refolded protein for further study was attempted using this renaturation buffer. However, when this was attempted, and the refolded protein purified by size-exclusion chromatography no monomeric protein was recovered. The scaling up of the refolding reaction and the subsequent purification probably failed due to the presence of these soluble oligomers, which may have been increased in abundance due to the higher concentration of protein required to perform the size-exclusion chromatography purification. A larger scale refolding which utilised refolding conditions which gave a close match between the level of soluble recovery and monomeric protein recovery might have resulted in the purification of soluble KIS; however, time constraints did not allow this experiment to be attempted.

### **3.3.6 The Refolding of PhK**

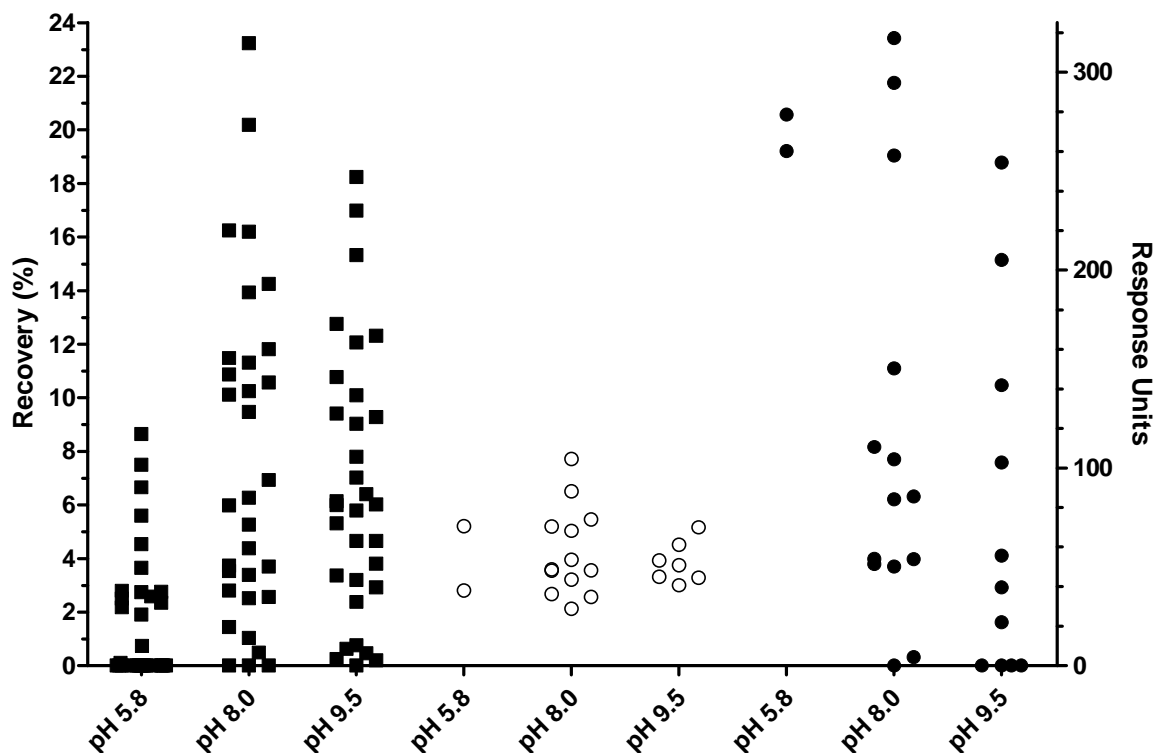
The refolding of the kinase domain of PhK was performed using the kinase refolding screen laid out previously (Chapter 2, Cowan *et al.*, 2008). The co-expression of the PhK kinase domain with additional GroEL subunits using the pGRO7 chaperone expression plasmid was able to produce PhK in a soluble form. However it was not possible to purify the solubly expressed PhK to a purity necessary for the use of the soluble protein to construct standard curves for the analytical readout methods. Therefore the procedures adopted for AKT2 and KIS were also used for these readout methods. Examples of the analytical readouts used to assay the refolding recovery of PhK are shown in Figure 3.14.



**Figure 3.14:** Example screen readout data for PhK refolding screens. (A) Capillary electrophoresis of PhK. LM – lower marker, UM – upper marker. (B) Analytical size exclusion chromatography of refolded PHK. Peak A – monomeric PhK. Peak B – Myoglobin. (C) Surface plasmon resonance analysis of refolded PhK. Curve 1 – refolded PhK sample. Curve 2 – refolded PhK sample + 10  $\mu$ M UTDC in solution.

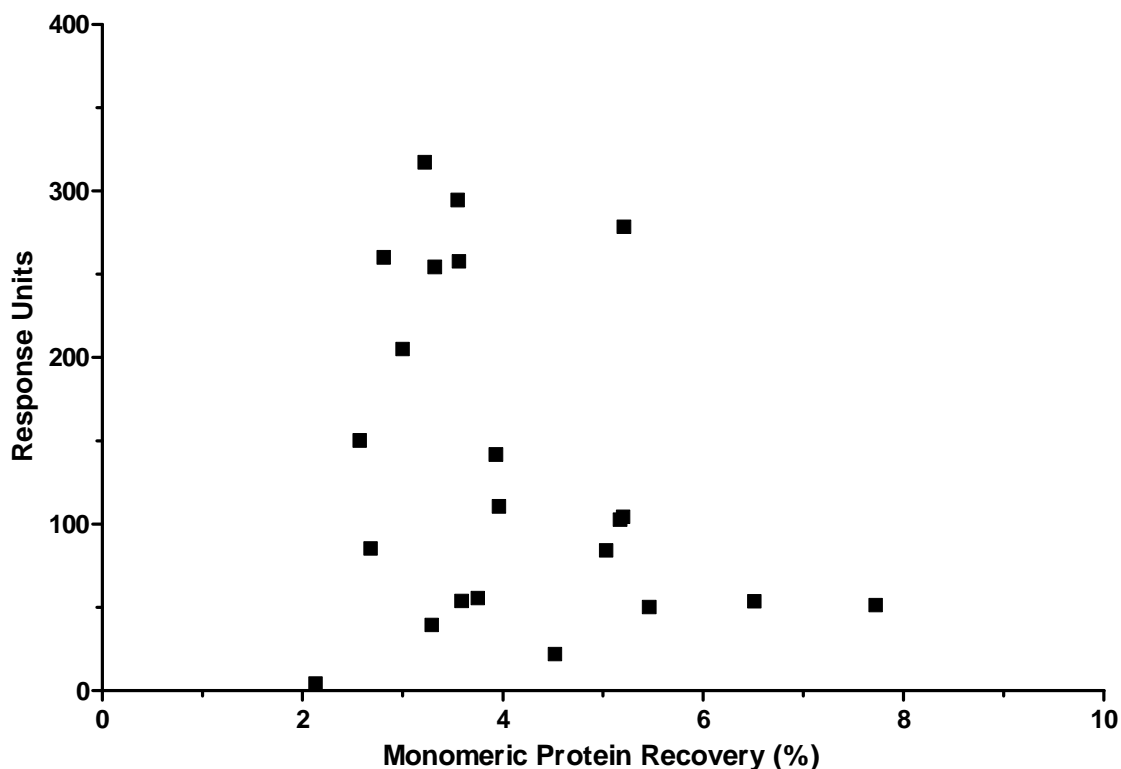
Since the PhK kinase domain had been previously refolded from inclusion bodies (Owen *et al.*, 1995), although no details had been given on the overall yield of the process, it was expected that the refolding screen would show high yields of refolding, similar to those seen for p38 $\alpha$ . The yields observed for the refolding of PhK are comparable to the yields obtained when refolding p38 $\alpha$  derived from inclusion bodies. The average recovery of soluble protein is higher under refolding conditions of high pH. The average soluble protein recovery at pH 5.8 was 1.79 % as opposed to an average recovery of 7.63 % at pH 8.0 and 6.69 % at pH 9.5. The average recovery at pH 5.8 is significantly different from the average soluble protein recovery at high pH, as determined by a single tailed T-test at a significance level of 5 %. The difference observed between the average soluble recovery at pH8.0 and pH9.5 is not significant (Figure 3.15). The pI of the PhK kinase domain is 6.06 (Table 3.1), approximately 1 unit greater than the low pH conditions, and 2 units lower than high pH conditions. The change in the ionization of the protein from pH 5.8 to pH 8.0 appears to have a significant effect on the recovery of refolding of PhK.

A higher number of conditions gave recoveries of monomeric protein and identifiable responses in the binding activity assay than was seen with both AKT2 and KIS. Twenty-two conditions gave these measurable responses (Figure 3.15). The majority (20) of these conditions occurred at the higher pHs used in the screen, with most of these results being obtained from a pH 8.0 50 mM Tris buffer (13 conditions) as opposed to pH 9.5 100 mM CAPSO (7 conditions). The average monomeric protein recovery at pH 8.0 and pH 9.5 is not significantly different (T-test at 5 % significance level).



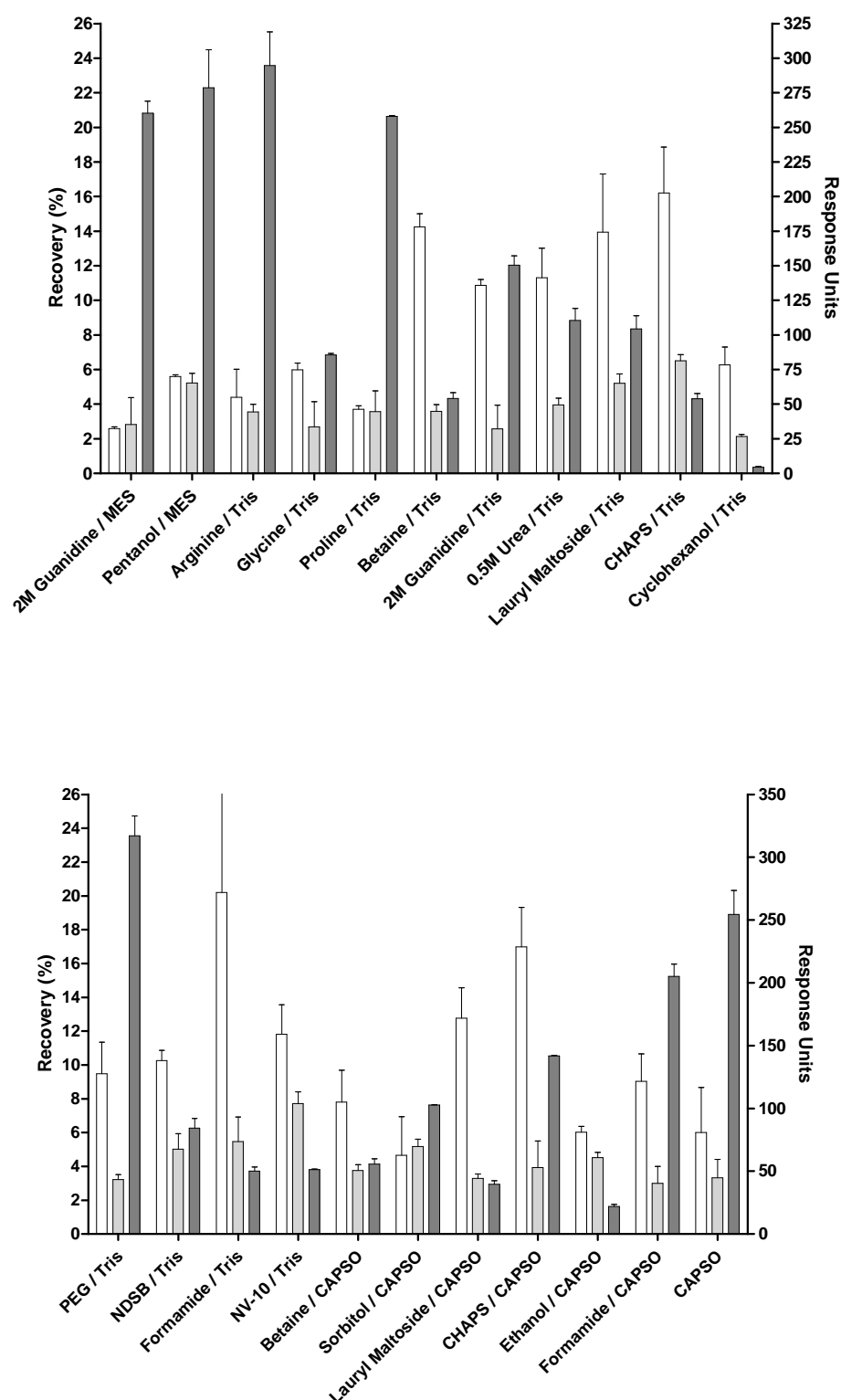
**Figure 3.15:** Recovery of refolded protein and response units obtained for the refolding of PhK in the kinase refolding screen. Results are grouped by pH of the renaturation buffer. ■ – Soluble Protein recovery, ○ – Monomeric protein recovery; ● – response units from SPR binding assay. Results are mean of 2 screens. Horizontal bar in the soluble protein recovery columns indicate the mean soluble protein recovery for all 32 conditions at each pH.

The quality of the folded protein recovery response units was tested by examining the correlation between the monomeric protein recovery and the response units from the SPR based binding assay. The correlation between the two analytical readouts for PhK was poor and not significant (Figure 3.16). The correlation between the soluble protein recovery and the monomeric protein recovery was also poor (Figure 3.17). This indicates the presence of soluble oligomers in various amounts under different refolding conditions. The chromatograms from the screening of the refolding of PhK confirm this, with peaks at the exclusion limit of the column used (data not shown). These problems indicate that the monomeric protein recovery is the best readout to be used to assess the refolding of PhK, due to the presence of soluble oligomers distorting the soluble protein recovery and the lack of a control protein for the surface plasmon resonance assay allowing testing the specificity and linear range of binding for PhK.



**Figure 3.16:** Correlation between monomeric protein recovery and response units from the SPR based folded protein assay for the refolding of PhK. Results are mean of two experiments.

Twenty conditions were noted to give measurable recoveries of refolded protein in the three analytical readout methods used. The recoveries obtained under these twenty conditions are shown in Figure 3.17. The recoveries of monomeric protein are low compared to the recoveries of soluble protein (Figure 3.15, 3.17). The highest recovery of monomeric protein achieved was 7.7 %. This recovery was achieved with a renaturation buffer of 0.1 % NV-10 in 50 mM Tris pH 8.0. The second best recovery was achieved under renaturation conditions of 20 mM CHAPS 50 mM Tris pH 8.0, where a soluble recovery of 6.5 % was achieved. The rest of the conditions tested gave soluble recoveries between 3 and 5 % (Figure 3.17). The majority of the additives that were effective in the supporting of the refolding of PhK were additives that suppress aggregation through various methods. Particularly prevalent were the additives from the detergents class, with lauryl maltoside and CHAPS supporting the refolding of PhK, both at pH 8.0 and pH 9.5. Formamide was also found to support the refolding of PhK at multiple pHs.



**Figure 3.17:** Recovery and response units obtained for the refolding of PhK for successful conditions. Soluble Protein recovery and Monomeric protein recovery plotted as percentage recovery, SPR binding activity assay results plotted as response units. Soluble protein recovery – white, monomeric protein recovery – light grey, SPR binding activity assay – dark grey. Error bars show SE of mean.

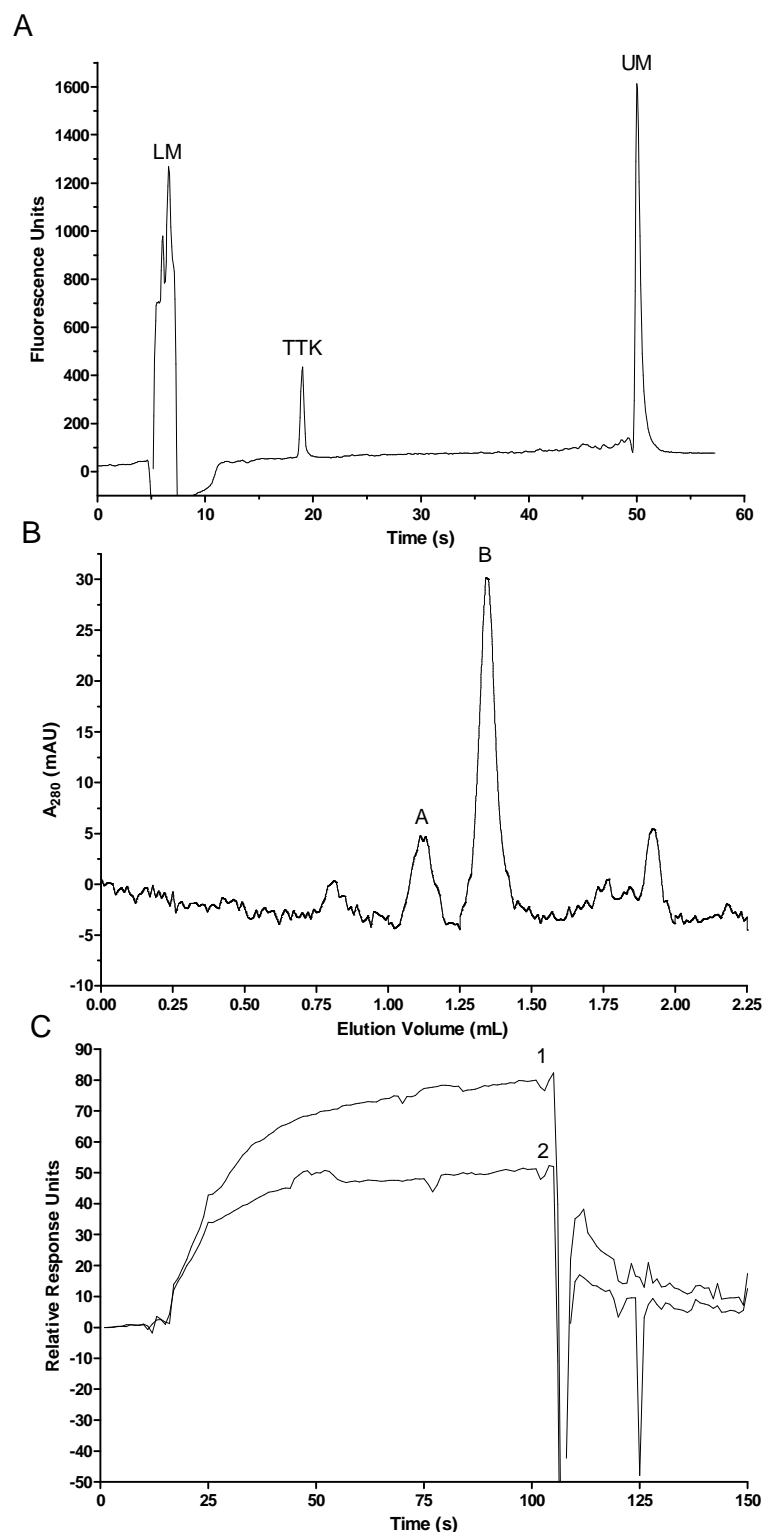


The correlation between the monomeric protein recovery and the soluble protein recovery was previously noted to be poor. This can be seen by an examination of the relative recoveries obtained through the two readout methods seen in Figure 3.17. Only in relatively few conditions, six in total, were the recoveries similar. In the rest of the cases the monomeric recovery was much lower than the soluble protein recovery. This difference suggests the presence of soluble oligomers of PhK. The renaturation buffers that best supported the refolding of PhK to a monomeric form were 2 M guanidine 100 mM MES pH 5.8; 5 mM pentanol 100 mM MES pH 5.8; 1 M proline 50 mM Tris pH 8.0; 20% sorbitol 50 mM Tris pH 8.0 and 10% ethanol 100 mM CAPSO pH 9.5.

The control condition at pH9.5, i.e. the condition lacking a specific refolding additive was able to give a monomeric protein recovery of 3.3 % (Figure 3.17), which is similar to the recovery of monomeric protein achieved under the same conditions with AKT2. The control conditions were not able to support a measurable recovery of protein with KIS.

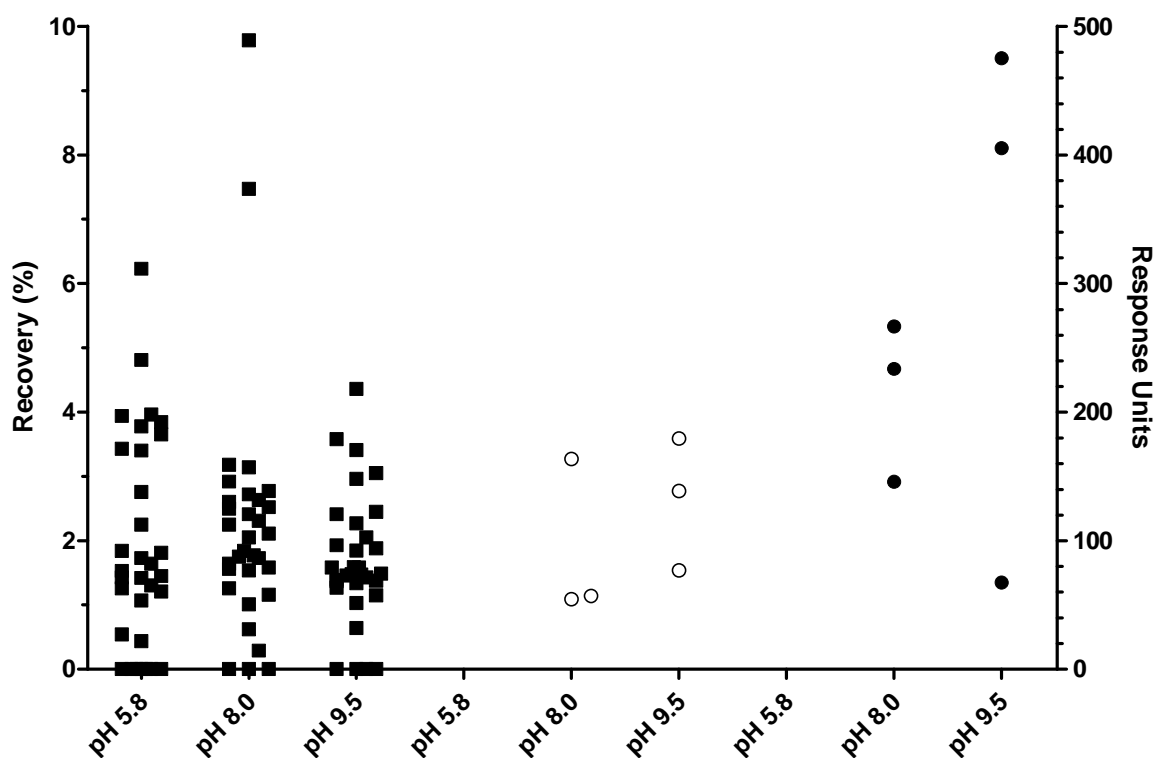
### **3.3.7 The Refolding of TTK**

The refolding of the TTK kinase domain, residues 514-804, was performed using the initial kinase refolding screen laid out previously (Chapter 2, Cowan *et al.*, 2008). Although it would prove possible to express and purify TTK in a soluble form, as laid out in chapter 4 of this thesis, it was decided to not use this soluble form of TTK as a control protein for the SPR based binding assay, and the analytical size exclusion chromatography standard curve, for reasons of data comparability among the four kinases tested using the refolding screen. Examples of the data obtained through the analytical readouts used to assay the refolding recovery of TTK are shown in Figure 3.18.

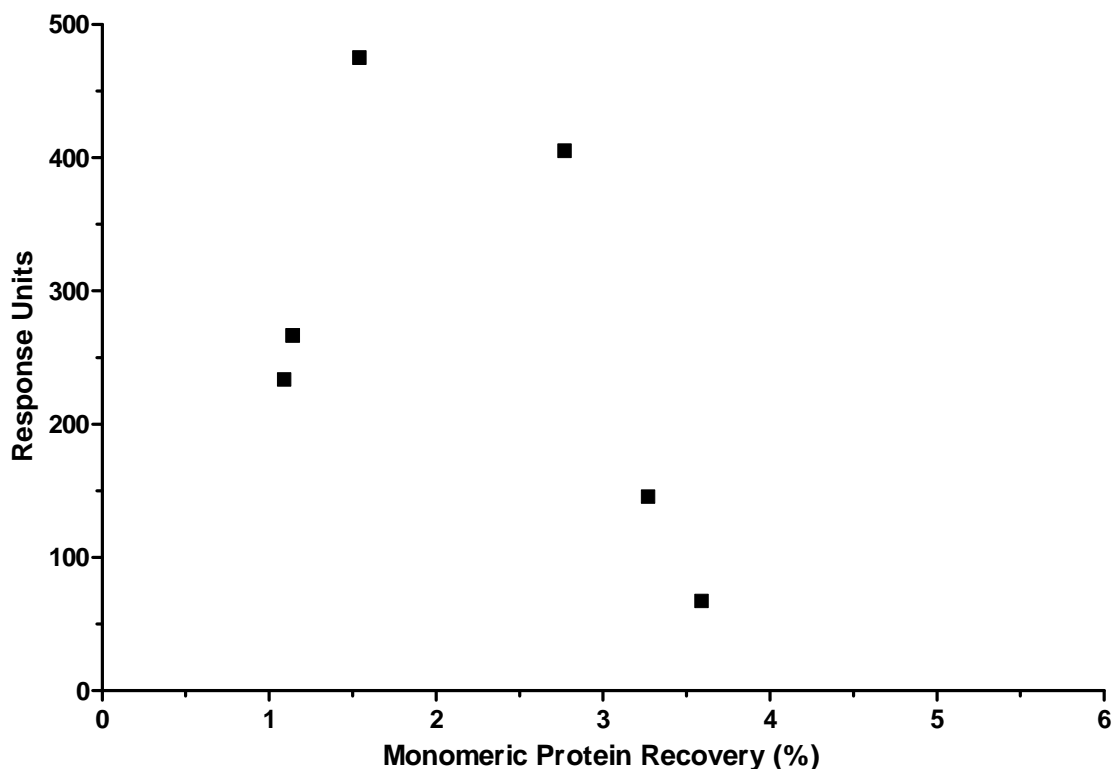


**Figure 3.18:** Example screen readout data for TTK refolding screens. (A) Capillary electrophoresis of TTK. LM – lower marker, UM – upper marker. (B) Analytical size exclusion chromatography of refolded TTK. Peak A – monomeric TTK. Peak B – Myoglobin. (C) Surface plasmon resonance analysis of refolded TTK. Curve 1 – refolded TTK sample. Curve 2 – refolded TTK sample + 10  $\mu$ M UTDC in solution.

The pH of the renaturation buffer has previously been noted to affect the recovery of soluble protein in the screen. For p38 $\alpha$ , KIS and PhK the average recovery of soluble protein was significantly lower at low pH when compared to the recovery of soluble protein at high pH. The average recovery of soluble protein upon the refolding of TTK (Figure 3.19) was not significantly different between pH5.8 and pH 8.0 or pH9.5. The difference between pH 8.0 and pH 9.5 was also not found to be significant. This result was in opposition to the results obtained for the majority of the kinase panel, where the recovery of soluble protein was significantly different between the different pHs used in the screen. The other kinases used in the screen, however, have pIs in the range of 5.48 to 6.65. The pI of TTK is higher, at 8.42 (Table 3.1). This change in the pI may be responsible for the different pattern of response to pH seen in the soluble protein recovery of TTK when compared to PhK and p38 $\alpha$ .

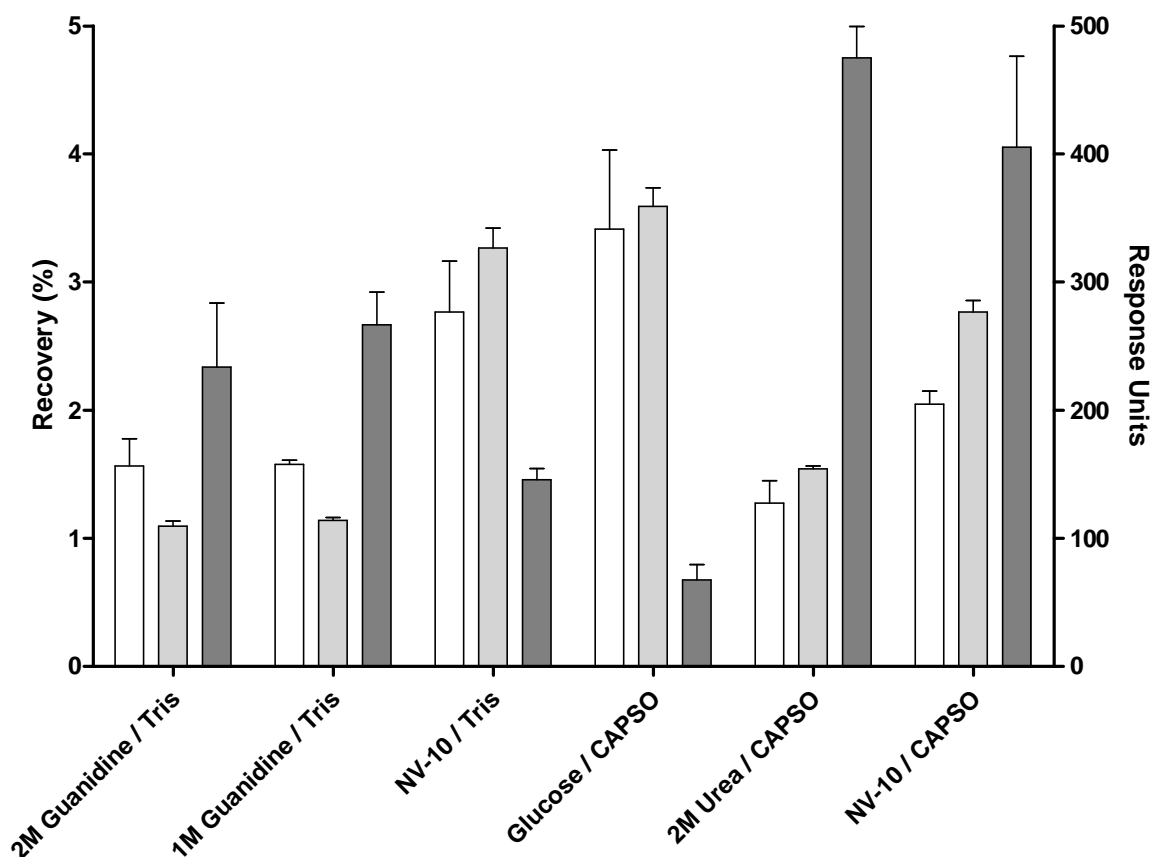


**Figure 3.19:** Recovery of refolded protein and response units obtained for the refolding of TTK in the kinase refolding screen. Results are grouped by pH of the renaturation buffer. ■ – Soluble Protein recovery, ○ – Monomeric protein recovery; ● – response units from SPR binding assay. Results are mean of 2 screens.



**Figure 3.20:** Correlation between monomeric protein recovery and response units from the SPR based folded protein assay for the refolding of TTK. Results are mean of two experiments.

The observed correlation between the monomeric protein recovery and the SPR response units is low, not significant, and negative (Figure 3.20) as determined by calculating the Pearson correlation coefficient and comparing the calculated value to a table of critical values at a 5 % significance level. There is however, a good correlation between the soluble protein recovery and the monomeric protein recovery (Figure 3.21). The number of refolding conditions which gave rise to a monomeric protein recovery and a measurable response in the SPR based binding activity assay was low. Only six conditions gave a measurable response in both assays. A large number of conditions gave low measurable responses in the SPR based activity assay of less than 20 response units. These conditions did not give a measurable response in the analytical size exclusion assay, probable due to the low protein concentration in these conditions, as indicated by the low measured response units. These results were excluded from the relevant columns on Figure 3.19. The recoveries of soluble protein, monomeric protein and the response units from the SPR based activity assay for the six successful conditions are summarised in Figure 3.21.

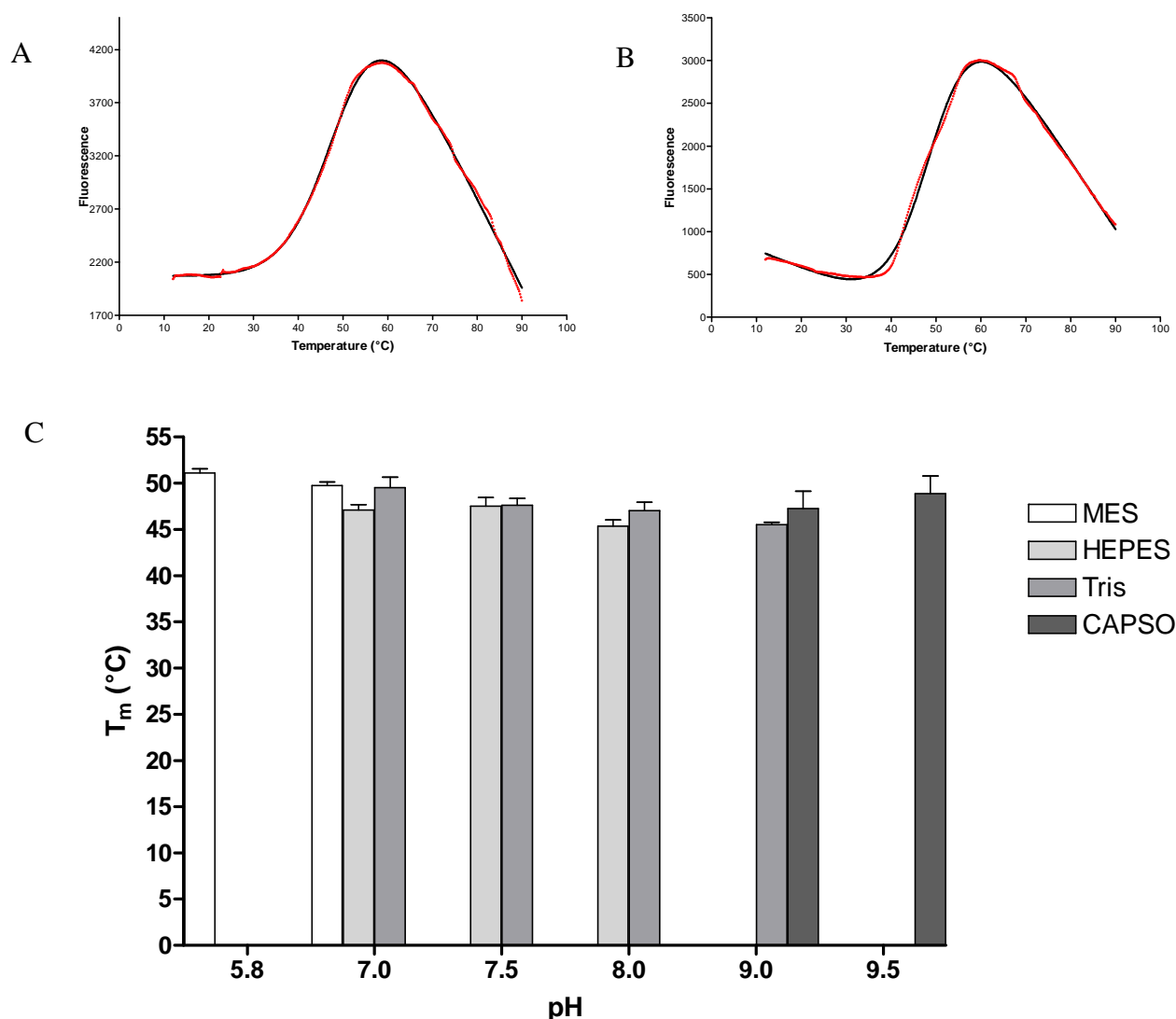


**Figure 3.21:** Recovery and response units obtained for the refolding of TTK for successful conditions. Soluble Protein recovery and Monomeric protein recovery plotted as percentage recovery, SPR binding activity assay results plotted as response units. Soluble protein recovery – white, monomeric protein recovery – light grey, SPR binding activity assay – dark grey. Error bars show SE of mean.

There does not appear to be the formation of significant amounts of soluble oligomers that were observed with other members of the kinase panel. The overall recoveries of refolded protein were; however, low, with a maximum recovery of ~3.5 %. This is below the cut-off that was applied for further analysis in the screening of p38 $\alpha$ , and is the lowest recovery observed with the kinase panel. The control conditions, lacking a refolding additive, were not able to support the refolding of TTK. However a commonality was observed among the additives which were capable of supporting the refolding of TTK. NV-10 was found to be capable of supporting the refolding of TTK at both pH 8.0 and pH 9.5. In addition, different conditions from the denaturant groups also supported the refolding of TTK from inclusion bodies. Interestingly guanidine was effective at pH 8.0 and urea at pH 9.5.

The concentrations required to support refolding were fairly high (1 – 2 M), indicating that the refolding was supported through the maintenance of a substantially unfolded state. The most effective additive was 1 M glucose at pH 9.5, although NV-10 at pH 8.0 was similarly effective. The nature of these two additives is fairly similar, NV-10 being a branched polysaccharide, and it appears that, in this case, glucose is able to perform a similar function to NV-10, though a substantially larger amount of glucose is required to have the same effect

Due to the availability of pure, soluble protein (Chapter 4), which was not available for the other members of the kinase panel it was possible to examine the stability of the TTK kinase domain at various pHs and in different buffers, in the same manner as was performed with p38 $\alpha$  to explain the refolding behaviour of p38 $\alpha$  relative to the pH. A thermal melting analysis of TTK was performed using the buffers 100 mM MES, 10 mM HEPES, 50 mM Tris and 100 mM CAPSO at a range of pHs that covered the range of the screen; from pH 5.8 to pH 9.5. The thermal unfolding of TTK was performed according to the method laid out in section 2.2.18 and the thermal unfolding transitions were fitted to equation 2.2 to determine the parameters of the unfolding transition. The results of this screening of the thermal unfolding of TTK are shown in Figure 3.22 and the mid points of the thermal melting transitions are summarised in table 3.2.



**Figure 3.22:** (A) Thermal unfolding transition of TTK in 100mM MES pH 5.8. Red circles show experimental data, solid line shows fit of Equation 2.1 to experimental data. (B) Thermal unfolding transition of TTK in 10 mM HEPES pH 7.0. Red circles show experimental data, solid line shows fit of Equation 2.1 to experimental data. (C) Average thermal melting points for TTK under conditions of various pH and buffers. Results are the mean of 5 unfolding transitions. Error bars indicate standard error.

**Table 3.2**  $T_m$  values for TTK under various pH and buffer conditions.

		$T_m$ (°C)			
pH	Buffer	100 mM MES	10 mM HEPES	50 mM TrisHCl	100 mM CAPSO
5.8		51.1 ± 0.45			
7.0		49.7 ± 0.38	47.1 ± 0.57	49.5 ± 1.13	
7.5			47.5 ± 0.97	47.6 ± 0.72	
8.0			45.3 ± 0.67	47.0 ± 0.89	
9.0				45.5 ± 0.24	47.2 ± 1.90
9.5					48.9 ± 1.90

The results of the thermal melting analysis of TTK are consistent with the lack of a significant difference between the average recovery of soluble protein across the pH range utilised in the refolding screen. It can be seen that, unlike p38 $\alpha$  (Figure 2.11), there is not a substantial change in the mid point of the thermal melting transition observed for TTK either with a variation in pH or with the buffer used. The conditions, however, which result in refolded protein (Figure 3.21) occur in the high pH conditions, which is surprising, since it might be expected from the thermal unfolding data (Figure 3.22C, Table 3.2) and the soluble protein recovery distribution (Figure 3.19) that low pH conditions would result in measurable refolded protein in the analytical size exclusion experiment and the SPR based binding activity assay. This is because there is no large change in the  $T_m$  of TTK under conditions of low pH, and the average recovery of soluble protein is not significantly different between low and high pH. The reason for this unexpected result is difficult to understand from the data collected, but it can be suggested that the unusual effect may be due to the interaction of the buffer and the refolding additives.

### 3.3.8 Fractional Factorial Screen Design

The initial screen that was used to screen the refolding of the members of the kinase panel is unable to test for the effect of a combination of refolding additives. In



addition, the analysis of the results of the refolding screen is limited in the statistical tests that can be applied. To improve the screen in this manner a factorial screen was applied to the problem of screening for the refolding of protein kinases. This allows for the examination of the effect of the combination of different refolding additives and the use of more powerful ANOVA statistical tests to analyse the results of the refolding screen.

Factorial screens can exploit a fractional design to reduce the number of experiments required to cover the parameter space of a screen. The ANOVA analysis is interpolating the missing data points and covering the whole parameter design space in the analysis. The number of factors analysed in the screen and the extent of the fractional nature of the screen used in the screen were chosen to balance the number of factors analysed, the coverage of the design space by the screen and the number of experiments required in each screen.

To balance these three requirements for the screen, a screen using six factors was chosen. A full factorial screen with this number of factors requires  $2^6$ , or 64 experiments per screen. This would not represent a substantial decrease of the number of experiments required for the screen over the initial screen used previously. The screen was converted to a half-fractional design from a full factorial design. This reduces the number of experiments required for the screen to 32, which is a substantial reduction of the size of the screen. The drawback of the use of a half-factorial screen is the aliasing of the results. Aliasing in a fractional factorial screen results in the ANOVA analysis of the screen being unable to distinguish between certain combinations of factors. The model of the screen used does not consider combinations of factors of more than three factors. It is unlikely that a combination of more than three factors would have a positive effect on the refolding of a protein kinase. In addition, some of the factor combinations with more than three factors can be directly examined by an examination of the raw results. The aliasing of the screen results in the ANOVA analysis of the screen being unable to distinguish pairs of three factor interactions. The aliased interactions are summarised in table 3.3

**Table 3.3:** Aliased model terms for the fraction factorial refolding screen. The pairs of terms given cannot be distinguished from the term used in the model constructed from the data and tested using ANOVA.

Model Term	Aliased Terms	
ABC	ABC	DEF
ABD	ABD	CEF
ABE	ABE	CDF
ABF	ABF	CDE
ACD	ACD	BEF
ACE	ACE	BDF
ACF	ACF	BDE
ADE	ADE	BCF
ADF	ADF	BCE
AEF	AEF	BCD

An analysis of the results of the initial screen was used to decide what screen conditions would be used in the fractional factorial screen. It was decided to add a new condition to the screen in the fractional factorial screen. The condition to be added was the inclusion of staurosporine, a generic kinase inhibitor, at a concentration of 5  $\mu$ M. The addition of a kinase inhibitor was expected to assist the refolding of kinases by stabilising the native state of the protein. Staurosporine was dissolved in DMSO at a concentration of 10 mM, leading to a final DMSO concentration in the screen of 0.05 % (v/v).

The other five conditions for the fractional factorial screen were selected from the conditions used in the broader refolding screen. The monomeric protein recovery was used to judge conditions that had been effective in increasing the recovery of monomeric protein. The recovery of monomeric protein in control conditions and in the presence of additives was compared, and additives which increased the recovery of monomeric protein over control conditions at the same pH were identified as candidates for inclusion in the factorial screen. The effect of pH on the refolding of protein kinases was different across the tested kinases. Some

kinases gave poor recoveries at pH 5.8, whereas some kinases did not show significant differences between the recoveries of soluble protein. AKT2 and KIS gave higher recoveries of monomeric protein at pH 9.5 when compared to pH 8.0. PhK, TTK and p38 $\alpha$  did not show a significant difference in the recoveries of soluble protein between pH 8.0 and pH 9.5. However, with TTK, the additive NV-10 gave a higher recovery of monomeric protein in 50 mM Tris pH 8.0 than in 100 mM CAPSO pH 9.5. On this basis, the buffer used was included as a factor, having two values of pH 8.0 and pH 9.5, with pH 8.0 being the base value. The remaining four factors in the fractional factorial screen were filled by examining the monomeric protein recovery and selecting factors that had been effective at promoting the refolding of several protein kinases above the control conditions. The conditions chosen on this basis were 1 M TMAO, 0.5 M urea, 0.1% NV-10 and 0.04% PEG3440.

An additional advantage of the use of a fractional factorial screen is that the software used to generate and to analyse the screen allows the randomisation of the screen layout to be easily performed, and does not require the re-ordering of the screen to allow analysis of the data. When this feature is combined with multiple screens for refolding a single protein this aids in minimising any edge effects in the screen. This feature was used in the fractional factorial screen for protein kinases.

The factors selected for the fractional factorial screen for the refolding of protein kinases were combined into a screen of 32 refolding conditions containing the six factors in the screen. A list of the combinations of factors that were used in the screen is laid out in Table 3.4. The order of this list does not represent the order in which the conditions were laid out in the fraction factorial refolding screen, as this order was randomised to minimise edge effects.

**Table 3.4:** Composition of the fractional factorial screen for protein kinase refolding. The composition of the 32 experiments in the half-fractional factorial screen are indicated by the presence or absence of 5 factors, and the pH of the buffer used in each experiment. The order of experiments in the screen was randomised when performing the screen to minimise edge effects.

Experiment		Screen Factors					Experiment		Screen Factors				
#	pH	TMAO	Urea	NV-10	PEG3440	Staurosporine	#	pH	TMAO	Urea	NV-10	PEG3440	Staurosporine
1	8.0	N	N	N	N	N	17	8.0	N	N	N	Y	Y
2	9.5	N	N	N	N	Y	18	9.5	N	N	N	Y	N
3	8.0	Y	N	N	N	Y	19	8.0	Y	N	N	Y	N
4	9.5	Y	N	N	N	N	20	9.5	Y	N	N	Y	Y
5	8.0	N	Y	N	N	Y	21	8.0	N	Y	N	Y	N
6	9.5	N	Y	N	N	N	22	9.5	N	Y	N	Y	Y
7	8.0	Y	Y	N	N	N	23	8.0	Y	Y	N	Y	Y
8	9.5	Y	Y	N	N	Y	24	9.5	Y	Y	N	Y	N
9	8.0	N	N	Y	N	Y	25	8.0	N	N	Y	Y	N
10	9.5	N	N	Y	N	N	26	9.5	N	N	Y	Y	Y
11	8.0	Y	N	Y	N	N	27	8.0	Y	N	Y	Y	Y
12	9.5	Y	N	Y	N	Y	28	9.5	Y	N	Y	Y	N
13	8.0	N	Y	Y	N	N	29	8.0	N	Y	Y	Y	Y
14	9.5	N	Y	Y	N	Y	30	9.5	N	Y	Y	Y	N
15	8.0	Y	Y	Y	N	Y	31	8.0	Y	Y	Y	Y	N
16	9.5	Y	Y	Y	N	N	32	9.5	Y	Y	Y	Y	Y

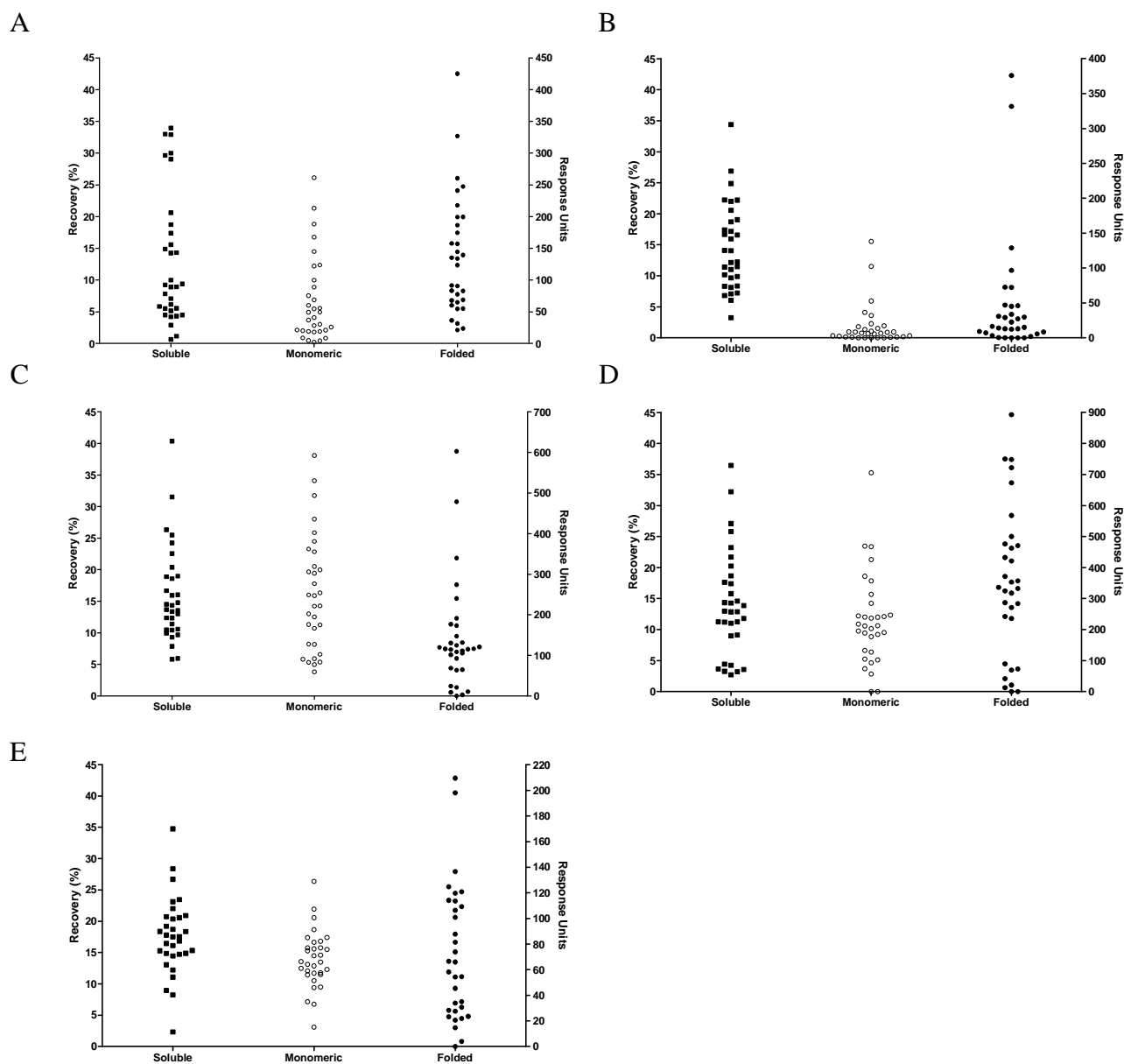
The screening procedure from the initial screen used previously (Figure 2.8) was used for the fractional factorial refolding screen. All the conditions in each screen were analysed by each analytical readout method used, since the ANOVA of the screen results required results for each experiment in the screen. A lack of data would compromise the analysis performed and result in less reliable results from the analysis. If the analytical method did not return a detectable value for the recovery of refolded protein, the recovery in the screen was assigned a value of 0 for the purposes of the analysis. To maintain similar data for all five kinases to be used in the fractional factorial screen the same methods of calibrating the analytical readouts were used for all five kinases. BSA was used to create a standard curve for the analytical size exclusion chromatography assay, and the recoveries of each kinase calculated from the relative molecular masses and calculated extinction coefficients. In addition, no standard curve was prepared for the SPR based binding activity assay, since three members of the kinase panel lacked available control protein and the results of the refolding of the five members of the kinase panel would not be comparable if the analytical readout of the screen were presented in different units.

### **3.3.9 Fractional Factorial Screen Results**

The fraction factorial screen created was applied to the five kinases which had been refolded using the initial refolding screen outlined previously (Chapter 2, Cowan *et al.*, 2008). The recoveries of monomeric protein, and the response units due to the binding of the refolded kinases to the UTDC were analysed using DesignExpert 7 (Stat-Ease) and the factors and combinations of factors that had a significant, positive or negative effect on the refolding of the five protein kinases used for the initial kinase refolding screen identified.

The recoveries of refolded protein observed in the fraction factorial refolding screen was generally higher than that seen in the initial screen design (Chapter 2). The recoveries of refolded protein, as measured by the three analytical readout methods described in chapter 2 are shown in Figure 3.23. The maximum recovery of refolded protein for AKT2 is raised from a recovery of 6 % with the initial screen design (Section 3.3.4) to a recovery of monomeric protein of 26 % (Figure 3.23A). Similarly the recovery of monomeric protein on the refolding of KIS was raised from ~12% in conditions of 0.5 M urea 100 mM CAPSO pH 9.5 to a refolding recovery of 15.5 % at maximum (Figure 3.23B). The recoveries of monomeric protein do appear, however, to cluster at lower values. This indicates that KIS

may be a more difficult protein to refold than other proteins tested in the screen, which produce higher maximum recoveries and a wider spread of recoveries from the fractional factorial screen.



**Figure 3.23:** Refolding recoveries of the five kinases tested using the fractional factorial refolding screen. (A) Refolding of AKT2, (B) refolding of KIS, (C) refolding of p38 $\alpha$ , (D) refolding of PhK, (E) refolding of TTK. Refolding recovery determined by capillary electrophoresis (■, soluble protein recovery), analytical size exclusion chromatography (○, monomeric protein recovery), and SPR binding activity (●, folded protein recovery). Results are mean of two screens. Maximum response units for each panel vary due to differing binding of proteins to the UTDC (Figure 2.6).

The refolding of p38 $\alpha$  has shown an increased recovery of monomeric protein in the fractional factorial screen as opposed to the screen described in chapter 2. The highest

monomeric protein recovery observed with the initial screen design was ~15 % (Figure 2.9A). The maximum monomeric protein recovery rises to 38 % in the fractional factorial screen (Figure 3.23C). The recovery of refolded protein is similar to the maximum recoveries obtained when refolding p38 $\alpha$  which had been sourced from protein expressed in a soluble form and subsequently unfolded with 8 M urea (Figure 2.9B). The range of recoveries of monomeric protein is also high, indicating that the screen should be capable of discriminating well between the effectiveness of different additives and combinations of additives.

The recovery of PhK, as determined through the recovery of monomeric protein, was dramatically improved by the combination of refolding additives, with the maximum monomeric protein recovery rising from 8 % to 35 % (Figure 3.17, 3.23D). A similar dramatic improvement in the maximum recovery of refolded protein was observed for TTK, with the maximum monomeric protein recovery rising from 3.59 % (Figure 3.21) to 26.4 % (Figure 3.23E). These results indicate that the fraction factorial screen has successfully identified conditions which lead to even higher recoveries of refolded protein when compared to the refolding screen described in chapter 2. This basic analysis of the results of the refolding screens does not, however, exploit the statistical tests which can be applied to the screen, and does not indicate if all of the factors present in the conditions which result in high recovery conditions are significant in increasing the refolding recovery. ANOVA of the screen results is used to elucidate this information.

The ANOVA of the screen readouts generates a model which attempts to explain the changes in the recovery of refolded protein from the base recovery, which for the refolding screen is in a renaturation buffer of 50 mM Tris pH 8.0. A limitation of the model that is generated to explain the differences in refolding yields between the conditions tested in the screen is that it must be hierarchical. That is, the model must include all terms that could make up a higher term. For example if a two factor term is included in the model, i.e. the effect of a combination of two factors is included in the model, then the model must incorporate the effect of the two single factors that were combined in the two factor combination. Likewise if, for example, the model incorporated the combination of factors ABC, then the model would have to include the following factors and combinations of factors; A, B, C, AB, AC and BC. The significance of all these factors can then be determined. This can quickly lead to a very complicated model if multiple combinations of three factors are included in the model.

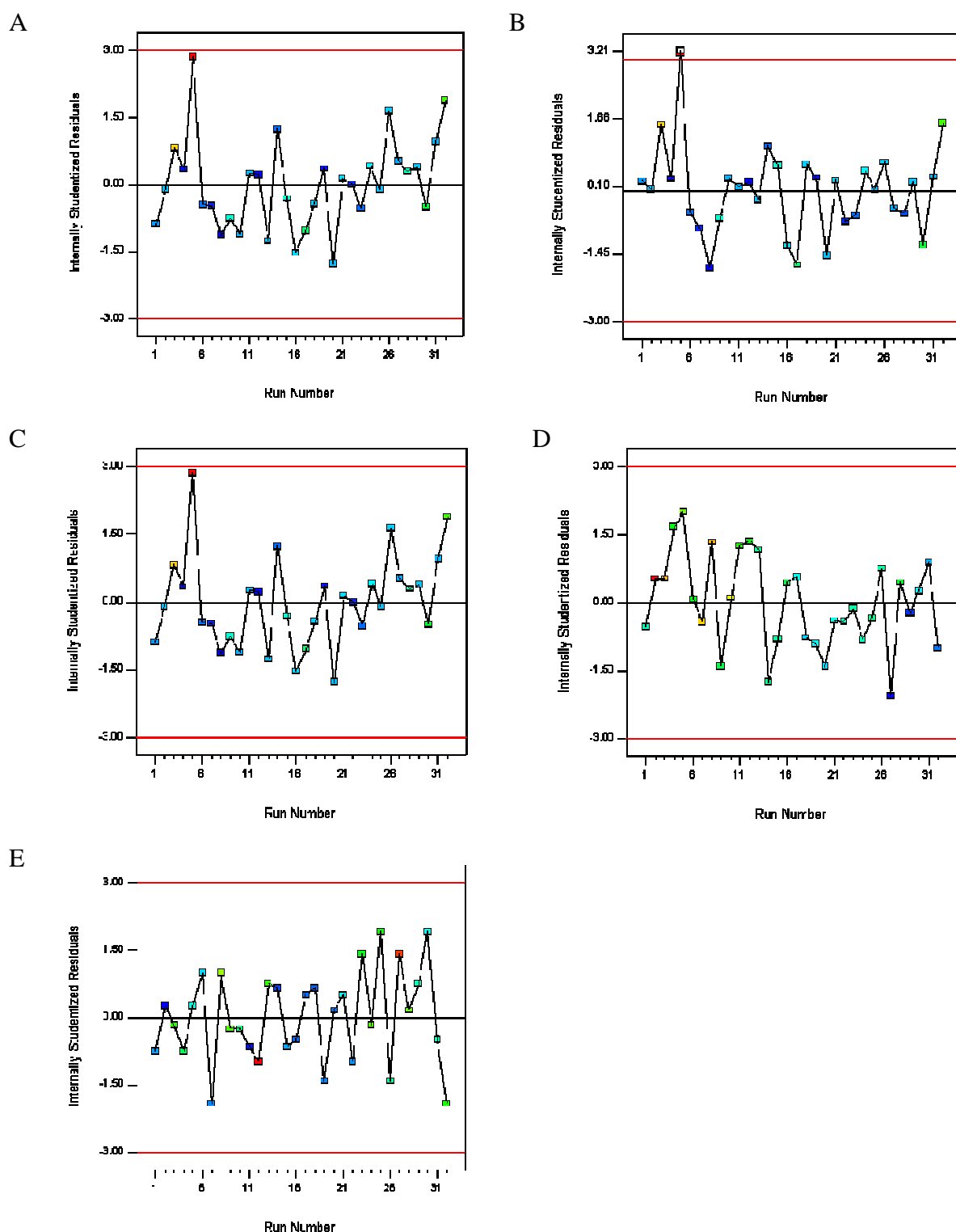
Several tests can be applied to examine the significance and appropriateness of the model that is generated from the ANOVA of the results. Overall F-tests for the significance of the model generated are performed in the ANOVA of the data generated. These tests determine if the model results are likely to be due to random noise in the experiment as opposed to being due to the changes in the conditions under which refolding has occurred. These model F-values are summarised in table 3.5.

**Table 3.5:** Model F-values and chance that the model F-value could be due to random chance for the ANOVA analysis of the fractional factorial refolding screen of five protein kinases. Model F-values indicated as significant by stars. Significance level of F-test was 5 %.

Protein Refolded	Model F-value	Chance that Model F-value could be Due to Random Noise
AKT2	4.30 *	0.24 %
KIS	10.69 *	<0.01 %
p38 $\alpha$	11.65 *	<0.01 %
PhK	3.70 *	0.67 %
TTK	2.92 *	1.91 %

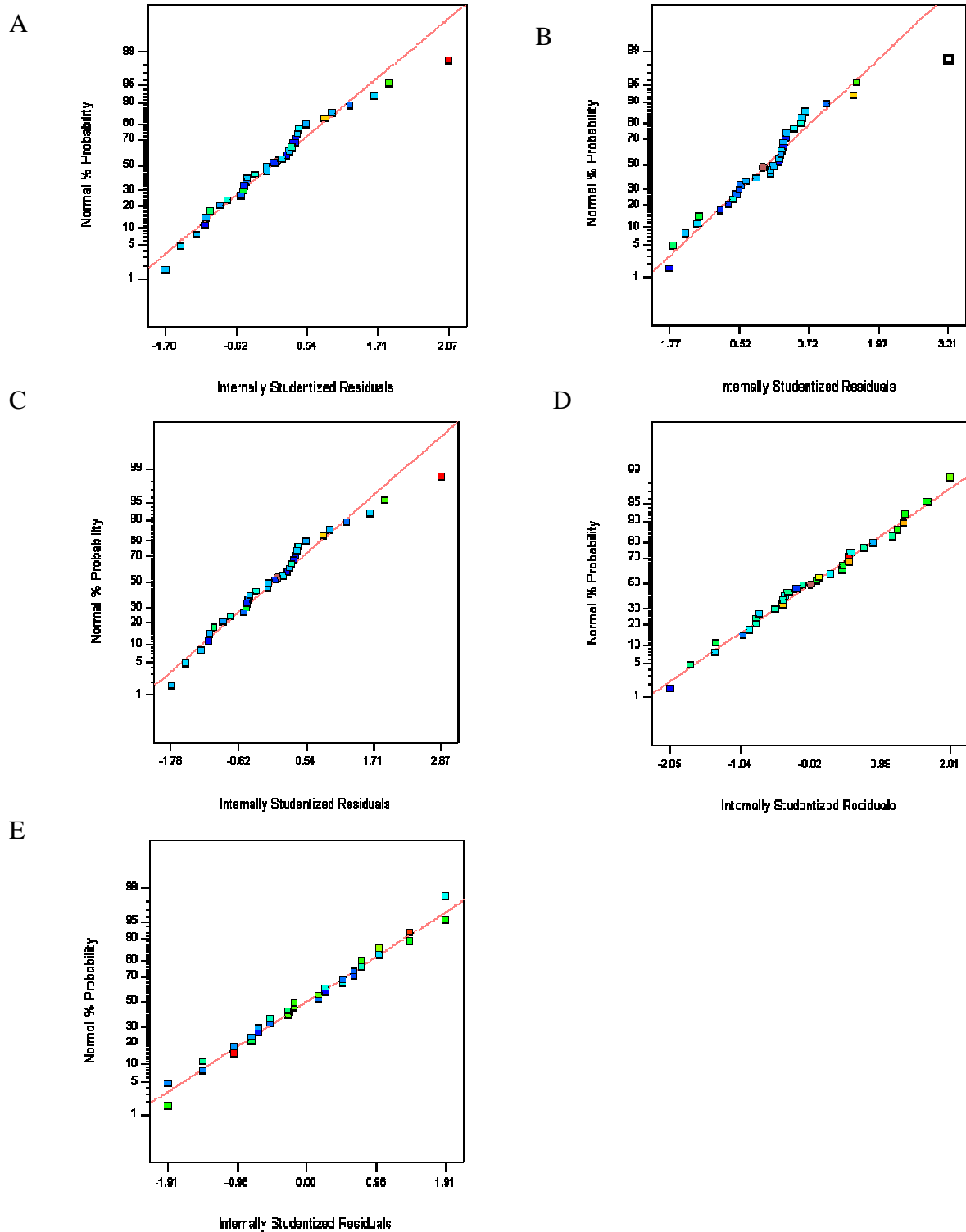
In addition, the residuals between the predicted results determined by the analysis of the screen and the actual results obtained in the experiments performed using the fractional factorial screen can be examined to determine if the model explains all the results. The residuals from the model are studentised, that is they are divided by an estimate of the standard deviation of the residual. The plot of the studentised residuals against the experiment number gives an indication of the quality of the fit of the model generated. A studentised residual of greater than 3 indicates a poor fit of the model to the data for that experiment. This may indicate an underlying effect not accounted for in the model, or a poor experimental result due to a mistake in the experimental procedure. These residual plots for the five kinases screened are shown in Figure 3.24 A-E.





**Figure 3.24:** Studentised residuals for the fit of fractional factorial screen model against actual data. Red lines indicate cut-off level for studentised residuals of 3. (A) AKT2 residuals. (B) KIS residuals. (C) p38 $\alpha$  residuals. (D) PhK residuals. (E) TTK residuals. Points coloured by value of data for each run number; Blue – lowest value ; Red – highest value.

The plots of residuals against experiment number indicate that the models generated for the refolding screen are well fitted to the data collected. A second measure of the quality of the fit of the refolding screen model to the data is to examine the normality of the residuals. If there is no systematic bias in the data, then it would be expected that the residuals would be normally distributed. A plot of the expected cumulative probability of the normal distribution against the residual value is generated by calculating the cumulative probability of each residual according to equation 3.1. If the residuals were perfectly normally distributed then all the residuals would lie on a line indicated on the normal plots. Systematic or large deviations from this line would indicate a problem with the data, indicating that a transform of the data should be used to improve the fit of the model, or that the model should be changed to better fit the data. The normal plots of residuals for the fractional factorial refolding screen for the five kinases to which the screen was applied are shown in Figure 3.25 A-E.



**Figure 3.25:** Normal probability plot of the residuals for the fit of fractional factorial screen model against actual data. Red lines indicate theoretical normal distribution of residuals. (A) AKT2 residuals. (B) KIS residuals. (C) p38 $\alpha$  residuals. (D) PhK residuals. (E) TTK residuals. Points coloured by value of data for each run number; Blue – lowest value ; Red – highest value.

The normal probability plots indicate that the residuals for the fractional factorial screen model are close to normally distributed, showing the quality of the fit. A single value has been excluded from the fit in the refolding screens performed on KIS. This high value has a studentized residual of greater than 3 and is not close to the theoretical normal distribution of residuals.

The results of the ANOVA of the screen results indicate the factors and combination of factors which produce a significant improvement in the recovery of refolded protein. The refolding screen performed on AKT2 identified both positive and negative factors involved in the refolding of AKT2. The significance level for factors was set at 5 %. Seven factors or combinations of factors were identified by the modelling of the refolding data to be significant in their effect upon the refolding of AKT2 compared to the base condition of 50 mM Tris pH 8.0. Of these seven terms, four had a positive impact on the refolding of AKT2 and three had a negative impact on the refolding of AKT2. The significant, negative terms were AD, BE and DE, or NV-10 / CAPSO, TMAO / PEG3440 and NV-10 / PEG3440. The significant positive factors were D, AE, BC and ADE, or NV-10, PEG / CAPSO, TMAO / urea and CAPSO / NV-10 / PEG3440. However, since the screen is aliased, CAPSO / NV-10 / PEG3440 cannot be distinguished from TMAO / urea / staurosporine. These results are summarised in table 3.6.

**Table 3.6:** Significant positive and negative factors in the refolding of AKT2 identified by the ANOVA of the results of a fractional factorial screen of AKT2 refolding.

Significant Negative Factors	Significant Positive Factors
NV-10 / CAPSO	NV-10
TMAO / PEG3440	PEG3440 / CAPSO
NV-10 / PEG3440	TMAO / Urea
	CAPSO / NV-10 / PEG3440 or
	TMAO / Urea / Staurosporine

The most interesting combination of positive additives was the combination of 1 M TMAO and 0.5 M urea. It is possible that this combination of factors, which are opposite in their effect on native protein, allow for easier conformational shuffling, which aids the protein in refolding. NV-10 was identified as a positive factor in the refolding of AKT2, but

was not so identified in the initial screen. It is possible that the more robust data analysis of the screen performed has identified an effect missed in the initial screen. It is also noteworthy that the combination of CAPSO and NV-10 had a negative effect on the recovery of refolded protein.

The refolding of KIS was screened and four factors, or combinations of factors were identified which were significant in altering the recovery of refolded protein from the base refolding yield. No factors which had a significant, negative effect on the refolding of KIS were identified through the ANOVA of the screen. The factors identified which had a positive impact on the refolding of KIS were D, F, AE, and ADE, or NV-10, staurosporine, PEG3440 / CAPSO and NV-10 / PEG3440 / CAPSO. The final term is aliased with TMAO / urea / staurosporine. Several of the positive factors are found in common between AKT2 and KIS. NV-10 PEG3440 / CAPSO and CAPSO / NV-10 / PEG3440 or TMAO / urea / staurosporine. The common feature of these refolding additives is that they have a similar mode of action. These additives function by inhibiting the aggregation of the refolding protein. However, the three factor term is aliased and could, in fact, be different between the two kinases.

p38 $\alpha$  was also screened through the fractional factorial screen, and was analysed in the same fashion as the other four kinases of the kinase panel without the benefit of soluble p38 $\alpha$  based standard curves for the analytical readout methods. This was to allow direct comparison between the five kinases used in the kinase panel. Two significant positive factors and two significant negative factors were found. The negative significant factors were B and ABE or TMAO and TMAO / PEG3440 / CAPSO. The significant positive factors were E and BD or PEG3440 and TMAO / NV-10. The base recovery of monomeric protein is high for the refolding of p38 $\alpha$ , and the effect of the significant positive additives on the refolding of p38 $\alpha$  was lower than the effect of the significant positive factors on the refolding of AKT2 and KIS.

The refolding of PhK was also performed using a fractional factorial screen, and the fit of the model generated on the analysis of the refolding recoveries to the actual data was significant (Table 3.5). The ANOVA of the screen data identified eight factors or combinations of factors which exerted a significant effect on the recovery of refolded protein from the refolding screen. Of these factors, three factors had a negative effect on the refolding of PhK and five factors had a positive effect on the refolding of PhK, increasing the

recovery of refolded protein. The factors which gave a negative effect were B, BC and BF or TMAO, TMAO / 0.5 M urea and TMAO / staurosporine. The positive factors were C, CD, CF ABC and ABF or 0.5 M urea, 0.5M urea / NV-10, 0.5 M urea / staurosporine, 0.5 M urea / TMAO / CAPSO or NV-10 / PEG3440 / staurosporine and TMAO / staurosporine / CAPSO or 0.5 M urea / NV-10 / PEG3440. These factors are summarised in table 3.7.

**Table 3.7:** Positive and negative factors identified as significant by ANOVA of the fractional factorial refolding screen performed on PhK.

Significant Negative Factors	Significant Positive Factors
TMAO	0.5 M Urea
TMAO / 0.5 M Urea	0.5 M Urea / NV-10
TMAO / Staurosporine	0.5 M Urea / Staurosporine
	0.5 M Urea / TMAO / CAPSO or
	NV-10 / PEG3440 / Staurosporine
	TMAO / Staurosporine / CAPSO or
	0.5 M Urea / NV-10 / PEG3440

Although the three factor terms are aliased and cannot be distinguished from the two sets of three factor terms given, TMAO is common as a negative factor with both factors with which it is included in a three factor term. It is likely therefore, that the true three factor terms are NV-10 / PEG3440 / staurosporine and 0.5 M urea / NV-10 / PEG3440. The common presence of the factors 0.5 M urea and NV-10 in the positive factors from the screen is parallel to the results from the initial screen performed on PhK (Figure 3.17), which show a positive effect on the refolding of PhK by 0.5 M urea and by NV-10.

The final protein kinase to be refolded using the fractional factorial screen was TTK. The screen model was found to be significant, with a low chance of the model being due to random noise (Table 3.4). The model identified seven significant factors affecting the recovery of refolded protein. Four negative significant factors were identified. These factors were CF, DF, ABC (DEF) and ABE (CDF) or 0.5 M urea / staurosporine, NV-10 / staurosporine, TMAO / 0.5 M urea / CAPSO or NV-10 / PEG3440 / staurosporine and TMAO / PEG3440 / CAPSO or 0.5 M urea / NV-10 / staurosporine (Table 3.8). In addition,

the combination of factors AD or NV-10 / CAPSO has a small, not significant negative effect.

**Table 3.8:** Positive and negative factors identified as significant by ANOVA of the fractional factorial refolding screen performed on TTK.

Significant Negative Factors	Significant Positive Factors
0.5 M Urea / Staurosporine	NV-10
NV-10 / Staurosporine	TMAO / CAPSO
TMAO / 0.5 M Urea / CAPSO or NV-10 / PEG3440 / Staurosporine	0.5 M Urea / PEG3440
TMAO / PEG3440 / CAPSO or 0.5 M Urea / NV-10 / Staurosporine	

The positive factors identified by the ANOVA of the screen were D, AB and CE or NV-10, TMAO / CAPSO and 0.5 M urea / PEG3440 (Table 3.8). There is a strong similarity between the factors that had a positive effect on the refolding of TTK identified in the fractional factorial screen, and those identified in the initial refolding screen (Figure 3.21). NV-10 was shown to be effective in both screens, and the lower recovery of refolded protein at pH 9.5 compared to pH 8.0 was reflected in the results of the fractional factorial screen. Staurosporine was a strongly negative factor in the refolding of TTK measured in the fractional factorial screen as it appears in most of the negative factors. Its effect is such that it counteracts positive additives such as NV-10. The selection of buffer did not have a significant impact on the recovery of refolded protein, paralleling the effect of pH and buffer seen with the initial refolding screen.

The screen of the refolding of the five protein kinases, AKT2, KIS, p38 $\alpha$ , PhK and TTK, through the fractional factorial screen developed has identified several positive and negative factors in the refolding of the kinases tested. The factors which were found to be significant in the refolding of the tested kinases are summarised in table 3.9.

**Table 3.9:** Significant positive and negative factors identified in ANVOA of fractional factorial screens performed on five protein kinases.

<b><u>Protein</u></b>	<b><u>Factors Affecting Refolding</u></b>		<b><u>Protein</u></b>	<b><u>Factors Affecting Refolding</u></b>	
	<b>Significant Negative Factors</b>	<b>Significant Positive Factors</b>		<b>Significant Negative Factors</b>	<b>Significant Positive Factors</b>
AKT2	NV-10 / CAPSO TMAO / PEG3440 NV-10 / PEG3440	NV-10 PEG3440 / CAPSO TMAO / 0.5 M Urea CAPSO/NV-10/PEG3440 or TMAO/0.5 M Urea/Staurosporine	PhK	TMAO TMAO / 0.5 M Urea TMAO / Staurosporine	0.5 M Urea 0.5 M Urea / NV-10 0.5 M Urea / Staurosporine 0.5 M Urea / TMAO / CAPSO or NV-10/PEG3440/Staurosporine TMAO/Staurosporine/CAPSO or 0.5 M Urea / NV-10 / PEG3440
KIS (1-313)		NV-10  Staurosporine PEG3440 / CAPSO NV-10 / PEG3440 / CAPSO or TMAO/0.5 M Urea/Staurosporine	TTK (514-804)	0.5 M Urea / Staurosporine NV-10 / Staurosporine TMAO / 0.5 M Urea / CAPSO or NV-10 / PEG3440 / Staurosporine TMAO / PEG3440 / CAPSO or 0.5 M Urea/NV-10/Staurosporine	NV-10 TMAO / CAPSO 0.5 M Urea / PEG3440
p38 $\alpha$	TMAO  TMAO / PEG3440 / CAPSO or 0.5 M Urea / NV-10 / Staurosporine	PEG3440  TMAO / NV-10			



### 3.4 Discussion

The refolding screen described in chapter 2 was developed and tested using p38 $\alpha$  as a model protein kinase. To demonstrate that the refolding screen is suitable for testing for the refolding of protein kinases the screen was applied to a series of human protein kinases and the results examined to determine if the refolding screen was suitable for determining conditions which would support the refolding of protein kinases which have proved difficult to express in *E. coli*. In addition the screen was developed to examine the effect of the combination of additives on the refolding of protein kinases. A fractional factorial screen was exploited to provide a testing of the interaction of a limited number of screen additives and to provide a more robust test of the significance of the effect of different additives on the refolding of the test protein kinases through ANOVA of the screening results.

Four protein kinases from the serine/threonine and dual specificity classes of protein kinases, with tyrosine kinases being excluded. The protein kinases selected were distributed through the kinome, in order to avoid close sequence similarity between the kinases selected. KIS and TTK are shown to be closely related in the analysis of the kinome produced by Manning *et al.* (2002). However the sequence identity of these two kinases is only 43 % despite their close relationship. Despite this close relationship between KIS and TTK there were important differences noted in the refolding of the two proteins in both screens under which refolding was examined. In the initial screen, a significant difference between the refolding of KIS at pH 9.5 and pH 5.8 was observed. However, the comparison of the average recoveries of soluble protein for the refolding of TTK did not result in a significant difference between pH 5.8 and pH 8.0 and pH 9.5. Despite their diverse sequences (Figure 3.1) the four protein kinases of the kinase panel that have a published crystal structure, AKT2, p38 $\alpha$ , PhK and TTK, adopt a common structure. The mainly  $\alpha$ -helical C-terminal lobe and mainly  $\beta$ -sheet N-terminal lobe of the kinase fold are seen to be common (Figure 3.2). It is also expected that the KIS kinase domain adopts a similar structure. Given the similarity of the folds of the protein kinases, and their relative sequence diversity the kinase panel selected forms an excellent tool for demonstrating the suitability of the refolding screens for all protein kinases, and for helping to address the question of whether the folding of the kinase domain follows a common route.

The refolding screen presented in chapter 2 was applied to the four additional protein kinases selected, AKT2, KIS, PhK and TTK. Soluble, control protein was not available for three of these protein kinases, and so the analytical readouts were modified slightly to account for this difference to the studies on the refolding of p38 $\alpha$ . The correlation between the monomeric protein recovery and the response units from the binding assay were used as a control for the quality of the SPR data. Likewise the comparison of the soluble protein recovery and the monomeric protein recovery was used to shed light on potential reasons for any discrepancies between the monomeric recovery and the response units from the SPR based assay.

The refolding of AKT2 resulted in a small number of conditions which gave measurable recoveries of protein in the analytical size exclusion assay, and measurable response units in the SPR-based binding assay (Figure 3.7). Although the number of conditions is low, there is a significant, positive correlation between the two assays. In addition the recovery of soluble protein is similar to the recovery of monomeric protein. This is indicative of the presence of soluble oligomers. It appears, however that these soluble oligomers are distinguished from the monomeric protein by the binding assay (Figure 3.8).

The refolding screen performed on KIS identified 10 conditions which promoted the refolding of KIS (Figure 3.13). In these conditions the correlation between the monomeric protein recovery and the SPR-based binding assay was strong and significant (Figure 3.12). The relation of the soluble protein recovery and the monomeric protein recovery was mixed (Figure 3.13).

The refolding of PhK performed as a positive control for the screen. The correlation of the monomeric recovery and response units from the SPR-based binding assay was negative and not significant. The SPR-based binding assay does not appear to be suitable for detecting the refolding of PhK (Figure 3.16) and the analysis of the refolding of PhK centred on the results of the analytical size exclusion assay. The refolding of PhK also produced large amounts of soluble oligomers, as indicated by the large differences between the soluble protein recovery and the monomeric protein recovery (Figure 3.17). However, some conditions did give similar levels of soluble and monomeric recovery (Figure 3.17).

The final kinase tested using the screen outlined in chapter 2 was TTK. The refolding of TTK did not show a dependence upon pH (Figure 3.19), a result which was supported by the thermal melting analysis of TTK in different buffers, which saw the thermal melting

midpoint of TTK being similar in a range of buffers and at a range of pHs (Figure 3.22, Table 3.2). The correlation between the monomeric protein recovery and the SPR based activity assay response units was not significant and negative (Figure 3.20). However, the soluble and monomeric protein recoveries were similar, showing that limited amounts of soluble oligomers were formed in refolding.

The five protein kinases tested in the initial refolding screen showed differences in their patterns of refolding and in the behaviour of the refolded protein. KIS (Figure 3.13), PhK (Figure 3.15) and p38 $\alpha$  (Figure 2.9) showed a dependence of their refolding yields on the pH of the buffer in which they were refolded. However, TTK (Figure 3.19) and AKT2 (Figure 3.8) did not show significant differences in the recoveries of soluble protein at different pHs. Based on the thermal unfolding of p38 $\alpha$  (Figure 2.11) and TTK (Figure 3.22) this difference is probably due to a difference in the stability of the native state in these conditions. p38 $\alpha$  appears to be partially unfolded at pH 5.8, whereas TTK appears to be unaffected by this pH. The use of high pH is, therefore, beneficial to the refolding of protein kinases, but is not essential in all kinases. The pIs of the kinases do vary, with TTK having a higher pI than the other kinases used (Table 3.1). However, AKT2 has a pI similar to those of KIS, PhK and p38 $\alpha$ . This suggests that an effect more specific than the changes in pI relative to the pH at which refolding has occurred is responsible for the differences in the refolding yields obtained. The role of histidine residues in the structure to the different kinases is a possible candidate for the reason for the differences between the response of the kinases to the pH of refolding.

To examine the effect of the interaction of additives, and to improve the robustness of the identification of refolding conditions a fractional factorial screen was used. To construct this screen a set of additives, four in total were selected on the basis of their effects in the initial screens on the refolding of the protein kinases tested. A wide variety of additives permitted the refolding of p38 $\alpha$ , so only those additives which gave an improvement in the recovery of monomeric protein over the control conditions at pH 8.0 and pH 9.5 were considered. No additive which was included in the initial protein refolding screen was effective at improving the recovery of monomeric protein for all five kinases tested. The best additives for affecting multiple kinases were PEG3440, which supported the refolding of KIS (Figure 3.13), p38 $\alpha$  and PhK (Figure 3.17) and 0.5 M urea which also supported the refolding of KIS (Figure 3.13), p38 $\alpha$  and PhK (Figure 3.17). NV-10 supported the refolding of PhK (Figure 3.17) and TTK (Figure 3.21) but resulted in a decrease in the recovery of monomeric

protein with p38 $\alpha$ . The recoveries of monomeric protein do not form similar patterns between any two protein kinases which were refolded with the initial screen.

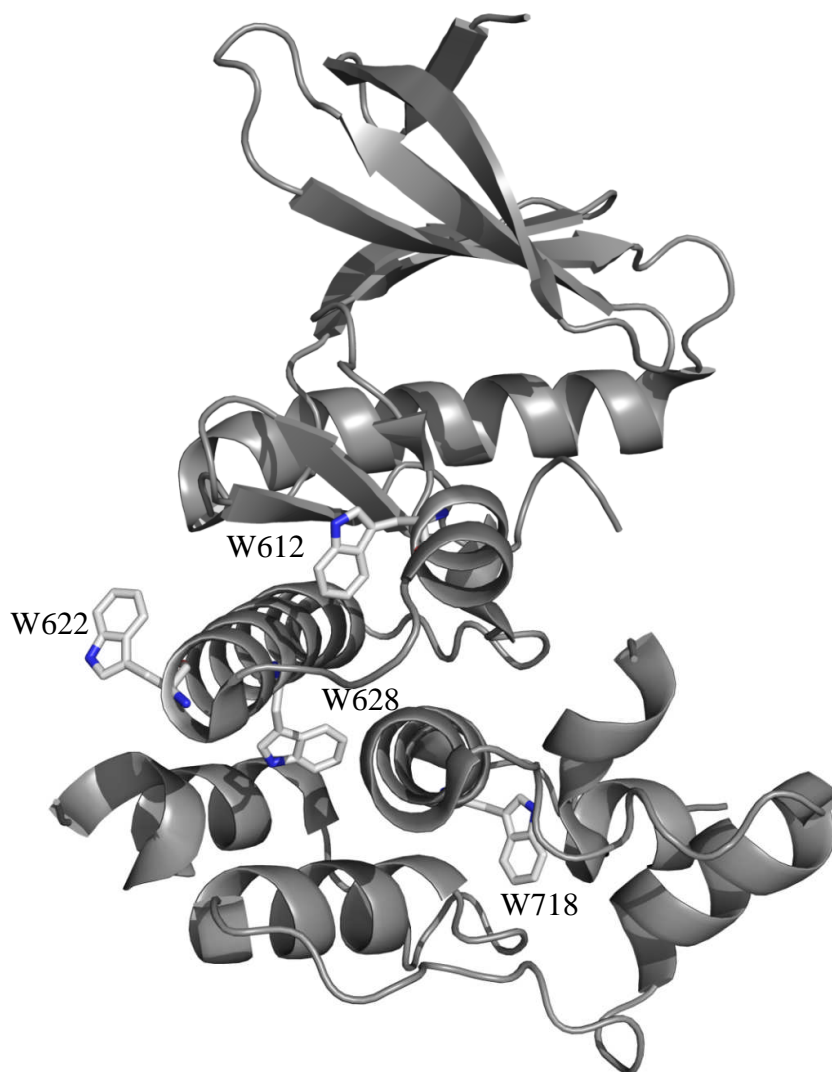
The screen was performed and analysed using the same procedures as the initial refolding screen so that the results could be directly compared. The maximum recovery of monomeric protein was increased from the initial screen, and several combinations of factors were found to be effective in increasing the refolding yields of some of the kinases tested (Table 3.9). The effect of the significant additives and combinations of additives were mixed. In several cases, the same additives were found to have a positive impact on the refolding on one protein kinase and a negative effect on the refolding of a second protein kinase. No one additive was a positive or negative factor in the refolding of all of the protein kinases tested (Table 3.9). Overall, it is clear that in the fraction factorial refolding screen no consistent effect can be observed on the refolding of protein kinases by a single additive or combination of additives.

The two refolding screens performed on the selected protein kinases have shown that it is possible to refold many protein kinases in a screen format, and to identify the extent of the refolding through various analytical techniques. The results have underlined the importance of having multiple analytical techniques which can be compared to identify which conditions lead to the recovery of correctly folded, monomeric protein. The fractional factorial screen has succeeded in producing detailed, statistically relevant results from a smaller screening format than was required initially. The combination of additives has been shown to be effective in supporting the refolding of protein kinases. The effect of additives on protein kinases has been shown to depend on the individual kinase being refolded, rather than having a common effect on protein kinases. This result undermines the concept of a single folding mechanism underlining the folding of the kinase domain. If a common mechanism of folding existed, it would be expected that more similarities in the pattern of refolding would be found, with respect to the response to the changes in pH and in different refolding additives. The lack of similarities in the best refolding conditions observed, does not, however, provide conclusive evidence about the folding pathways, since the additives may have little effect on the folding pathway, instead acting to prevent aggregation. The further examination of the folding pathway of a protein kinase is necessary to determine the existence and extent of any similarities.

## Chapter 4. Mutagenesis, Expression and Characterisation of Human TTK tryptophan Mutants

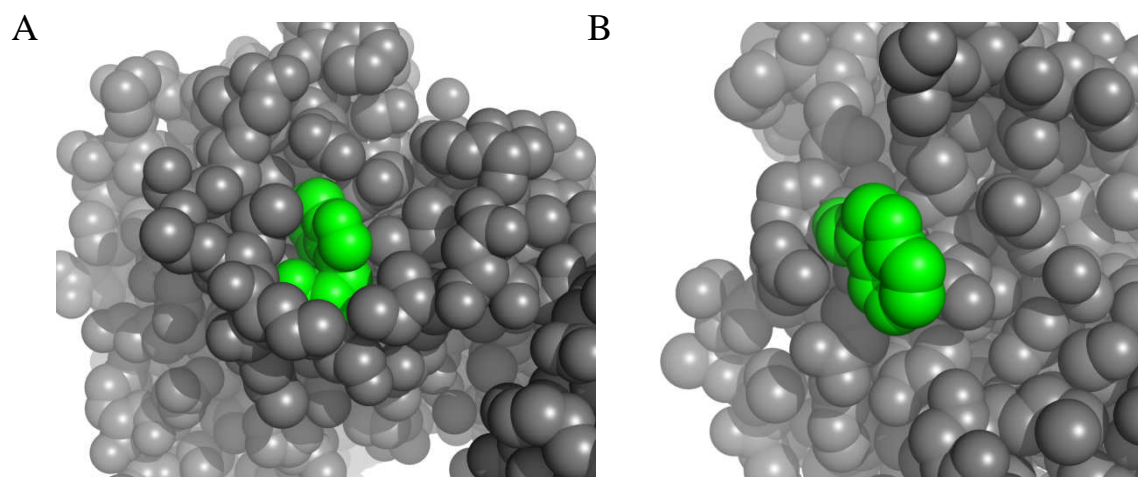
### 4.1 Introduction

This chapter describes the expression and characterisation of wild type TTK kinase domain and tryptophan to phenylalanine substitution mutants of the TTK kinase domain. The creation of tryptophan to phenylalanine substitution mutants is key for the study of the folding of TTK, and will allow comparison of the folding of TTK and the folding of p38 $\alpha$  examined by Davies (2004).



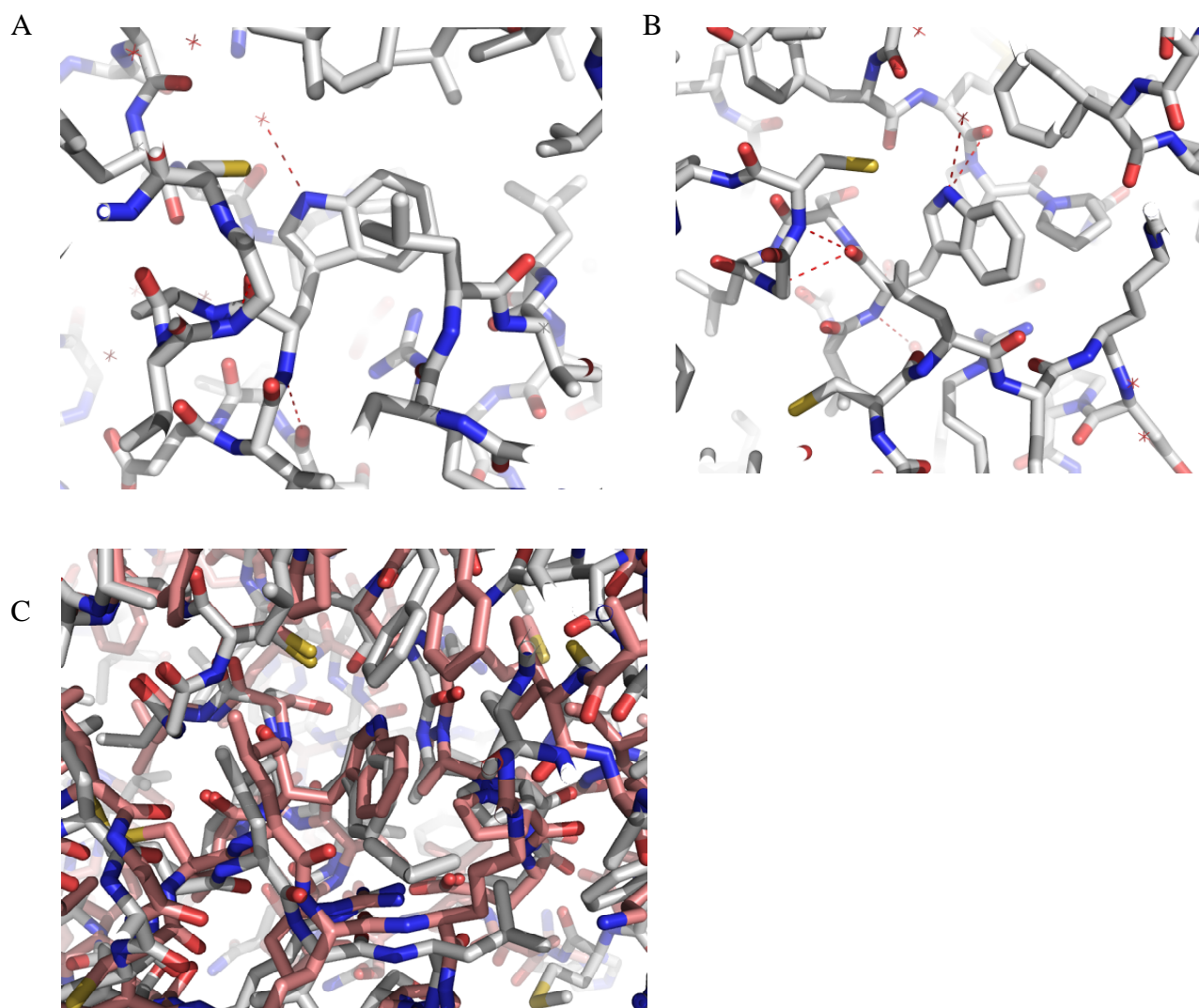
**Figure 4.1:** Crystal structure of human TTK kinase domain with native tryptophan residues highlighted. The structure consists of 2 lobes, a  $\beta$ -sheet rich N-terminal lobe, and an  $\alpha$ -helical C-terminal lobe. All four native tryptophan residues are contained in the C-terminal lobe. Structure from pdb file 3CEK, rendered using ray tracing module of Pymol (DeLano, 2008).

The TTK kinase domain contains four native tryptophan residues, W612, W622, W628 and W718 (Figure 4.1). These four residues are all contained within the C-terminal lobe of the kinase domain, and are found in different environments. For example, the crystal structure of the TTK kinase domain indicates that the residues 612 and 622 are found in a solvent exposed position (Figure 4.2), and residues 628 and 718 are buried in the core of the C-terminal lobe of the protein (Figure 4.1).



**Figure 4.2:** Space filling representation of the local environment around tryptophan residues 612 (A) and 622 (B) showing that these residues are exposed to solvent in the native structure Tryptophan residues highlighted in green. Structure from pdb file 3CEK, rendered using ray tracing module of Pymol (DeLano, 2008).

Residue 718 is the structural and sequence homologue of tryptophan 207 of p38 $\alpha$  (Figure 3.1). This residue is found to be conserved throughout the kinome and was shown to be key for the correct folding p38 $\alpha$  (Davies, 2004). In addition if the local environments of W718 (TTK) and W207 (p38 $\alpha$ ) are examined in the respective crystal structures, it can be seen that the environments of these residues is conserved as well as the residues themselves (Figure 4.3). Therefore, the folding similarities of these two proteins, as studied by their respective single tryptophan mutants may reveal important insights into whether common folding pathways exist in the kinome.



**Figure 4.3:** Local environment of the core tryptophans of the kinase domains of (A) p38 $\alpha$  and (B) TTK, showing hydrogen bonds to the core tryptophan (red dotted lines). Panel C – structure alignment of p38 $\alpha$  and TTK showing close alignment of the core tryptophans and the similarity of their local environments. p38 $\alpha$  - grey, TTK – pink. Red crosses represent ordered water molecules. Structures from pdb files 1WFC and 3CEK respectively, rendered using ray tracing module of Pymol (DeLano, 2008).

The fluorescence of the protein is used to follow the folding of the protein under conditions of denaturation and renaturation. If an excitation wavelength of 295 nm is used, then the fluorescence is dominated by the emission of the tryptophan residues in the protein. The fluorescence of the tryptophan side chain is highly sensitive to the local environment and undergoes a substantial change in the fluorescence yield and in the wavelength of maximum emission when the residue changes from a solvated to a hydrophobic environment. This sensitivity makes tryptophan residues an excellent probe for folding studies. However, many large proteins contain several tryptophan residues, and the contribution of these different

residues in different local environments makes the changes in the emission spectra of the tryptophans on folding difficult to interpret. To make the spectra easier to interpret, the number of tryptophan residues must be decreased. The conservative mutation of a tryptophan to a phenylalanine residue is used to decrease the number of tryptophans, ideally to a single tryptophan, whilst having a minimal effect on the structure of the protein.

Site directed mutagenesis was used to create tryptophan to phenylalanine substitutions in the TTK kinase domain. These mutants were purified and the removal of the purification tag, and the phosphorylation state of the expressed proteins determined by ESI-MS. Thermal melting and CD spectra analysis were used to determine if the tryptophan to phenylalanine substitutions had had a deleterious effect on the fold of the protein.

## ***4.2 Materials and Methods***

### ***4.2.1 Materials***

Reagents for site directed mutagenesis, including *E. coli* strain XL1 Blue supercompetent cells were purchased from Stratagene (La Jolla, CA, USA). Restriction enzymes used were purchased from New England Biolabs (Ipswich, MA, USA). Oligonucleotide primers were purchased from Eurogentec (Southampton, UK). Chromatography media was from GE Healthcare (Amersham, UK). Sypro-Orange was from Invitrogen (Carlsbad, CA, USA). All other chemicals were purchased from Sigma Aldrich (Poole, UK)

### ***4.2.2 Site Directed Mutagenesis***

Site-directed mutagenesis was carried out using the QuikChange II site directed mutagenesis kit (Stratagene). This kit uses a *DpnI* digestion of methylated and hemi-methylated DNA. PCR mutagenic primers were designed using the manufacture's online tools, and synthesized by Eurogentec. These primers are used to introduce the mutation via PCR replication of a methylated template DNA. This produces double stranded plasmids, with methylated or hemi-methylated plasmids not containing the desired mutation. This methylated DNA is digested by *DpnI* before the reaction mixture was used to transform XL1-blue supercompetent cells.



Briefly, forward and reverse oligonucleotide primers of 36 to 50 bases were used in a PCR with template DNA, which had been produced in an *E. coli* strain that methylates DNA. The thermal cycling conditions that were used are; 95 °C for 30 seconds, followed by 16 cycles of 95 °C for 30 seconds, 55 °C for 1 minute and 68 °C for 5 minutes 30 seconds. Following this the reaction was cooled to 20 °C. The resultant PCR product was digested with *DpnI* for 1 hour at 37 °C. XL1-blue supercompetent cells were transformed with 4 µL of the digested DNA as described in section 2.2.2. Plasmid DNA was then isolated using the method described in section 2.2.3. Incorporation of the desired mutations was confirmed by restriction digest using enzymes with cut sites incorporated into the sequence by the mutation performed, or by DNA sequencing.

#### **4.2.3 Expression of Wild Type TTK and TTK Tryptophan Mutants**

Aliquots of *E. coli* strain BL21\* (DE3) competent cells, transformed with a plasmid encoding λ-phosphatase and prepared according to section 2.2.3, were transformed with the wild type TTK (514-820) expression construct, or with a tryptophan mutant expression construct according to section 2.2.2. Coexpression of λ-phosphatase with the TTK kinase domain is known to be required for the kinase domain to be expressed in a soluble form in *E. coli* (AstraZeneca unpublished data). For expression of TTK, a single colony was picked from the agar plate and used to inoculate a 75 mL culture of Terrific Broth (Section 2.2.4) containing Kanamycin at 50µg/mL and Tetracycline at 12.5 µg/mL. This culture was incubated overnight at 37 °C. Expression cultures of 600 mL of Terrific Broth containing Kanamycin at 50 µg/mL and Tetracycline at 12.5 µg/mL were inoculated with 5 mL of the previous culture and incubated in a shaking incubator at 37 °C and 180 rpm. When the OD<sub>600</sub> of the culture reached ~0.45 the temperature was reduced to 20 °C. At OD<sub>600</sub> ~0.6 the expression of TTK was induced by the addition of IPTG to a final concentration of 0.1 mM. The culture was incubated for 20 hours and the biomass harvested after this time by centrifugation at 6000 g for 20 minutes at 4 °C.

#### **4.2.4 Purification of Wild Type TTK and TTK Tryptophan Mutants**

Cell paste obtained from 7.2 litres of cell culture was resuspended in 1 litre of 50 mM Tris, 200 mM NaCl, 10 mM Imidazole, 4 mM DTT, pH 8.0 using a homogeniser. Lysozyme (SigmaAldrich) was added to a final concentration of 0.5

mg/mL, and the resuspended cells incubated for 3 hours at 4 °C. The lysis of the *E. coli* cells was then completed by sonication. This was performed by four rounds of 30 s sonication followed by 2 minutes relaxation time. Samples were kept on ice between periods of sonication. Insoluble material was then separated by centrifugation at 35000 g at 4 °C for 30 minutes.

The recovered supernatant containing soluble TTK was then applied to a 35 mL NiNTA superflow column (GE Healthcare) equilibrated in 50 mM TrisHCl, 200 mM NaCl, 10 mM Imidazole, 4 mM DTT, pH 8.0 using a P1 peristaltic pump (Pharmacia) at a flow rate of 5 mL/min. Once this was completed the column was transferred to an AKTA purifier LC system equilibrated in the same buffer as the column. Bound proteins were then eluted by raising the concentration of Imidazole in the buffer using a step elution at a flow rate of 5 mL/min. Six steps were used, step 1 – 100 mL buffer 10 mM Imidazole; step 2 – 100 mL buffer 20 mM Imidazole; step 3 – 70 mL buffer 50 mM Imidazole; step 4 – 70 mL buffer 100 mM Imidazole; step 5 – 70 mL buffer 250 mM Imidazole; step 6 – 100 mL buffer 500 mM Imidazole.

Fractions containing TTK were collected and pooled. The protein concentration was estimated by measurement of the  $A_{280}$  of the solution, assuming that 1  $A_{280}$  (10mm pathlength) = 1 mg/mL [Protein]. TEV protease (AstraZeneca) was added to the pooled fractions, with the amount of added TEV depending on the estimated total protein content of the pooled fractions. TEV protease was added for the purpose of cleaving the N-terminal 6His tag from the expressed protein (Figure 4.4). The pooled fractions with TEV protease were dialysed overnight against 20 mM TrisHCl, 5 mM DTT, pH 7.4 using dialysis tubing with a 6-8 kDa molecular weight cut-off.

The pooled fractions were applied to a 5 mL NiNTA Superflow column equilibrated in 20 mM Tris 5 mM DTT pH 7.4. Cleaved TTK was recovered from the flow through, and uncleaved TTK, cleaved 6His tag and TEV protease were eluted from the column using a step elution with increasing Imidazole concentrations as previously.

Fractions containing cleaved TTK were concentrated using a centrifugal concentration device with a 10 kDa molecular weight cut-off. The concentrated solution was applied to a 15 mL Source 30S cation exchange column equilibrated in 20 mM TrisHCl, 5 mM DTT pH 7.4 at a flow rate of 2.5 mL/min and TTK was

eluted with a linear gradient of 0-250 mM NaCl over 30 min. The binding of TTK to the column was of low affinity, so the flow through was reapplied to the column after the elution of bound TTK to maximise the recovery of TTK.

#### ***4.2.5 Tryptophan Fluorescence of TTK***

Fluorescence emission spectra were measured using a Perkin-Elmer LS50B spectrofluorimeter. Data were collected and the instrument controlled by FLWinLab 2 software (Perkin-Elmer). A 10 mm square quartz cell was used, with spectra being recorded at 20 °C. Excitation and Emission slits were set to a bandwidth of 5 nm, with an excitation wavelength of 295 nm. Emission spectra were collected at a scanning speed of 60 nm per minute from 305 nm to 450 nm and at a resolution of 0.5 nm. Four scans were averaged to produce a single spectrum. Appropriate buffer blanks were subtracted from each spectrum.

#### ***4.2.6 Circular Dichroism of TTK***

Far-UV circular dichroism spectra of TTK (514-804) and tryptophan mutants were collected using a JASCO J-810 spectropolarimeter. A 1 mm pathlength quartz cuvette was used and the spectra were recorded at 20 °C. Spectra were collected at a scanning rate of 50 nm per minute with a bandwidth and resolution of 1 nm. The time constant was 0.5 s. 16 scans were averaged to produce a single spectrum.

The spectra were collected from 260 nm to between 210 nm and 190 nm depending on the GdnHCl concentration in the sample analysed. As the concentration of denaturant decreased from 6M the lower limit of the spectra was lowered in a manner that maximised the information content of the spectra, whilst minimising the time that the HT voltage for the photomultiplier tube was at maximum (950 V).

#### ***4.2.7 Curve Fitting for Tryptophan Emission Spectra***

Spectral broadening of the TTK fluorescence spectra was determined by fitting the tryptophan emission spectra to a log normal distribution, described by equation 4.1, 4.2 and 4.3 (Burstein *et al.*, 2001), to obtain the  $\lambda_{\text{max}}$  and the spectral width of the spectrum.

$$f = I_{\max} e^{\left( \frac{\ln 2}{\ln^2 p} \ln^2 \left( \frac{a - \lambda}{a - \lambda_{\max}} \right) \right)} \quad (4.1)$$

$$a = \lambda_{\max} + \left( \frac{p(v^+ - v^-)}{p^2 - 1} \right) \quad (4.2)$$

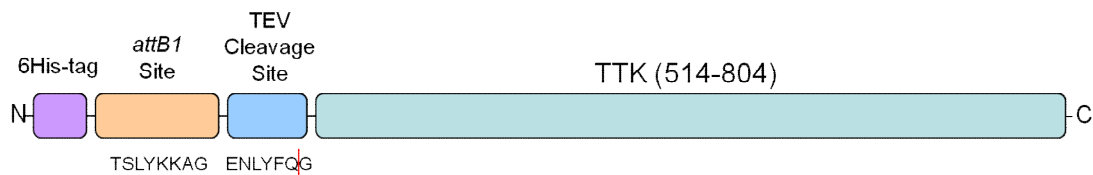
$$p = \frac{(\lambda_{\max} - v^-)}{(v^+ - \lambda_{\max})} \quad (4.3)$$

Where  $I_{\max}$  is the maximum intensity,  $\lambda_{\max}$  is the wavelength of maximum fluorescence intensity,  $v^+$  and  $v^-$  are the wavelengths at half peak height, and  $p$  is a measure of the asymmetry of the curve. The values derived from curve fitting were compared to standard values to determine if spectral broadening of the tryptophan fluorescence had occurred.

## 4.3 Results

### 4.3.1 Wild Type TTK Expression Construct

An expression construct for the kinase domain of wild type TTK was supplied by AstraZeneca. It contains a sequence representing residues 514-804 of the human TTK full length sequence which had been codon optimised for expression in *E. coli* and was flanked by *attB* sites. The insert contains, N-terminal of the TTK kinase domain sequence, an *attB1* site for Gateway cloning and a sequence encoding a TEV protease cleavage site. This insert had been recombined into a pT7#3.3 expression vector, which added an N-terminal 6His tag and contained a tetracycline resistance gene. The expressed protein, therefore contained an N-terminal 6His-tag, *attB1* Gateway sequence, TEV protease cleavage site and human TTK residues 514-804. It is therefore referred to as N6His-GW-TEV-TTK (514-804). This construct is illustrated in figure 4.4.



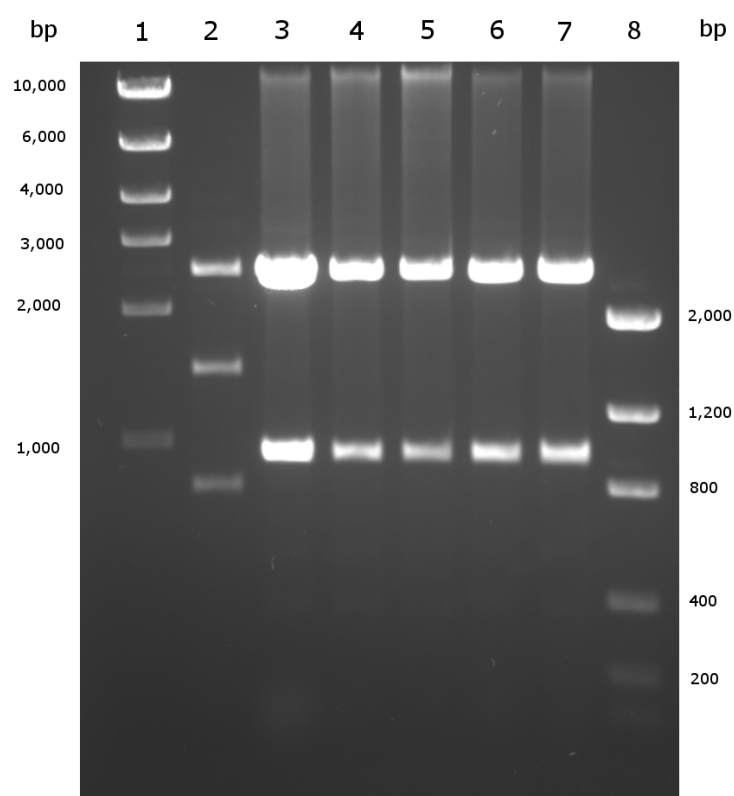
**Figure 4.4:** Diagrammatic representation of the TTK kinase domain expression construct used. Major elements are shown, namely the N-terminal 6 Histidine tag, *attB1* site, TEV protease recognition site, and the TTK kinase domain, residues 514-804 from the full length Human TTK sequence. Amino acid sequences of the *attB1* site and TEV protease recognition site are shown in single letter code. Cleavage point by TEV protease indicated by vertical red line

TTK(514-804) G A N E C I S V K G R I Y S I L K Q I G S G G S S K V F Q V L N E K K Q I Y A I K Y V N L E E A D N Q T L D S Y R N E I A Y L N K L Q Q H S D K  
 TTK(514-804) I I R L Y D Y E I T D Q Y I Y M V M E C G N I D L N S W L K K K S I D P W E R K S Y W K N M L E A V H T I H Q H G I V H S D L K P A N F L I V  
 TTK(514-804) D G M L K L I D F G I A N Q M Q P D T T S V V K D S Q V G T V N Y M P P E A I K D M S S S R E N G K S K S I S P K S D V W S L G C I L Y Y M T  
 TTK(514-804) Y G K T P F Q Q I I N Q I S K L H A I I D P N H E I E F P D I P E K D L Q D V L K C C L K R D P K Q R I S I P E L L A H P Y V Q I Q T H P V N Q  
 TTK(514-804) M A K G

**Figure 4.5:** The amino acid sequence of the cleaved human TTK kinase domain construct used. The sequenced is displayed in single letter amino acid code. The initial residue is not found in human TTK, but remains from the TEV protease recognition site.

#### 4.3.2 Creation of TTK Tryptophan Mutants using Site-Directed Mutagenesis

The supplied expression construct was unsuitable for site-directed mutagenesis, due to the size of the vector, and RNA secondary structure in the vector. To create a template suitable for site-directed mutagenesis, the design of the Gateway cloning system was exploited. A BP reaction (section 3.2.2.1) was performed, using the pT7#3.3 TTK plasmid as the source vector, and pDONR221 as the destination vector. This created a plasmid which was suitable for modification by site-directed mutagenesis. A BSRG1 digest was performed on the plasmid preparations produced to check for the incorporation of the coding insert into the destination vector (Figure 4.5).



**Figure 4.6:** 1% Agarose gel showing BSRG1 digest of TTK containing pDONOR221 plasmids produced by BP reaction. Lane 1 – High Mass DNA ladder (Invitrogen), Lane 2 – unreacted pDONOR221 digested with BSRG1, Lanes 3-7 – Reacted pDONOR221 minipreps showing the incorporation of TTK containing insert by band shift, Lane 8 – Low Mass DNA ladder (Invitrogen).

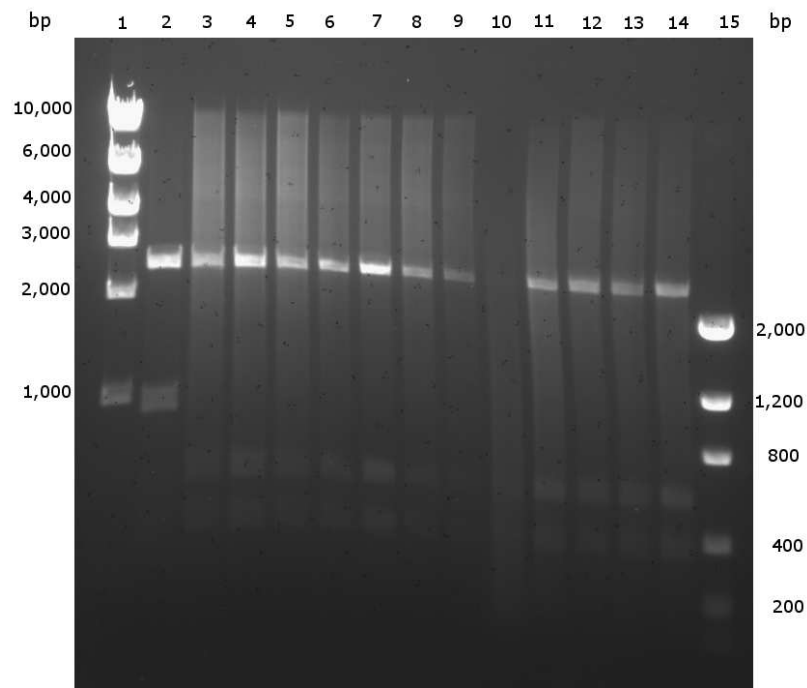
Wild type TTK contains 4 tryptophan residues, W612, W622, W628 and W718. These residues are all located in the C-terminal lobe of the TTK kinase domain.(Figure 4.1) A series of tryptophan mutants were created by exchanging tryptophan residues in the TTK kinase domain for phenylalanine, which was expected to result in a minimal effect of the structure of the TTK kinase domain. Table 4.1 describes the mutants that were created.

**Table 4.1:** Summary of the TTK tryptophan mutants created shown the tryptophan to phenylalanine mutations and the remaining tryptophans in the sequence.

Mutant	Tryptophan to Phenylalanine mutations	Remaining Tryptophans
TTK <sup>WT</sup>	None	W612 / W622 / W628 / W718
TTK <sup>W612F W622F</sup>	W612F W622F	W628 / W718
TTK <sup>W612F W622F W628F</sup>	W612F / W622F / W628F	W718
TTK <sup>W718F</sup>	W718F	W612 / W622 / W628

To create TTK tryptophan mutants, forward and reverse oligonucleotide primers were designed using the online tools provided by Stratagene, and synthesized by Eurogentec. These primers incorporated the codon for phenylalanine (TTC) in place of the codon for tryptophan (TGG) at the required places. The oligonucleotide primers that were designed to create a W628F mutation overlapped with the W622 site, and so contained the W622F mutation as well, since a construct containing a W628F mutation, but not a W622F mutation was not envisaged, and it was important to not reverse the earlier mutation through poor primer design.

The oligonucleotide primers for the W612F mutation and the W622F mutation introduced new restriction sites into the plasmid through the nucleotide changes involved. An XmnI site was introduced with the W612F mutation and a BstBI site was introduced with the W622F mutation. In each case a double digest using BSRG1 and the restriction enzyme for the mutation performed was used to demonstrate the successful incorporation of the new restriction site, and therefore the tryptophan to phenylalanine mutation into the sequence (Figure 4.6). The TTK<sup>WT</sup> template was used for the TTK<sup>W612F W622F</sup> and TTK<sup>W718F</sup> mutants, whilst the TTK<sup>W612F W622</sup> construct as the template for the TTK<sup>W612F W622F W628F</sup> mutant.



**Figure 4.7** A 1% agarose gel showing the double digestion of putative TTK<sup>W612F</sup><sub>W622F</sub> with BstB1 and BSRG1. Lane 1 – High MW DNA ladder, Lane 2 – Control Plasmid (pDONOR221 with TTK<sup>W612F</sup>) showing two bands expected when W622F mutation is absent, Lanes 3-14 – TTK<sup>W612F</sup><sub>W622F</sub> putative plasmids showing 3 bands indicating incorporation of the W622F mutation except in lane 10, Lane 15 – Low Mass DNA ladder.

The tryptophan mutations describe above were incorporated into the sequence using the QuickChange site-directed mutagenesis kit according to the manufacturer's instructions (method outlined in section 4.2.2). Once the mutagenesis had been performed, and the incorporation of the mutation confirmed, an expression vector was created for the mutant by recombining the pDONOR221 entry vector containing the mutated TTK sequence and a pT7#3.3 N6His vector by the LR reaction (Section 3.2.2.2). The expression vector was checked for the incorporation of the insert by a BSRG1 digest (data not shown).

#### ***4.3.3 Expression of Wild Type TTK and TTK Tryptophan Mutants***

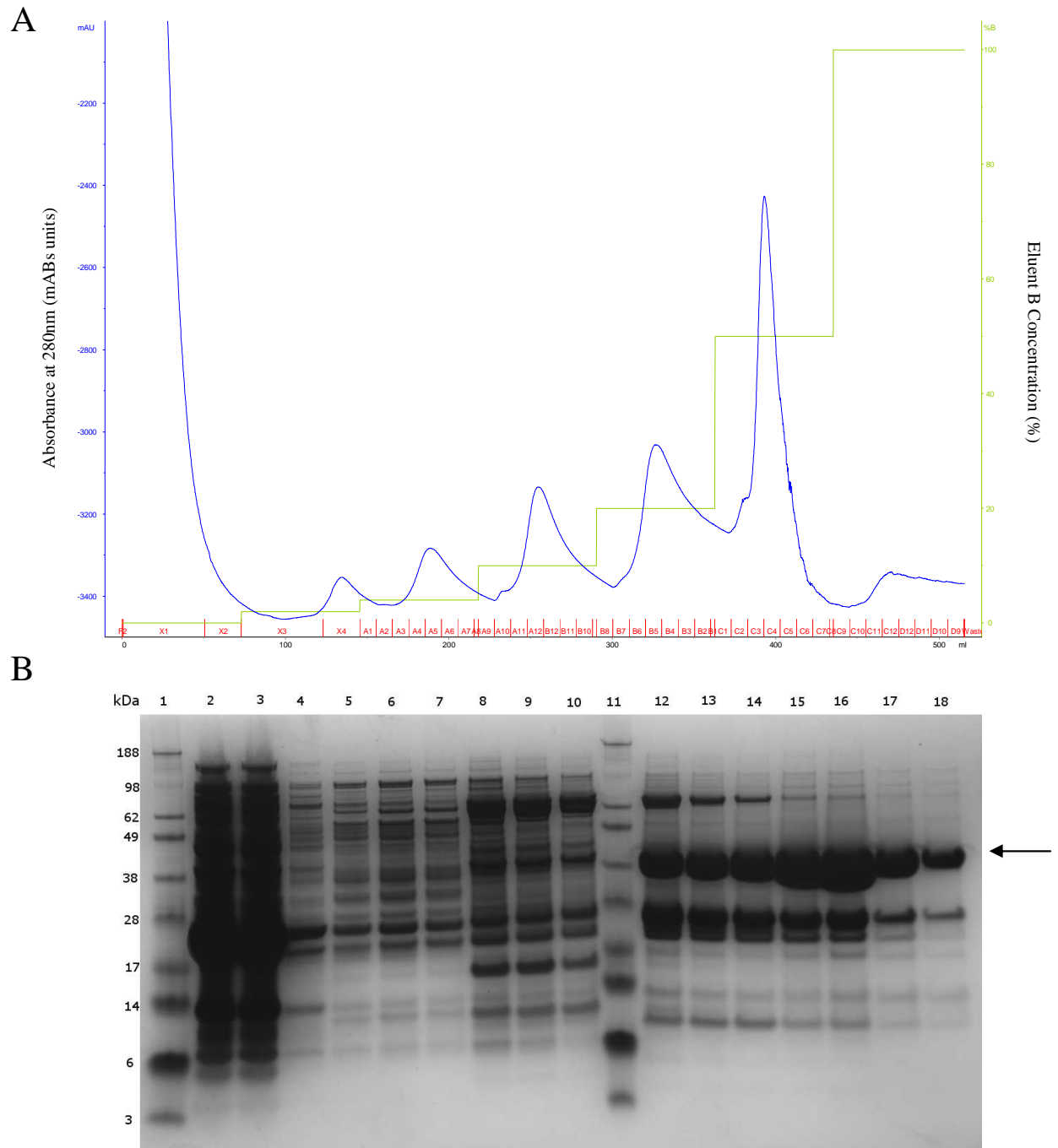
*E. coli* BL21 (DE3) cells that had been pre-transformed with a plasmid coding for λ-Phosphatase, and subsequently made competent were transformed with TTK<sup>WT</sup> or TTK mutant plasmids (Table 4.1) that were created as described in section 4.3.4. These transformed cells were cultured and the TTK constructs expressed as described in section 4.2.3.



#### ***4.3.4 Purification of Wild Type TTK and TTK Tryptophan Mutants***

Wild Type TTK kinase domain and the tryptophan mutants described in table 4.1 were purified in the manner described in section 4.2.4. Using this purification scheme it was possible to purify TTK to a high purity suitable for folding studies. The W718F mutant of TTK was not able to be purified, since the protein accumulated as inclusion bodies on expression.

TTK was initially purified by immobilised metal affinity chromatography (IMAC) using NiNTA superflow resin. TTK eluted at imidazole concentrations of 100 mM and 250 mM (Figure 4.8). Fractions containing the protein of interest were identified using SDS-PAGE (section 2.2.7).

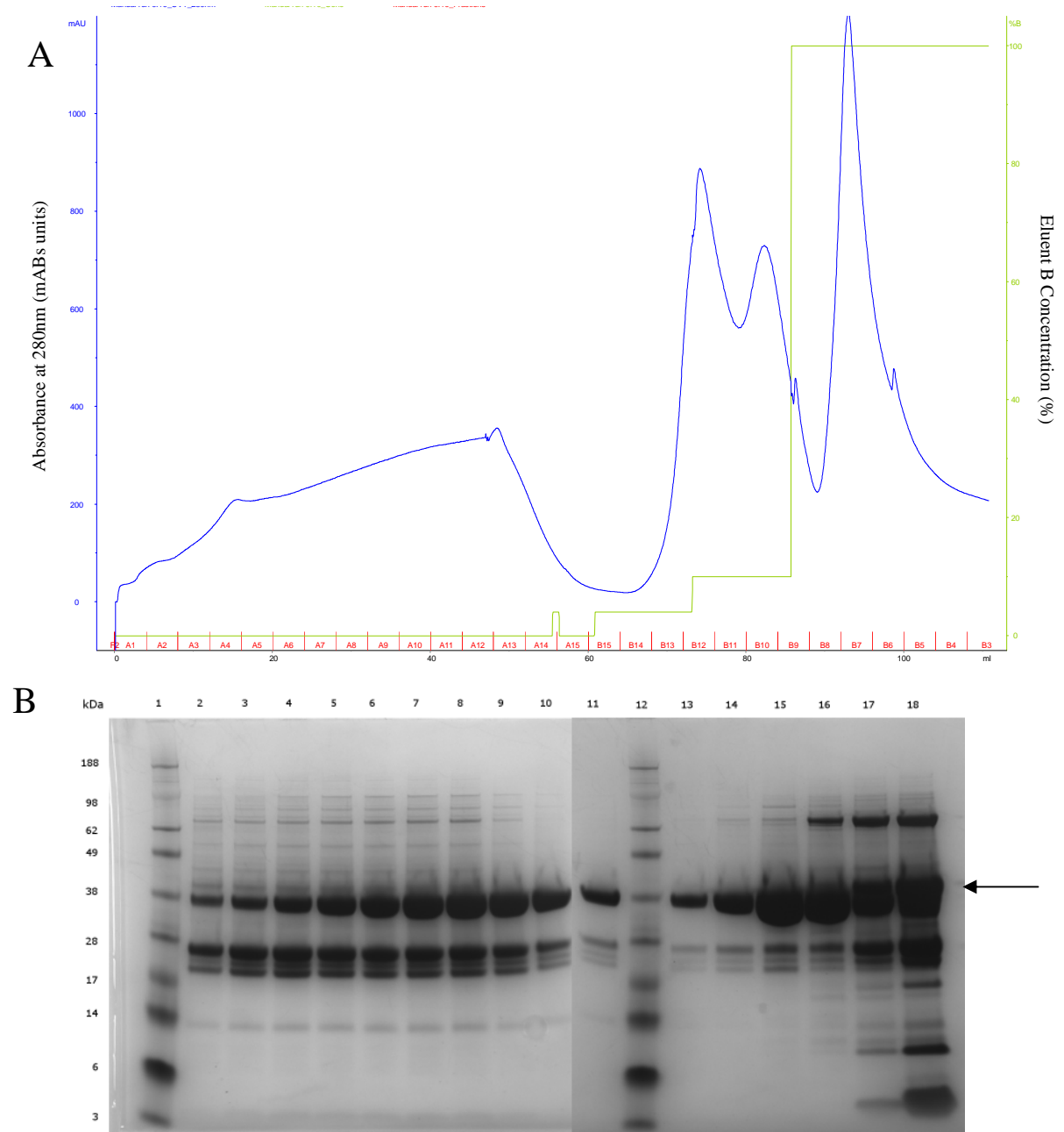


**Figure 4.8:** Purification of TTK<sup>WT</sup> by Immobilised Metal Affinity Chromatography. (A) – chromatogram of the elution of TTK from Ni-NTA superflow column in increasing concentrations of Imidazole. (B) SDS-PAGE gel of selected fractions from the elution of TTK. Lane 1 – MW markers (See-Blue Plus 2, Invitrogen); Lane 2 – Crude extract; Lane 3 – Flow through; Lane 4 and 5 – 10 mM Imidazole wash; lane 6 and 7 – 20 mM Imidazole wash; lane 8 to 10 – 50 mM Imidazole wash; Lane 11 – MW markers; Lane 12 to 14 – 100 mM Imidazole wash; Lane 15 to 16 – 250 mM Imidazole wash; Lane 17 and 18 – 500 mM Imidazole wash. Position of the TTK band is indicated by arrow.

Fractions containing TTK as identified by SDS-PAGE were pooled and a rough estimate of the protein concentration made by assuming that  $1 A_{280} = 1$  mg/mL. This estimation was used to calculate the amount of TEV protease which would be required to complete the cleavage of the eluted protein, when incubated overnight at 4 °C.

The cleavage of the 6His-tag, Gateway sequence and the majority of the TEV protease site was combined with a dialysis of the pooled fractions into a low salt buffer without imidazole allowing the subsequent purification steps of a subtractive IMAC step, and cation exchange chromatography to occur without further desalting or dialysis.

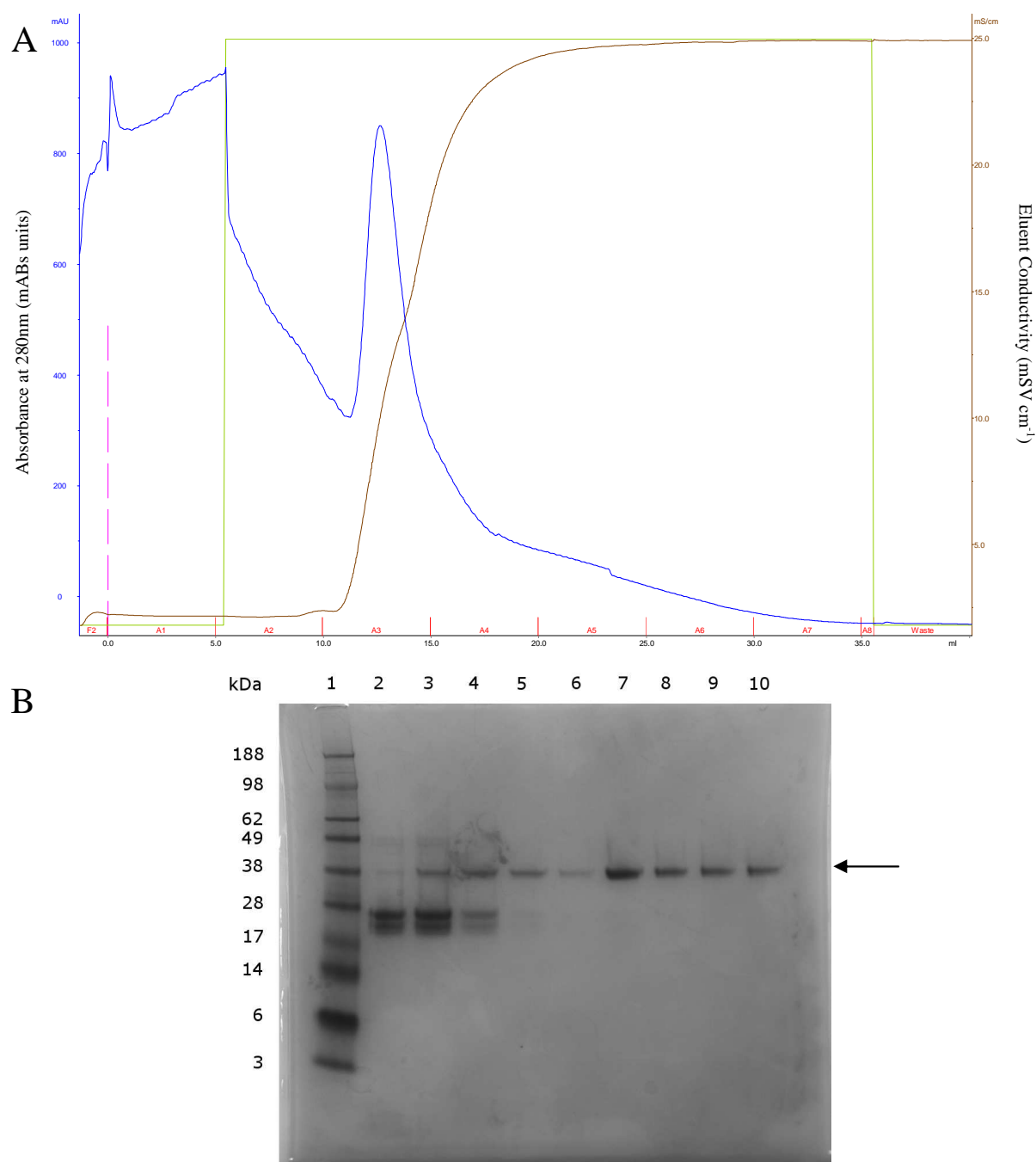
The dialysed fractions were applied to a NiNTA superflow column and separated by binding to immobilised Ni ions. Cleaved TTK eluted from the column in the flow-through (Figure 4.9), and cleaved tag, uncleaved TTK and TEV protease eluted in high imidazole concentrations. The fractions that contained cleaved TTK were identified by SDS-PAGE and pooled.



**Figure 4.9:** Purification of TTK by subtractive immobilised metal affinity Chromatography. (A) – Chromatogram of the elution of cleaved and uncleaved TTK, TEV protease and cleaved tag from Ni-NTA superflow column (5 mL). (B) – SDS-PAGE gel of selected fractions from the elution of TTK. Lane 1 – MW markers (See-Blue Plus 2, Invitrogen); Lane 2 to 11 – flow through showing unbound, cleaved TTK; Lane 12 - MW markers; Lane 13 and 14 – 20 mM imidazole wash; Lane 14 and 15 – 50 mM imidazole wash; lane 16 and 17 – 400 mM imidazole wash. Arrow indicates position of TTK band.

The final purification step purified TTK by cation exchange chromatography using a Source S column. TTK eluted from the column in the presence of low concentrations of NaCl (Figure 4.10). Significant amounts of TTK did not bind to

the column, so the flow-through was re-applied to the column to maximise the yield of purified TTK.

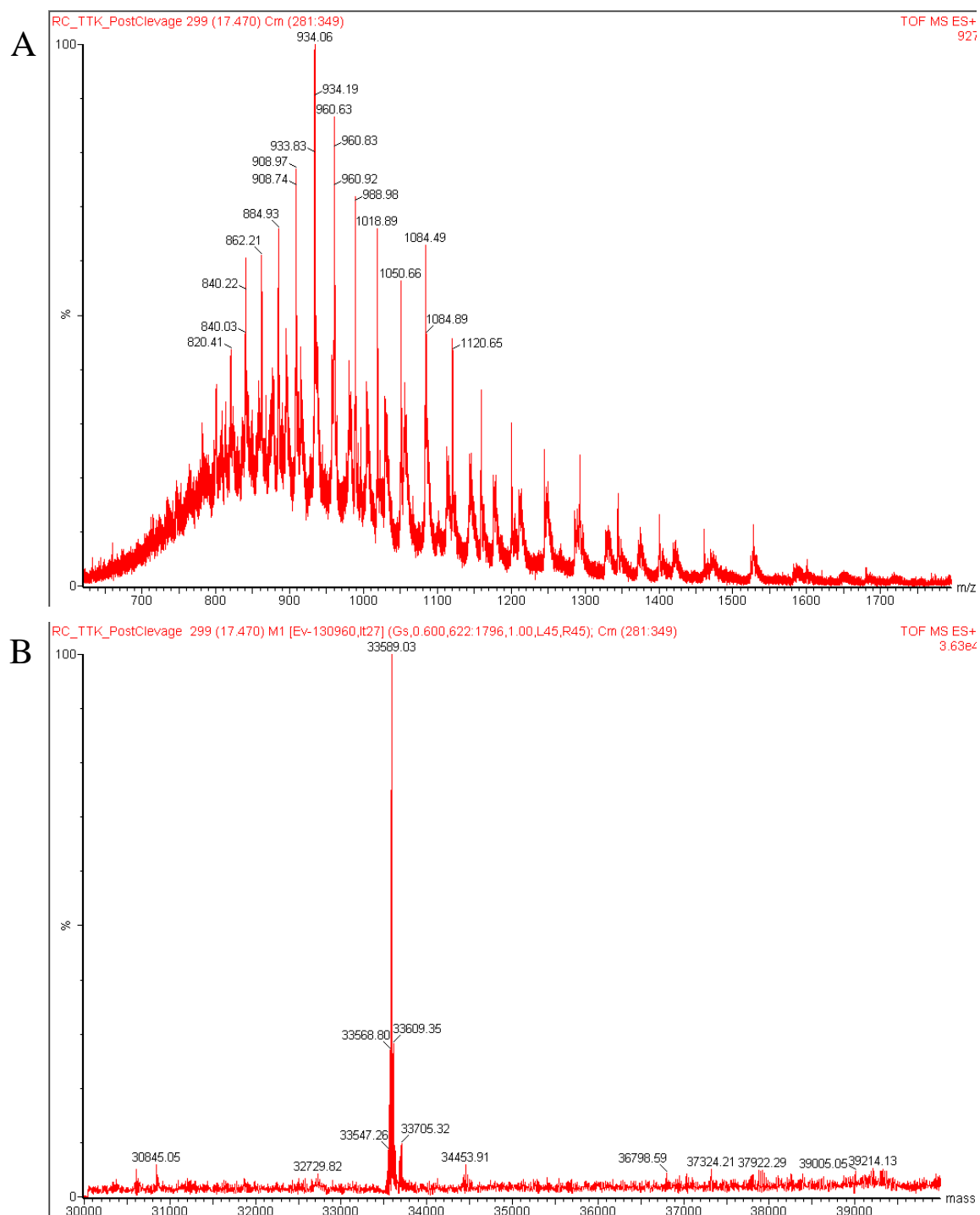


**Figure 4.10:** Purification of TTK by cation exchange chromatography. Panel A – Elution profile of TTK from Source S column in 20 mM Tris 5 mM DTT pH 7.4. Panel B – SDS-PAGE analysis of fractions from elution profile shown in Panel A. Lane 1 MW markers (SeeBlue Plus 2, Invitrogen), Lane 2-5: Flow Through Fractions, Lanes 6-10: Fractions from NaCl gradient elution, showing pure TTK eluting from the column. Arrow indicates TTK band.

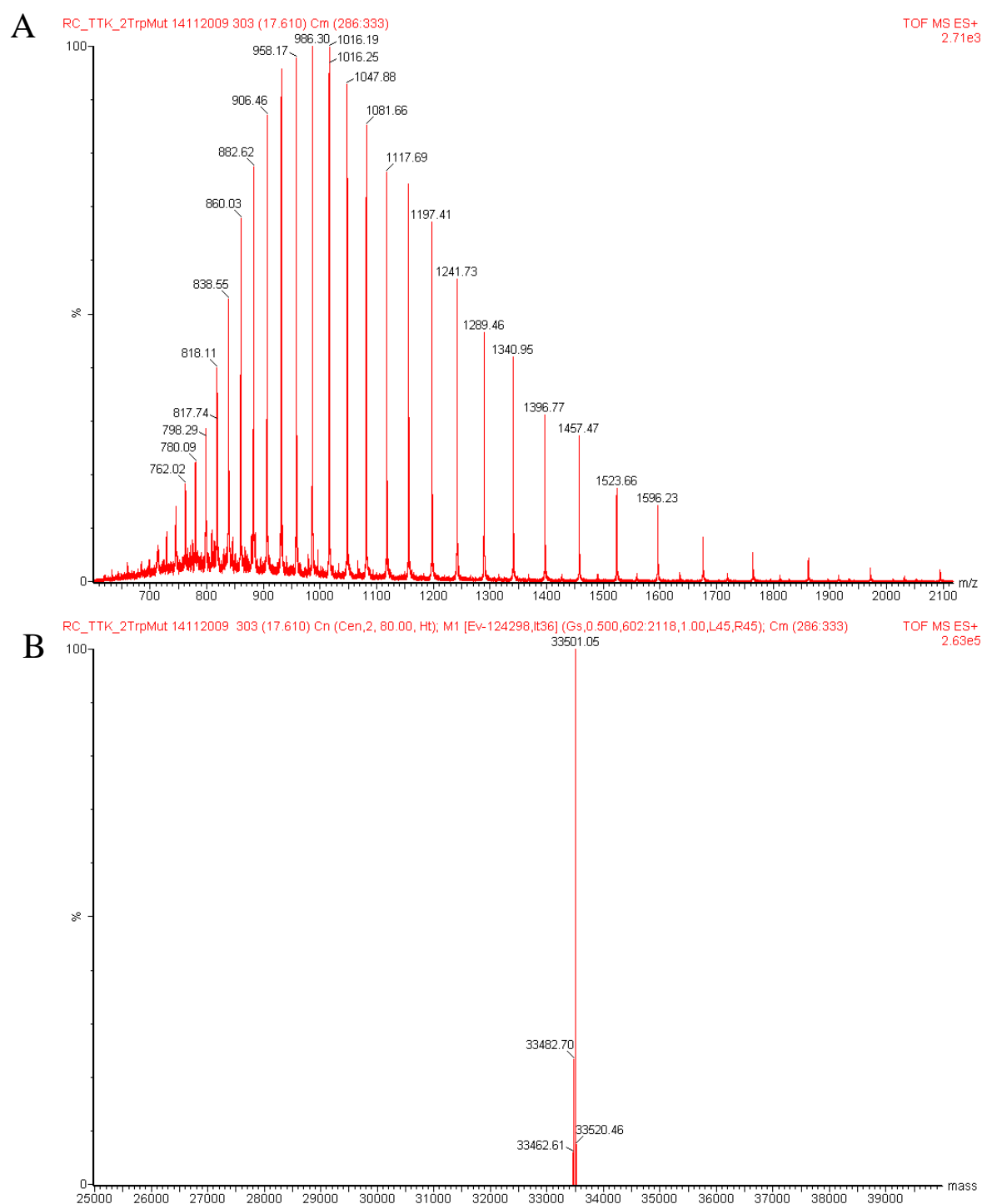
Fractions identified as containing TTK were identified by SDS-PAGE and pooled. The pooled fractions were concentrated using centrifugal concentrator device with a 10 kDa cut-off to a concentration of 3-5 mg/mL. Purified TTK was stored in buffer at 4 °C and used immediately.

#### ***4.3.5 Mass Spectrometry of TTK and TTK Tryptophan Mutants***

ESI-MS was used to measure the molecular mass of purified TTK and TTK tryptophan mutants. The measurement was performed after the 6His tag had been cleaved. The wild type TTK construct had a measured mass of 33589.03 Da (Figure 4.11). This is in agreement with the calculated mass of 33589.9 Da. The molecular masses of the TTK tryptophan mutants (Figure 4.12 and 4.13) were also in agreement with the calculated masses for those constructs (Table 4.2).

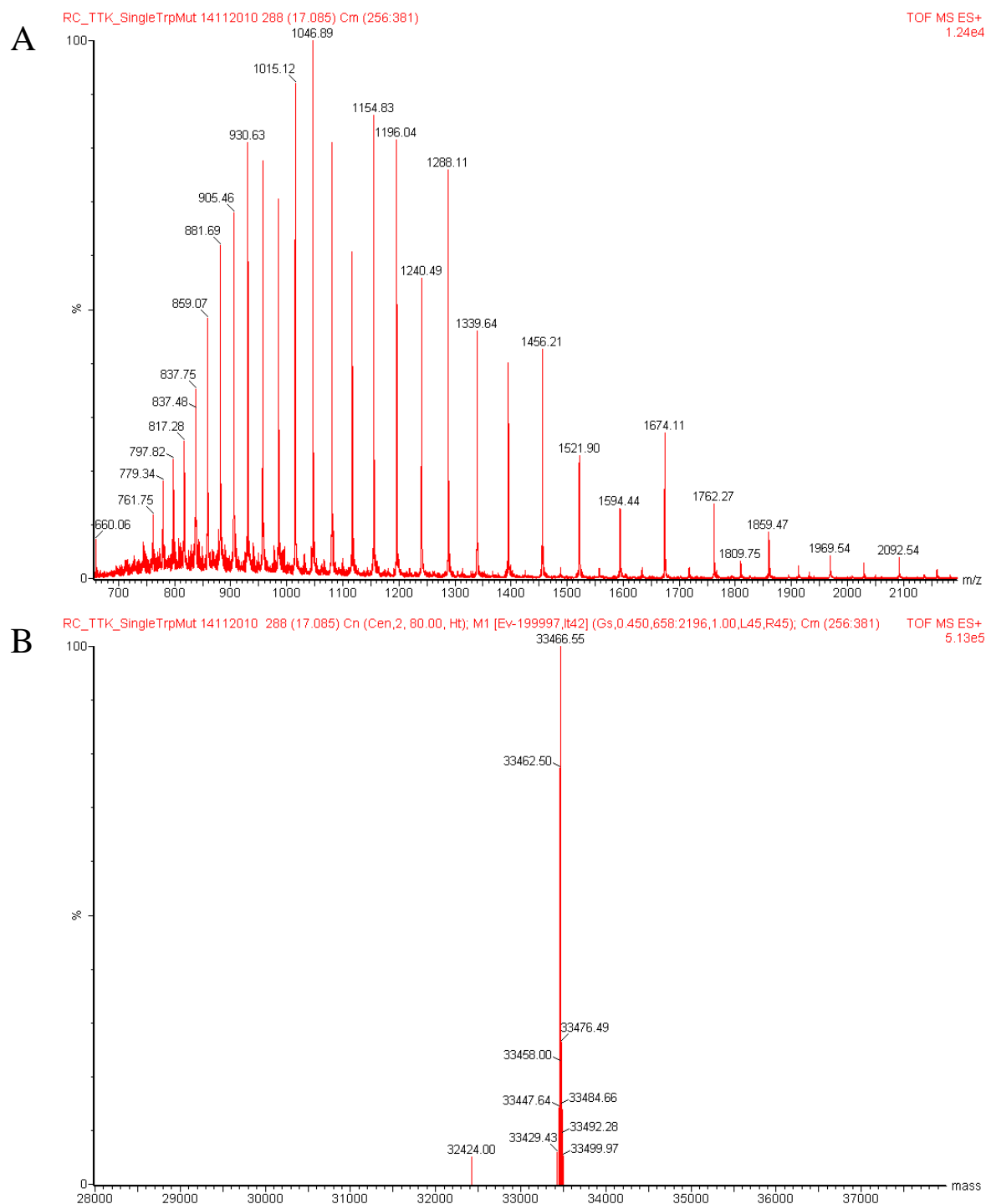


**Figure 4.11:** ESI-TOF mass spectrometry of wild type TTK. (A) – mass spectrum of wild type TTK, (B) maximum entropy deconvolution of mass spectrum in panel (A), showing the measured molecular mass of 33589.03 Da.



**Figure 4.12:** ESI-TOF mass spectrometry of TTK<sup>W612F W622F</sup>. (A) – mass spectrum of TTK<sup>W612F W622F</sup>, (B) maximum entropy deconvolution of mass spectrum in panel (A), showing the measured molecular mass of 33501.05 Da.





**Figure 4.13:** ESI-TOF mass spectrometry of TTK<sup>W612F W622F W628F</sup>. (A) – mass spectrum of TTK<sup>W612F W622F W628F</sup>, (B) maximum entropy deconvolution of mass spectrum in panel (A), showing the measured molecular mass of 33466.55 Da.

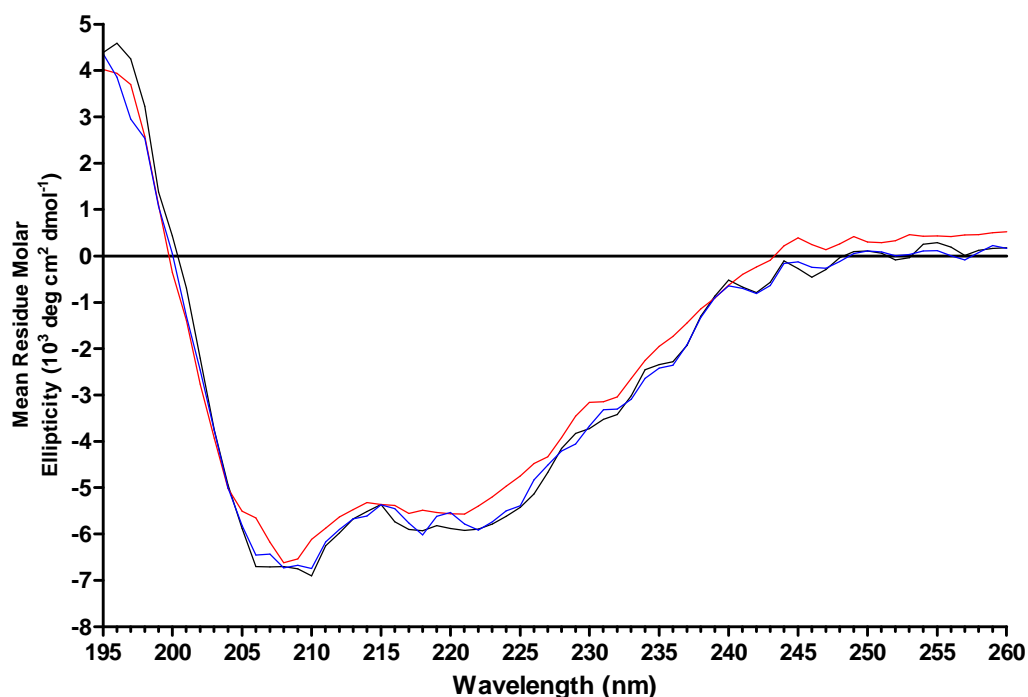
**Table 4.2:** Molecular mass of TTK and TTK tryptophan mutants as measured by ESI-MS and calculated from the sequence.

Protein Construct	Measured Molecular Mass	Calculated Molecular Mass
Wild Type TTK	33,589.03±0.34	33,589.9
TTK <sup>W612F W622F</sup>	33,501.05±0.34	33,511.8
TTK <sup>W612F W622F W628F</sup>	33,466.55±0.33	33,472.8

The expression and purification of wild type TTK resulted in a yield of purification of 4.2 mg of protein per litre of expression culture. This fairly low yield of purification was measured after a single round of cation exchange chromatography, in which TTK bound to the column with low affinity, requiring multiple rounds of cation exchange chromatography through reapplication of the flow through to the column to increase the yield of purified TTK. The yields of purification at this stage were 4.0 mg/L and 3.9 mg/L for TTK<sup>W612F W622F</sup> and TTK<sup>W612F W622F W628F</sup> respectively. The estimations of protein content of the pooled fractions after IMAC chromatography were also similar, approximately 260 to 280 mg of partially purified protein from 7.2 L of expression culture.

#### ***4.3.6 Far-UV CD Spectra of TTK and TTK Tryptophan Mutants***

The far-UV CD spectra of wild type TTK and TTK tryptophan mutants were measured and compared to identify if there are any major changes in the CD spectra observed on the replacement of tryptophan residues with phenylalanine. The wild type CD spectra were measured between 195 nm and 260 nm. The spectra show minima at 208 nm and 222 nm (Figure 4.14). These minima are indicative of a protein with a high alpha helical content.



**Figure 4.14:** CD spectra of TTK and tryptophan mutants. Black line – wild type TTK, Red line – TTK<sup>W612F W622F</sup>, Blue line – TTK<sup>W612F W622F W628F</sup>. Far-UV CD spectra were gathered using a 1mm pathlength quartz cuvette at 20 °C and a protein concentration of 0.1 mg/mL.

The CD spectra of the tryptophan to phenylalanine mutants are in good agreement with the CD spectra of the wild type protein, with minima at similar wavelengths (Table 4.3), and similar calculated secondary structure content. On this basis it appears that the replacement of tryptophan with phenylalanine does not cause substantial structural perturbation.

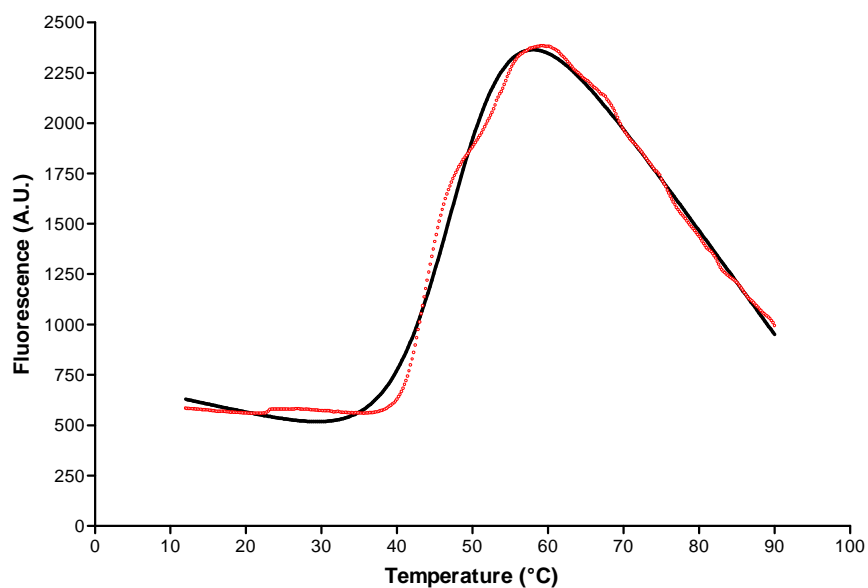
**Table 4.3:** Secondary structure content predictions for TTK and TTK tryptophan mutants calculated by K2D (Andrade *et al.*, 1993) from circular dichroism spectra. Values in brackets for wild type TTK derived from crystal structure.

Protein	$\alpha$ -Helix	$\beta$ -Sheet	Random coil
Wild Type TTK	27%(32%)	18%(16%)	55%(52%)
TTK <sup>W612F W622F</sup>	30%	15%	55%
TTK <sup>W612F W622F W628F</sup>	29%	17%	54%

Slight differences in the calculated secondary structure content of the crystal structure and the secondary structure content of the expressed protein, since the construct which is resolved in the crystal structure is shorter than the expressed construct and a number of regions were not resolved in the crystal structure. It is likely that these regions are flexible, and would appear as a random coil conformation in CD experiments. It would therefore be expected that the random coil percentage of the expressed construct would be higher than that of the crystal structure, and the  $\alpha$ -helix and  $\beta$ -sheet component correspondingly lower. This is indeed what is observed (Table 4.3).

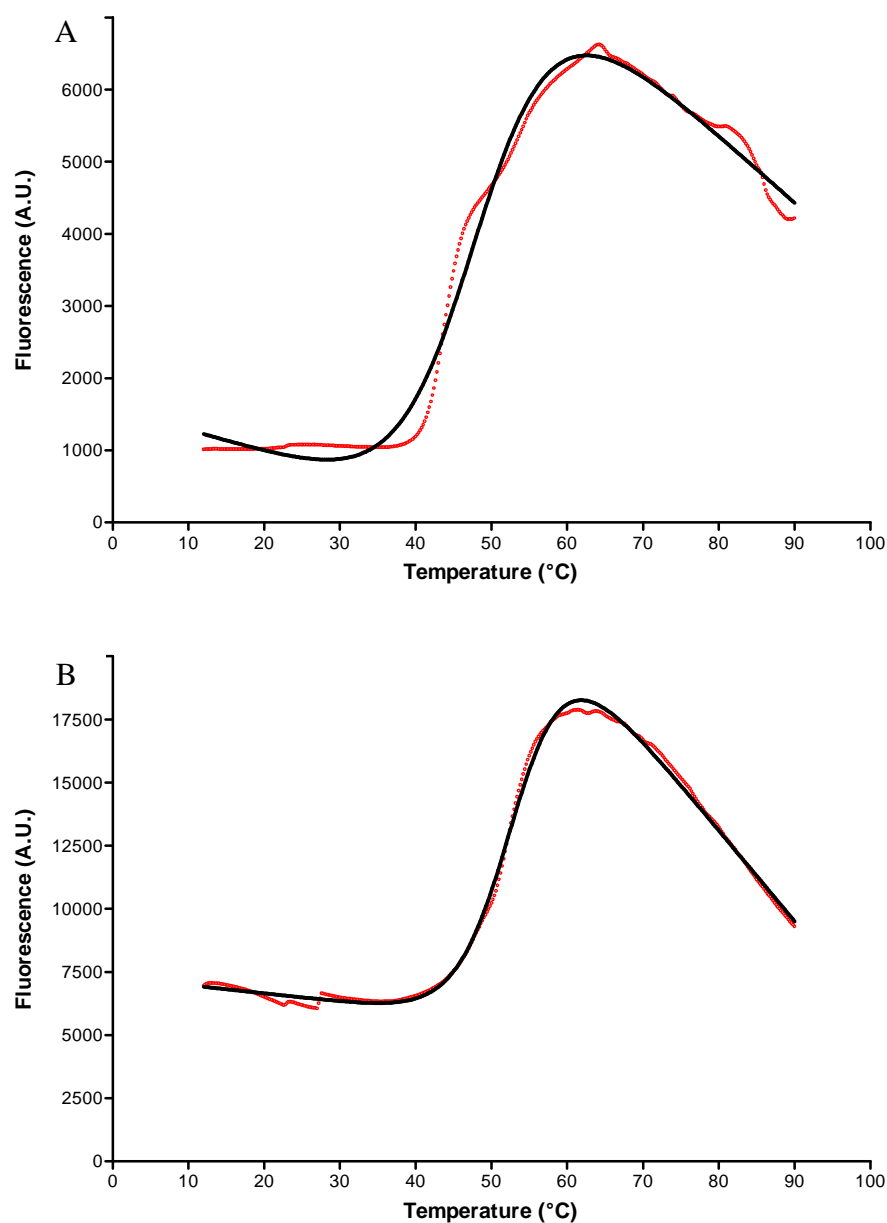
#### ***4.3.7 Thermal Melting Analysis of TTK and TTK Tryptophan Mutants***

The tryptophan to phenylalanine substitutions that were carried out on TTK may have a deleterious effect on the fold or the stability of the kinase domain. To check for this, the thermal melting curves of TTK and its tryptophan mutants were determined using the fluorescent dye, Sypro-Orange, which binds to hydrophobic areas of the protein which are exposed as the protein unfolds, and is quenched by water. Thermal melting curves of TTK and the tryptophan mutants described in section 4.3.2 were obtained using the procedure outlined in section 2.2.18. The thermal melting curves were fitted to equation 2.1 and the thermodynamic parameters calculated. The melting curves have a typical shape, with fluorescence minima and maxima defining a melting transition. Figure 4.15 shows a typical melting curve for wild type TTK. Samples are cooled to 12 °C before the experiment begins. The temperature is then raised to 90 °C in 0.1 °C steps. It can be seen that there is an initial baseline between 12 °C and 40 °C. The unfolding transition then occurs between 40 °C and 60 °C. This is followed by a decrease in fluorescence intensity between 60 °C and 90 °C.

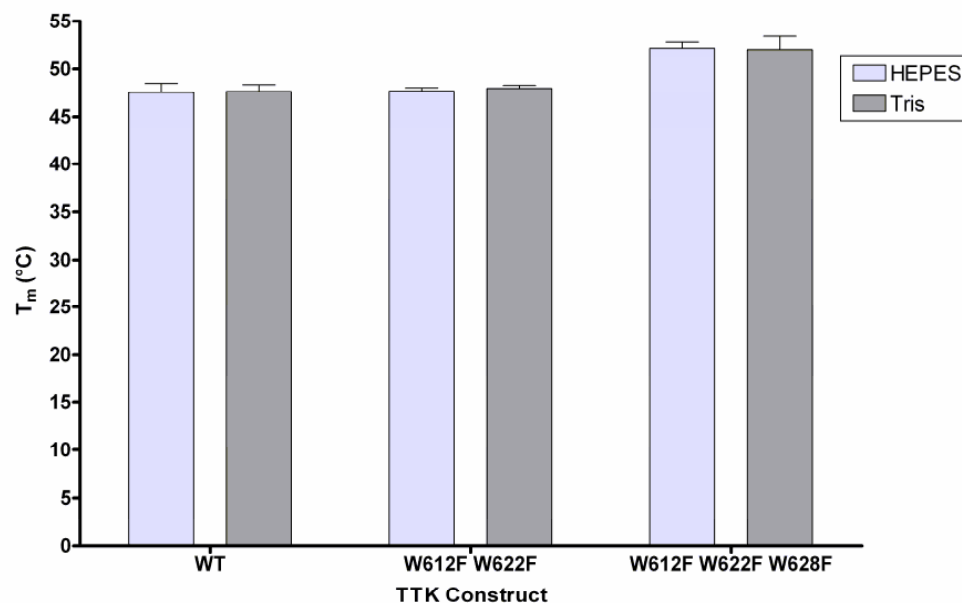


**Figure 4.15:** Thermal unfolding curve of TTK monitored by Sypro-Orange fluorescence. Protein concentration was 0.2 mg/mL, the recorded fluorescence data is shown as red circles and the black line represents the fitting of the data to equation 2.1. The three phases observed are an initial baseline between 12 and 40 °C, an unfolding transition between 40 °C and 60 °C, and a decrease in recorded fluorescence between 60 °C and 90 °C. Data from a single unfolding experiment.

The measured  $T_m$  value for wild type TTK is  $47.5 \pm 0.97$  °C and the change in enthalpy at the transition temperature ( $\Delta H_m$ ) is  $46.1 \pm 1.03$  kcal mol<sup>-1</sup>. These parameters were also measured for the tryptophan mutants of TTK. A summary of these results are shown in figure 4.16, figure 4.17 and table 4.4.



**Figure 4.16:** Thermal unfolding curves of TTK tryptophan mutants. (A) TTK<sup>W612F W622F</sup>, (B) TTK<sup>W612F W622F W628F</sup>. Red circles fluorescence data, black line represents fitting of equation 2.1 to fluorescence data. Protein concentration was 0.2 mg/mL. Curves shown are from a single refolding experiment.



**Figure 4.17:** Average  $T_m$  for wild type TTK and TTK tryptophan mutants. Thermal melting analysis performed at 0.2 mg/mL protein in 10 mM HEPES pH 7.4 and in 50 mM Tris pH 7.5. Results are mean of 5 experiments, bars show standard error.

The thermal melting analysis of the TTK tryptophan mutants indicate that the fold of the tryptophan mutants has not been destabilised by the substitution of tryptophan residues with phenylalanine. There has been an increase in the stability of the W612F W622F W628F mutant over the wild type of  $\sim 5^\circ\text{C}$  (Table 4.4).

**Table 4.4:** Mid points of thermal melting transitions for TTK kinase domain and tryptophan mutants. Data shown in 10 mM HEPES pH7.4.

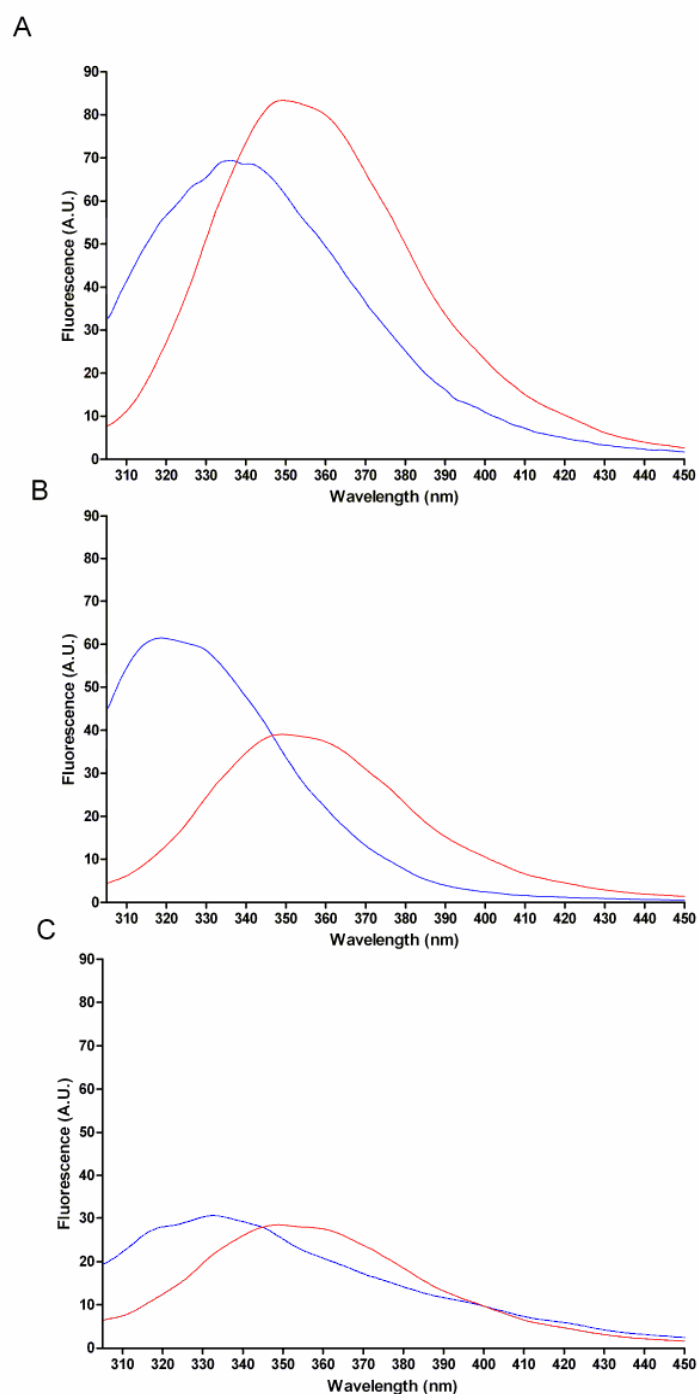
TTK Construct	$T_m$ ( $^\circ\text{C}$ )	$\Delta H_m$ ( $\text{kcal mol}^{-1}$ )
WT	$47.5 \pm 0.97$	$46.1 \pm 1.03$
W612F W622F	$47.6 \pm 0.35$	$42.9 \pm 1.30$
W612F W622F W628F	$52.2 \pm 0.64$	$44.3 \pm 0.88$

The changes in enthalpy of the unfolding transitions are consistent across the tryptophan mutants, suggesting that no major change in the stability of the fold has occurred.

#### ***4.3.8 Tryptophan Fluorescence Spectra of TTK and TTK Tryptophan Mutants***

At an excitation wavelength of 295 nm the fluorescence emission spectrum of TTK is dominated by tryptophan residues. The contribution from other fluorescent residues, phenylalanine and tyrosine, is minimised. Wild Type, native TTK has a maximum fluorescence intensity ( $\lambda_{\text{max}}$ ) of 336.5 nm (Figure 4.18A). When the wild type protein is unfolded in 6 M GdnHCl, a red shift in the  $\lambda_{\text{max}}$  to 349 nm and an increase in the fluorescence intensity is observed. This is consistent with buried tryptophan residues being exposed to solvent, and quenching of tryptophan fluorescence being released upon unfolding.





**Figure 4.18:** Tryptophan emission spectra of TTK and TTK tryptophan mutants. (A) – Wild Type TTK, (B) – TTK<sup>W612F W622F</sup>, (C) – TTK<sup>W612F W622F W628F</sup>. Native spectra shown in blue, Unfolded spectra (6 M GdnHCl) in red. Protein concentration 0.1 mg/mL.

The tryptophan to phenylalanine substitutions performed would allow the folding of the C-terminal lobe of the TTK kinase domain to be followed more

specifically, first by eliminating the solvent exposed tryptophan residues, and then by studying the core kinase domain tryptophan W718.

The W612F W622F mutant of TTK shows a slight decrease in the fluorescence intensity of the native state, indicating that the mutated residues were strongly quenched in the wild type native state (Figure 4.18B). There is also a large blue shift in the  $\lambda_{\text{max}}$  of the native state to 319 nm. When this mutant is unfolded with 6 M GdnHCl there is a red shift of the  $\lambda_{\text{max}}$  to 349 nm and a decrease in the observed fluorescence intensity. The fluorescence intensity of the denatured state of the W612F W622F mutant is approximately half that of the wild type denatured state, which is consistent with the replacement of two tryptophan residues.

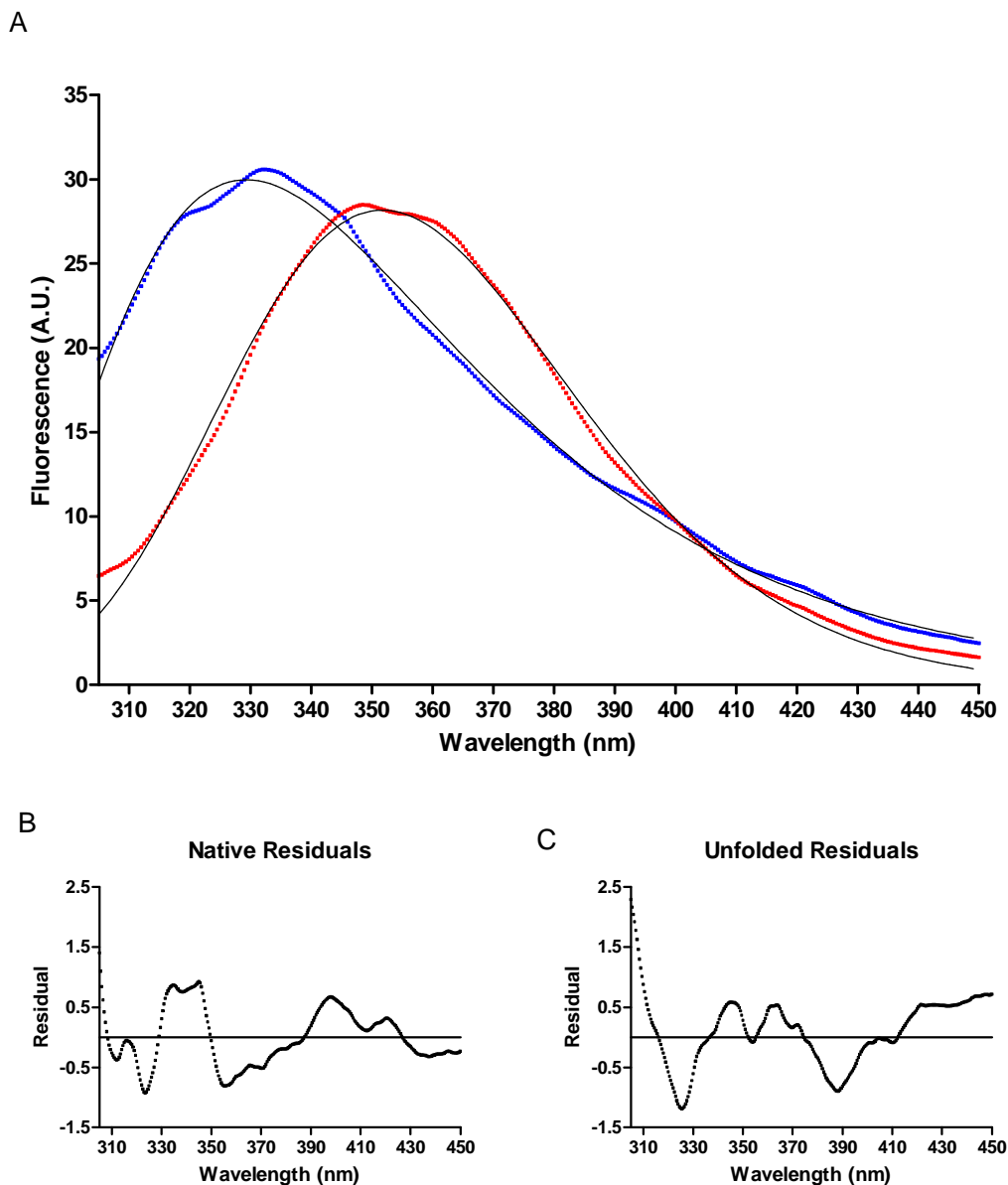
The W612F W622F W628F mutant was created to allow the folding of the conserved core tryptophan, W718, to be studied, and compared to W207 in p38 $\alpha$ . The fluorescence spectra of the native state of TTK<sup>W612F W622F W628F</sup> shows a decrease of approximately 50 % in the fluorescence intensity when compared to the W612F W622F mutant (Figure 4.18C). There is also a red shift in the  $\lambda_{\text{max}}$  from the W612F W622F mutant to a  $\lambda_{\text{max}}$  of 332.5 nm. Upon unfolding, there is a red shift to a  $\lambda_{\text{max}}$  of 348.5 nm and a decrease in fluorescence intensity.

**Table 4.5**  $\lambda_{\text{max}}$  for different TTK constructs under native and denaturing conditions. Protein denatured with 6 M GdnHCL.

TTK Construct	Native $\lambda_{\text{max}}$	Unfolded $\lambda_{\text{max}}$
WT	336.5 nm	349 nm
W612F W622F	319 nm	349 nm
W612F W622F W628F	332.5 nm	348.5 nm

#### 4.3.9 Spectral Broadening of Tryptophan Emission Spectra

Previous studies on the folding of p38 $\alpha$  had shown a broad spectrum for the equivalent single tryptophan mutant, p38 $\alpha$ <sup>W207</sup>, which was indicative of there being two environments in the native state. To determine if a similar arrangement existed for the single tryptophan mutant of TTK, TTK<sup>W612F W622F W628F</sup>, the native tryptophan emission spectra was fitted to a log normal distribution using equation 4.1 (Figure 4.19). The unfolded protein spectra was also fitted to the same distribution. The parameters for the log normal distribution are summarised in table 4.6.

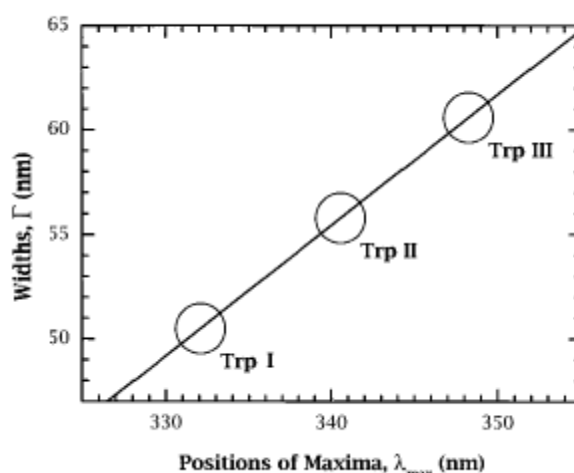


**Figure 4.19:** Fitting of log normal distribution to fluorescence spectra to estimate parameters for curve. (A) Native (blue) and unfolded (red) fluorescence spectra of TTK<sup>W612F W622F W628F</sup>, fitted log normal distribution shown as solid lines. (B) Residuals for fitted curve to native spectra. (C) Residuals for fitted curve to unfolded spectra.

**Table 4.6:** Log normal distribution parameters for the fitting of native and unfolded TTK<sup>W612F W622F W628F</sup> spectra to equation 4.1.

Spectra	$\lambda_{\text{max}}$ (nm)	$\nu_p$ (nm)	$\nu_n$ (nm)
Native	$329.2 \pm 0.1$	$302.1 \pm 0.2$	$377.9 \pm 0.2$
Unfolded	$351.7 \pm 0.1$	$321.5 \pm 0.1$	$389.8 \pm 0.2$

To determine if the spectral width of the spectra of TTK<sup>W612F W622F W628F</sup> indicates that there is heterogeneity in the tryptophan spectra, the calculated values of the  $\lambda_{\text{max}}$  and the spectra width are compared to expected values for tryptophan residues under various degrees of solvent exposure. The plot shown in Ladokhin *et al.* (2000) (Figure 4.20) is used to compare the expected values to the calculated values.



**Figure 4.20:** Tryptophan fluorescence spectrum position-width analysis. Trp I, II and III represent the classes of tryptophan that correlate with the extent of solvent exposure (adapted from Ladokhin *et al.*, 2000).

The spectra width of the unfolded protein was used as a control for the fitting process. The spectra width of unfolded protein was calculated as  $68.3 \pm 0.3$  nm with a  $\lambda_{\text{max}}$  of  $351.7 \pm 0.1$  nm. This spectral width is above the line indicated for a single tryptophan residue; however, the difference between the expected value at 352 nm and the calculated value is 5 nm. This value is small enough that it may be considered that the unfolded spectra consist of a single tryptophan in a single environment.

The spectral width of the native protein was calculated as  $75.8 \pm 0.4$  nm at a  $\lambda_{\text{max}}$  of  $329.2 \pm 0.1$  nm. The position of the native TTK<sup>W612F W622F W628F</sup> indicates that there is substantial spectral broadening of the native tryptophan spectra. This indicates that there is heterogeneity in the tryptophan environments, similar to the tryptophan spectra of the equivalent residue in p38 $\alpha$ .

#### **4.4 Discussion**

The study of protein folding requires the production of soluble protein in mg amounts, at a high purity. In addition, for the study of the folding of a protein by tryptophan fluorescence it is advantageous to be able to produce tryptophan mutants which can clarify the folding of the wild type protein. The presence of multiple tryptophans in a protein makes the interpretation of the folding observed by tryptophan fluorescence more complicated, since the different tryptophan residues are present in different environments, giving rise to different fluorescence signals. The ideal method of study for a protein containing multiple tryptophans would be to study the folding using single tryptophan mutants, where all but a single tryptophan residue have been replaced with phenylalanine or other residues. The use of single tryptophan mutants is necessary since the presence of multiple tryptophans, each in their own environment causes difficulties in the interpretation of the results obtained, and tends to broaden the folding transitions observed, possibly hiding intermediates.

Previous studies on the folding of p38 $\alpha$  (Davies, 2004) had identified a single, core tryptophan, W207, which was absolutely essential for folding, and was found to be conserved throughout the kinase domain (Davies, 2004). To allow a direct comparison with this result, a single tryptophan mutant of TTK was created containing only the homologous residue, W718. This residue was confirmed as essential by creating a mutant in which only the W718 residue had been mutated. It was not possible to produce this mutant as a soluble protein, indicating that like in p38 $\alpha$  this residue is essential for the folding of the kinase domain. An examination of the environment of the two residues shows that not only is the residue conserved in the sequence, but is also structurally conserved, and that its environment is also structurally conserved (Figure 4.3). The requirement of this residue for the correct folding of TTK prevents the creation of other single tryptophan mutants of TTK. Therefore, only one single tryptophan mutant was created and studied.

TTK was purified to a high purity using a 4 stage purification process, following co-expression with  $\lambda$ -phosphatase in *E. coli*, involving an initial IMAC step, cleavage of the 6His tag and other unwanted features by TEV protease, a subtractive IMAC step to remove uncleaved TTK and the protease, and finally an ion

exchange purification. Co-expression with  $\lambda$ -phosphatase was necessary since the expression of TTK alone results in the protein accumulating in inclusion bodies, presumably due to phosphorylation (AstraZeneca unpublished data). This purification strategy resulted in protein of a high purity which was suitable for use in biophysical experiments. ESI-MS was used to confirm the identity of the purified protein and the masses derived were in good agreement with the calculated masses from the sequence. This agreement indicated that the protein was purified in a unphosphorylated form.

Wild type TTK contains four native tryptophans, of which two are solvent exposed in the native state (Figure 4.2). The contribution of these tryptophans to the emission spectrum of the native state was expected to be minimal and their response to the folding and unfolding of the native state low. A mutant was successfully created that replaced these two residues with phenylalanine allowing study of the buried tryptophan residues without a contribution from these exposed residues. The tryptophan emission spectra of the native state was affected by these substitutions with the native  $\lambda_{\text{max}}$  decreasing (Figure 4.18). This indicates that the solvent exposed residues had a significant contribution to the fluorescence spectrum of the native state. The comparison of the spectra of the wild type and the W612F W622F mutants of TTK in the presence of 6M GdnHCl showed a two fold drop in the fluorescence intensity between the two proteins (Figure 4.18) and a  $\lambda_{\text{max}}$  of 349nm (Table 4.5), indicating that both protein were fully unfolded under these conditions, and that the number of tryptophan residues in the protein was reduced by half.

The far-UV CD spectra of the wild type TTK and tryptophan mutants were compared to identify if the tryptophan to phenylalanine substitutions carried out had affected the secondary structure content of the protein. The CD spectra of the proteins were similar, and the calculated secondary structure content of the proteins were also similar to each other and to the secondary structure content calculated from the crystal structure (Table 4.3). In addition, the thermal melting of the wild type protein and tryptophan mutants was assessed using the Sypro-orange dye binding method. The thermal melting temperature was not decreased by the tryptophan to phenylalanine substitutions performed, although the stability of the W612F W622F W628F mutant was higher than the wild type and W612F W622F mutant (Table 4.4). These results indicate that the mutations performed have not grossly affected the structure, although for a complete demonstration of this it would

be necessary to crystallise the mutants and solve their structures via X-ray crystallography.

Previous studies on p38 $\alpha$  had shown spectral broadening of the tryptophan emission spectra of the native single tryptophan mutant. This was interpreted to indicate that two conformational states were being populated (Davies, 2004). The spectra of the single tryptophan mutant of TTK shows significant spectral broadening (Figure 4.19, Table 4.6). This indicates that there are multiple environments for this tryptophan residue. An examination of the B-factors associated with the crystal structure does not identify W718 or its environment as being a particularly mobile element in the context of the TTK crystal structure. The heterogeneity of the environment of tryptophan 718 is more likely to be generated by larger scale motions in the kinase domain, possibly associated with substrate or ATP binding. Studies performed on cyclophilin A by Eisenmesser *et al.* (2005) have shown that the binding of substrates can be facilitated by a transient conformational shift from an inactive form of a protein to the substrate bound structure that occurs before substrate binding and is required for substrate binding. It can be proposed that such an effect is responsible for the heterogeneity found in the single tryptophan spectra of TTK<sup>W612F W622F W628F</sup> and p38 $\alpha$ .

The characterisation performed on the wild type TTK and tryptophan mutants indicates that they have been produced as soluble, correctly folded protein suitable for use in examining the equilibrium folding of TTK. The tryptophan to phenylalanine substitutions performed have been indicated to have minimal effect on the structure and stability of the TTK fold and the yields of expression and purification.

## **Chapter 5. Equilibrium Folding of Human TTK Protein Kinase Domain**

### ***5.1 Introduction***

This chapter investigates the equilibrium folding of TTK. Using site directed mutagenesis tryptophan to phenylalanine substitutions (Chapter 4) were made to allow the study of the TTK kinase fold and to allow a direct comparison to be made to the folding studies performed on p38 $\alpha$  by Davies (2004).

The TTK kinase domain contains four native tryptophan residues, W612, W622, W628 and W718 (Figure 4.1). These four residues are all contained within the C-terminal lobe of the kinase domain, and are found in different environments. For example, the crystal structure of the TTK kinase domain indicates that the residues 612 and 622 are found in a solvent exposed position (Figure 4.2), and residues 628 and 718 are buried in the core of the C-terminal lobe of the protein (Figure 4.1).

Using site directed mutagenesis two tryptophan to phenylalanine substitution mutants were created from the wild type TTK kinase domain. These mutants were W612F W622F and W612F W622F W628F (Table 4.1). The W718 residue was shown to be required for folding since a mutant containing this mutation could not be expressed as soluble protein. Wild type TTK and two tryptophan mutants were expressed and purified to a high purity from *E. coli* (Chapter 4). These mutants were observed to be of the correct molecular mass (Table 4.2) and to have similar CD spectra (Figure 4.14) and similar thermal melting properties (Figure 4.17, Table 4.4) to those of wild type protein.

Using the twin probes of far-UV circular dichroism, which examines the secondary structure content of proteins, and tryptophan fluorescence the folding of the kinase domain of TTK is examined, and the existence and nature of any folding intermediates on the equilibrium folding pathway are identified.

### ***5.2 Materials and Methods***

#### ***5.2.1 Materials***

All chemicals were purchased from Sigma Aldrich (Poole, UK).



### ***5.2.2 Equilibrium Folding and Unfolding***

To perform equilibrium folding and unfolding experiments on wild type TTK and the tryptophan mutants created, two solutions were required for each experiment. For unfolding experiments, the first solution contained native, folded protein at a concentration of 0.1 mg/mL in a 10 mM sodium phosphate pH 7.4 buffer and the second solution contained protein at a concentration of 0.1 mg/mL, unfolded by 6 M GdnHCl 10 mM sodium phosphate pH 7.4. For refolding experiments, the first solution contained protein at a concentration of 0.1 mg/mL, unfolded by 5.25 M GdnHCl 10 mM sodium phosphate pH 7.4 and the second solution contained native, folded protein at a concentration of 0.1 mg/mL in a 10 mM sodium phosphate pH 7.4 buffer. The protein was denatured by different concentrations of GdnHCl since both concentrations resulted in fully unfolded protein, and the concentrations used resulted in a lower consumption of purified protein in the folding experiments.

Equilibrium experiments were performed by exchanging a measured volume of solution in a quartz cuvette to increase or decrease the concentration of denaturant. The total volume of the unfolding and refolding experiments was maintained at a constant 3 mL. After mixing, the sample solution was allowed to equilibrate for 15 minutes prior to a spectrum being recorded. Several samples were allowed to equilibrate overnight after having been analysed after 15 minutes of equilibration. The samples were then reanalysed using the same techniques and the spectra compared to identify if 15 minutes was sufficient time to allow the protein to reach equilibrium. The sample was maintained at a temperature of 20 °C during the experiment. The solution in the quartz cuvette was analysed via tryptophan fluorescence spectroscopy and the measured volume that was removed was analysed by far-UV circular dichroism spectroscopy.

### ***5.2.3 Data Normalisation***

Both far-UV CD spectra (see section 4.2.6) and tryptophan fluorescence spectra (see section 4.2.5) were measured at a series of denaturant concentrations. The spectra were analysed, and features of the spectra were identified that described the changes that the TTK kinase domain underwent upon folding or unfolding. The changes in these features were used to plot transition curves for unfolding or refolding.

To allow a direct comparison between the different methods of reporting on the extent of the folding of the TTK kinase domain it is necessary to transform the data into the same units. The data was normalised as a fraction of folded protein to achieve this, according to equation 5.1 (Pace, 1986).

$$f_N = \frac{(y_D - y_x)}{(y_D - y_n)} \quad (5.1)$$

Where  $f_N$  is the fraction of native protein present at denaturant concentration  $X$ ,  $y_D$  is a measured value for the feature selected under denaturing conditions, i.e.  $\lambda_{\max}$  or signal intensity at a given wavelength,  $y_n$  is the measured value under native conditions, and  $y_x$  is the measured value at denaturant concentration  $X$ .

#### **5.2.4 Analysis of Equilibrium Folding Curves**

##### **5.2.4.1 Two State Folding Transitions**

Two state folding transitions were observed for some spectra. In a two state transition there is a direct conversion of folded to unfolded protein with no intermediate state. There is a single equilibrium constant,  $K_{N \rightarrow U}$ .

The proportion of folded protein present is described by assuming that the fractions of folded and unfolded protein sum to 1 and that their ratio is given by the Boltzman factor. From these assumptions we derive equations 5.2 and 5.3.

$$P_N = \frac{1}{1 + e^{\frac{-\Delta G}{RT}}} \quad (5.2)$$

$$P_U = \frac{e^{\frac{-\Delta G}{RT}}}{1 + e^{\frac{-\Delta G}{RT}}} \quad (5.3)$$

Where,  $P_N$  is the fraction of folded protein,  $P_U$  is the fraction of unfolded protein,  $\Delta G$  is the observed free energy change,  $R$  is the molar gas constant and  $T$  is the temperature in degrees Kelvin.

The fraction of folded protein is the sum of the amounts of each population. The fraction folded is given by.

$$f = S_N P_N + S_U P_U \quad (5.4)$$

The Linear free energy model is used to describe the dependence of the stability of the protein on the denaturant concentration. This model assumes that the difference in the number of accessible sites for denaturant molecules to bind to the protein is directly proportional to the accessible surface area. The relationship between  $\Delta G$  and the denaturant concentration is:

$$\Delta G = m(C_m - [GdnHCl]) \quad (5.5)$$

where  $C_m$  is the mid point of the folding transition, in  $\text{mol dm}^{-3}$ . To fit the data to the two state equilibrium model equation equations 5.2 5.3 and 5.5 are substituted into equation 5.4.

#### 5.2.4.2 Three State Folding Transitions

If the folding transitions observed were determined to be three state, then the data was analysed according to the following method. A three state folding transition is described by equation 5.6, which assumes the accumulation of an intermediate state (I), with two equilibrium constants  $K_{N \rightarrow I}$  and  $K_{I \rightarrow U}$  for the transitions between the native and intermediate state, and the intermediate and unfolded state respectively.



The observed free energy change ( $\Delta G_{\text{obs}}$ ) between the native states, intermediate states and the unfolded state was calculated using equation 5.7 (Tanford, 1970)

$$\Delta G_{\text{obs}} = -RT \ln K \quad (5.7)$$

Where  $R$  is the molar gas constant,  $T$  is the temperature in degrees Kelvin and  $K$  is the equilibrium constant.

If a linear dependence of  $\Delta G_{\text{obs}}$  on the denaturant concentration is assumed, then the free energy of folding ( $\Delta G$ ) can be calculated if this dependence is extrapolated to a denaturant concentration of zero. This is described by equation 5.8 (Greene and Pace, 1974; Pace, 1986; Santoro and Bolen, 1988), where the constant  $m$  is the gradient of  $\Delta G_{\text{obs}}$  vs the denaturant concentration.

$$\Delta G_{\text{obs}} = \Delta G - m[GdnHCl] \quad (5.8)$$

The combination of equations 5.7 and 5.8 then allows for the equilibrium constant,  $K$ , to be calculated.

$$K = e^{-\left(\frac{\Delta G - m[GdnHCl]}{RT}\right)} \quad (5.9)$$

At any denaturant concentration, the fraction folded,  $f_N$ , can be described by equation 5.10 (Morjana *et al.*, 1993). The fraction folded data was fitted to this equation using non-linear least squares regression, using Prism 4 software (Graphpad).

$$f_N = \frac{S_N + S_I K_{N \rightarrow I} + S_U K_{N \rightarrow I} K_{I \rightarrow U}}{1 + K_{N \rightarrow I} + K_{N \rightarrow I} K_{I \rightarrow U}} \quad (5.10)$$

where:

$$K_{N \rightarrow I} = e^{\frac{-(\Delta G_{N \rightarrow I} - m_{N \rightarrow I}[GdnHCl])}{RT}} \quad (5.11)$$

$$K_{I \rightarrow U} = e^{\frac{-(\Delta G_{I \rightarrow U} - m_{I \rightarrow U}[GdnHCl])}{RT}} \quad (5.12)$$

$S_N$ ,  $S_I$  and  $S_U$  are the signal intensities that represent the native, intermediate and unfolded states respectively. The equilibrium constants described in equations 5.9 and 5.10 were determined by substituting equations 5.11 and 5.12 into equation 5.10.

## **5.3 Results**

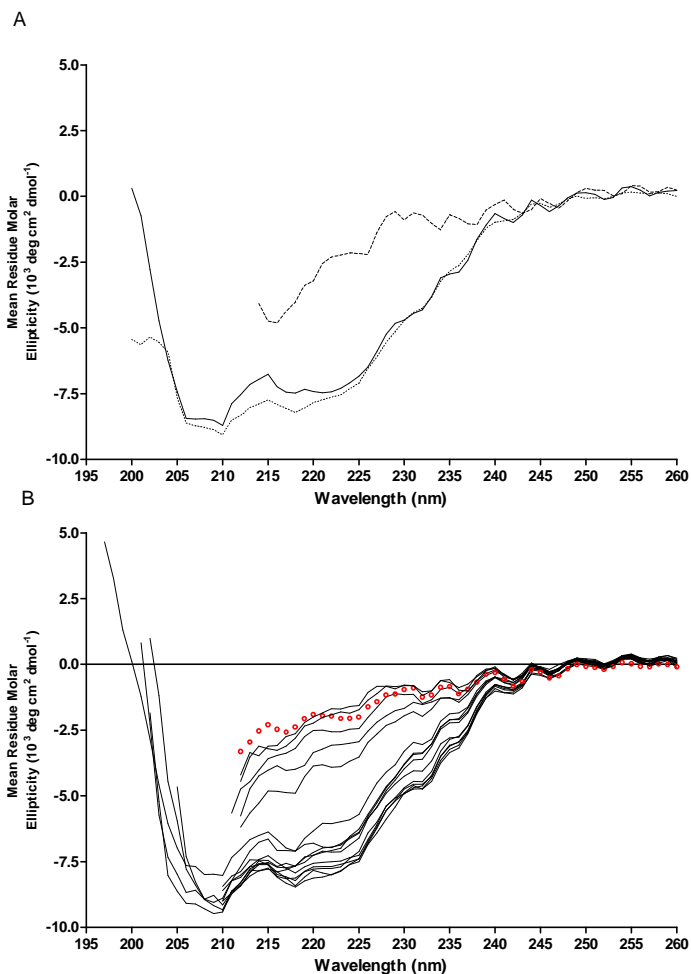
### **5.3.1 Equilibrium Folding and Unfolding of Wild Type TTK**

The folding of wild type TTK was examined by the gradual addition or removal of the denaturant, Guanidine hydrochloride, and examining the extent of folding via the probes of circular dichroism and intrinsic tryptophan fluorescence as described in sections 4.2.7 and 4.2.6 respectively.

As described in section 4.3.7 the CD spectra of wild type TTK shows distinctive minima at 208 nm and 222 nm which are consistent with a protein with a significant  $\alpha$ -helical content. The 222 nm minima of  $\alpha$ -helical content is used as a sensitive probe for the secondary structure content of the protein. The 208 nm band is unusable due to interference from high concentrations of GdnHCl.

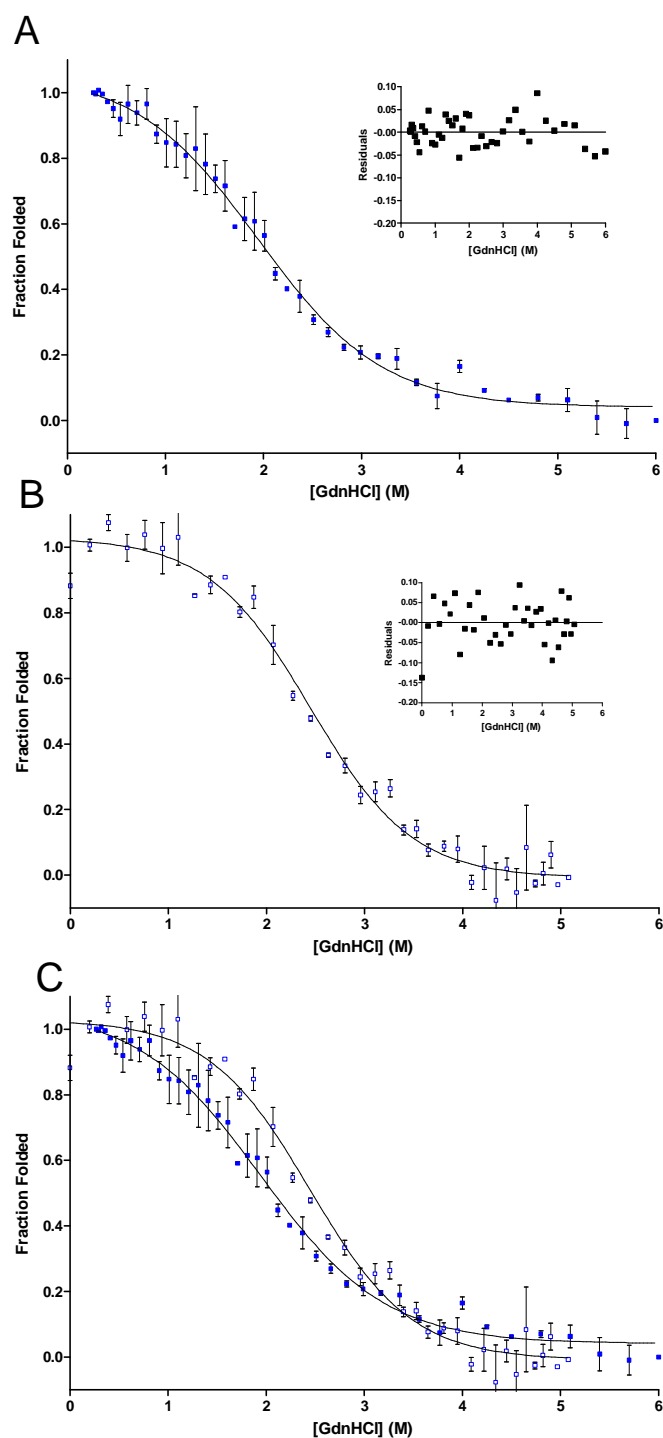
Upon unfolding in 6 M GdnHCl there is a substantial decrease in the intensity of the band at 222 nm, consistent with the protein unfolding to a mostly random coil conformation (Figure 5.1A). If denaturant is added stepwise to folded TTK kinase domain then there is a sequential decrease in the intensity of the 222 nm band (Figure 5.2B). The decrease in the intensity of the 222 nm band is small upon the initial addition of GdnHCl to the protein. Up to concentrations of ~1 M GdnHCl there is little change in the intensity of the CD signal. Once the denaturant concentration rises beyond this, the 222 nm band decreases in intensity, and achieves an intensity similar to that of the protein in 6 M GdnHCl by a denaturant concentration of 4.09 M. If the protein unfolded in 6 M GdnHCl is refolded by the sequential addition of folded protein in native buffer to dilute out the denaturant then a native like CD spectra is obtained, indicating that the protein has fully refolded from the unfolded state (Figure 5.1A). A sequence of spectra similar to those seen when unfolding wild type TTK (Figure 5.1B) is observed when performing this refolding reaction.

For several denaturant concentrations. The CD spectra were compared after incubation times of 15 minutes and overnight (data not shown). There was no significant difference in the observed spectra between the two times, indicating that the incubation time of 15 minutes was sufficient for the protein to reach equilibrium under the conditions used.



**Figure 5.1:** Far-UV CD spectra of wild type TTK kinase domain. (A) Far-UV CD spectra of wild type TTK kinase domain under native (solid line) and denaturing (dashed line) conditions. Far-UV CD spectra of refolded wild type TTK kinase domain shown by dotted line. Denatured spectra shows loss of secondary structure consistent with protein adopting a random coil structure. (B) – Far-UV CD spectra of wild type TTK kinase domain under conditions of varying concentrations of denaturant. Concentrations used; 5 M (red circles), 4.55 M, 4.09 M, 3.53 M, 2.96 M, 2.63 M, 2.07 M, 1.58 M, 1.42 M, 1.27 M, 0.94 M, 0.76 M, 0.58 M, 0.39 M, 0.2 M and native. Spectra collected at 20 °C and at a protein concentration of 0.1 mg/mL.

Figure 5.2 shows the refolding and unfolding curves obtained when plotting the fraction folded of the protein, determined using equation 5.1, from the far-UV CD signal at 222 nm. The unfolding and refolding curves obtained are fitted to two state folding transitions and are nearly super-imposable (Figure 5.2C).



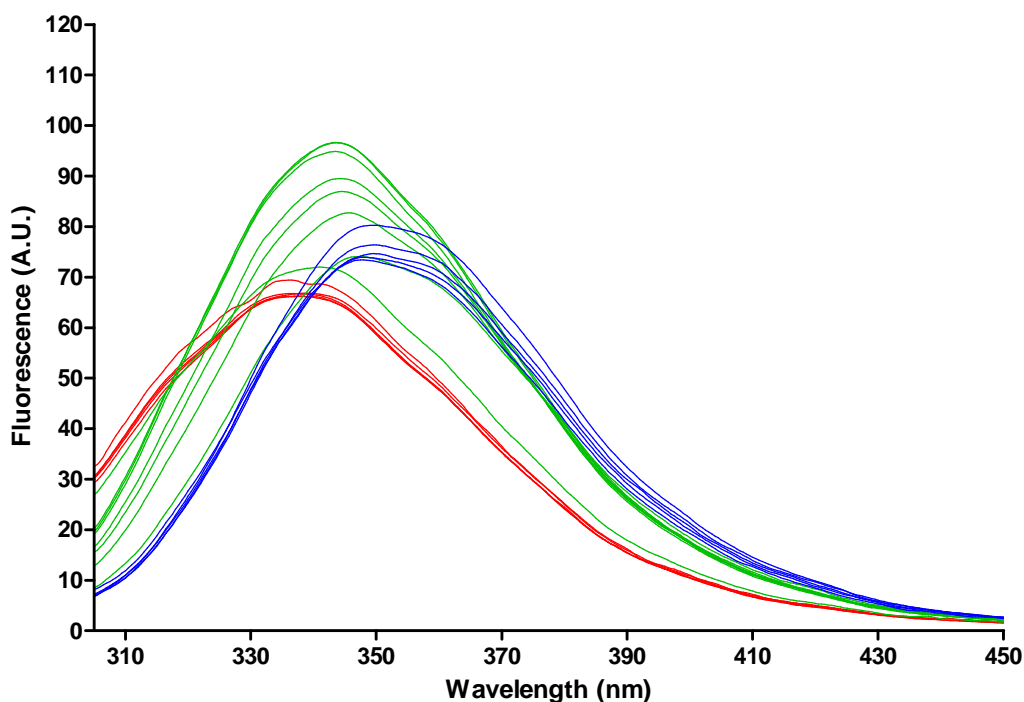
**Figure 5.2:** Unfolding and refolding of wild type TTK induced by GdnHCl and monitored by Far-UV CD. Fraction folded calculated by change in intensity at 222nm following equation 5.1. (A) Refolding of wild type TTK, solid line shows fit to a two state folding model (Equation 5.7). Insert shows residuals of fit. (B) Unfolding of wild type TTK, solid line shows fit to a two state folding model (Equation 5.7). Insert shows residuals of fit. (C) Overlay of refolding and unfolding and fits to two state folding models. Refolding shown as filled squares and unfolding as open squares.

The refolding and unfolding curves were successfully fitted to a two state folding model (Equation 5.4). The refolding transition occurs between ~4 M and ~0.5 M. The mid point of the transition is  $1.95 \pm 0.04$  M GdnHCl. The associated  $m$  value is  $0.91 \pm 0.06$  kcal mol<sup>-1</sup> M<sup>-1</sup> (Table 5.1). The unfolding transition occurs between ~1 M and ~4 M GdnHCl. The mid point of the transition is  $2.45 \pm 0.06$  M GdnHCl with an associated  $m$  value of  $1.11 \pm 0.11$  kcal mol<sup>-1</sup> M<sup>-1</sup> (Table 5.1).

The unfolding and refolding of wild type TTK kinase domain was also observed using intrinsic tryptophan fluorescence. As seen previously (Figure 4.17A) the protein in the native state showed a  $\lambda_{\text{max}}$  of fluorescence of 336.5 nm (Table 4.5). when wild-type TTK is unfolded by the addition of 6 M GdnHCl there is a red shift in the  $\lambda_{\text{max}}$  of fluorescence to a  $\lambda_{\text{max}}$  of 349 nm (Table 4.5), which is typical for tryptophan residues in a fully solvated environment. There is also an increase in the fluorescence intensity at the  $\lambda_{\text{max}}$  (Figure 4.17).

Upon the addition of denaturant to the native protein in a stepwise fashion (Section 5.2.2) there is an increase in the fluorescence intensity and a red shift in the  $\lambda_{\text{max}}$  towards the unfolded  $\lambda_{\text{max}}$  (Figure 5.3). Similar changes are observed in reverse when the denaturant is diluted out.



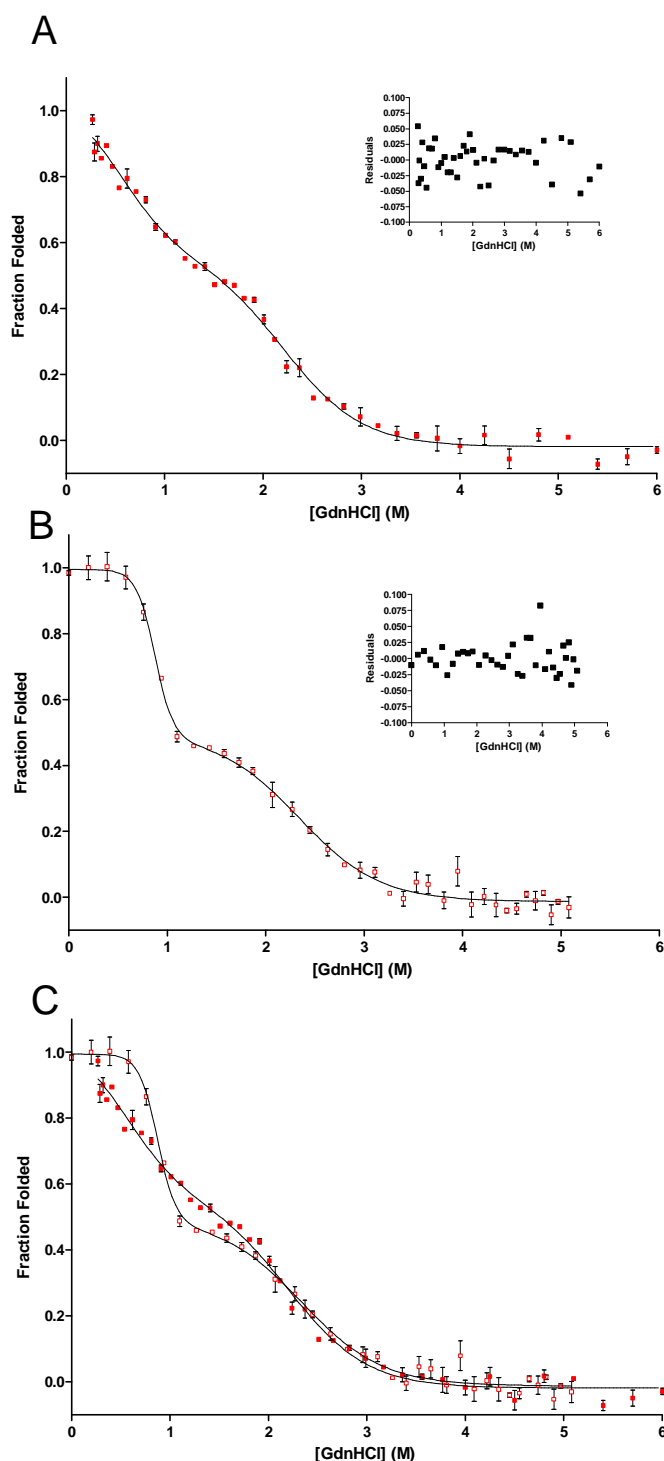


**Figure 5.3:** Fluorescence spectra of wild type TTK under conditions of various denaturant concentration. Concentrations of denaturant for spectra in blue – 5 M, 4.55 M, 4.09 M, 3.53 M, 2.96 M; in green – 2.63 M, 2.07 M, 1.58 M, 1.42 M, 1.27 M, 0.94 M; and in red – 0.76 M, 0.85 M, 0.39 M, 0.2 M, and native.

The fluorescence intensity of wild type TTK shows an increase in intensity when the denaturant concentration is decreased from the unfolded state. The fluorescence intensity reaches a peak at ~1.5 M GdnHCl. The maximum intensity then decays to reach the native state. This behaviour of the protein when unfolding suggests the formation of an equilibrium intermediate on the folding pathway.

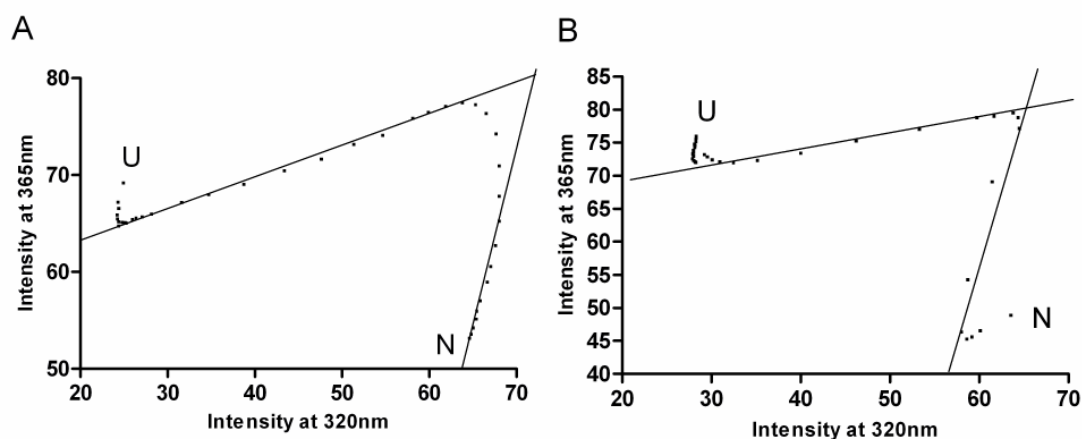
As shown in Figure 5.3, the fluorescence intensity of wild type TTK behaves in a complex fashion. The fluorescence intensity at the native  $\lambda_{\text{max}}$  rapidly rises above the intensity of the native state and then falls below the intensity of the native state as the protein unfolds. This causes a problem for describing the folding using fluorescence intensity at the native  $\lambda_{\text{max}}$ , a commonly used measure of the extent of folding when probing the folding of a protein using fluorescence spectroscopy. Therefore, the  $\lambda_{\text{max}}$  for each spectra was recorded, and used to determine the extent of folding for each spectra. The unfolding and refolding transitions were not well fitted to a two state folding model. A visual inspection of the folding transitions appears to indicate the formation of an equilibrium folding intermediate. The folding

transitions were, therefore, fitted to a three state folding model, which includes an intermediate state in addition to the native and the unfolded state. The fraction folded was calculated by equation 5.1 and the resultant data fitted to equation 5.10.



**Figure 5.4:** Refolding and unfolding of wild type TTK induced by GdnHCl and monitored by change in  $\lambda_{\max}$  of tryptophan fluorescence. (A) Refolding of TTK, solid line shows fit to three state folding model (Equation 5.10). Insert shows residuals. (B) Unfolding of wild type TTK, solid line shows fit to three state folding model (Equation 5.10). Insert shows residuals. (C) Overlay of refolding and unfolding transitions. Solid lines show three state folding model fits. Refolding shown as filled squares and unfolding as open squares.

To support the use of a three state folding model in describing the folding of wild-type TTK, phase diagram analysis was used (Kuznetsova *et al.*, 2003). Phase diagram analysis is a technique which is particularly sensitive to the presence of intermediates in the folding transitions. The analysis however, cannot be used to derive any thermodynamic parameters for the folding transitions examined. It can, therefore, be used to qualitatively identify the presence of folding intermediates. The technique can be applied to many combinations of folding probes, comparing, for example, near and far UV CD or CD and fluorescence. To apply the technique to the probe of tryptophan fluorescence, fluorescence intensities on opposite sides of the  $\lambda_{\text{max}}$  are plotted against each other. A simple two state folding transition will then produce a straight line plot. If folding intermediates are populated in the folding transition, then these intermediates appear as inflections in the phase diagram. Inspection of these inflection points can identify the position of the intermediate on the folding transition. Folding phase diagrams were constructed for wild type TTK folding, probed by tryptophan fluorescence for both the refolding (Figure 5.5A) and unfolding (Figure 5.5B) transitions using the fluorescence intensities at 320 nm and 365 nm.



**Figure 5.5:** Folding phase diagrams for wild type TTK. (A) Phase diagram for the refolding of wild type TTK. (B) Phase diagram for the unfolding of wild type TTK. Positions of the unfolded and native states have been indicated. Phase diagrams constructed using fluorescence intensity at 320 nm and 365 nm.

The phase diagrams for the folding of wild type TTK indicate that both the unfolding and refolding of wild type TTK occur through an equilibrium

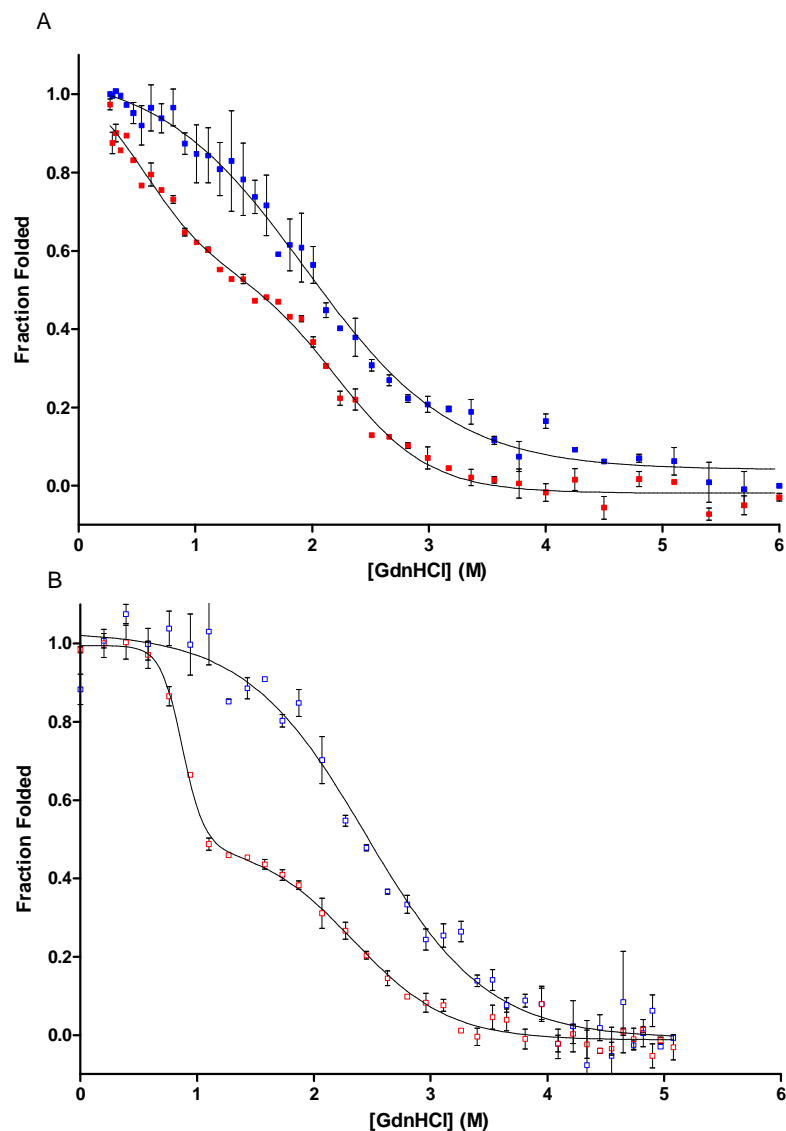
intermediate. This supports the use of a three state folding model to describe the refolding and unfolding transitions of wild type TTK. The inflections close to the unfolded and native states are not regarded as being significant.

The refolding transition occurs in two stages. The first transition occurs between 3.5 M GdnHCl and ~1.5 M GdnHCl. The mid point of this transition is at ~2.4 M GdnHCl. The free energy change,  $\Delta G_{U \rightarrow I}$ , associated with this transition is  $-3.24 \pm 0.78 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.46 \pm 0.27 \text{ kcal mol}^{-1} \text{M}^{-1}$  (Table 5.2). The second transition occurs between ~1.5 M GdnHCl and ~0.2 M GdnHCl with a mid-point of ~0.75 M GdnHCl. The free energy change,  $\Delta G_{I \rightarrow N}$ , associated with this transition is  $-1.08 \pm 1.04 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.97 \pm 1.26 \text{ kcal mol}^{-1} \text{M}^{-1}$  (Table 5.2).

The unfolding transition of wild type TTK occurs in two very distinct phases with a clear accumulation of an intermediate state. The first transition occurs between native conditions and 1.4 M GdnHCl, with a mid point of ~0.8 M. The free energy change,  $\Delta G_{N \rightarrow I}$ , associated with this transition is  $5.38 \pm 0.94 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $6.15 \pm 1.09 \text{ kcal mol}^{-1} \text{M}^{-1}$  (Table 5.2). The second transition occurs between ~1.5 M GdnHCl and ~4 M GdnHCl with a mid-point of ~2.5 M GdnHCl. The free energy change,  $\Delta G_{I \rightarrow U}$ , associated with this transition is  $3.33 \pm 0.54 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.41 \pm 0.20 \text{ kcal mol}^{-1} \text{M}^{-1}$  (Table 5.3).

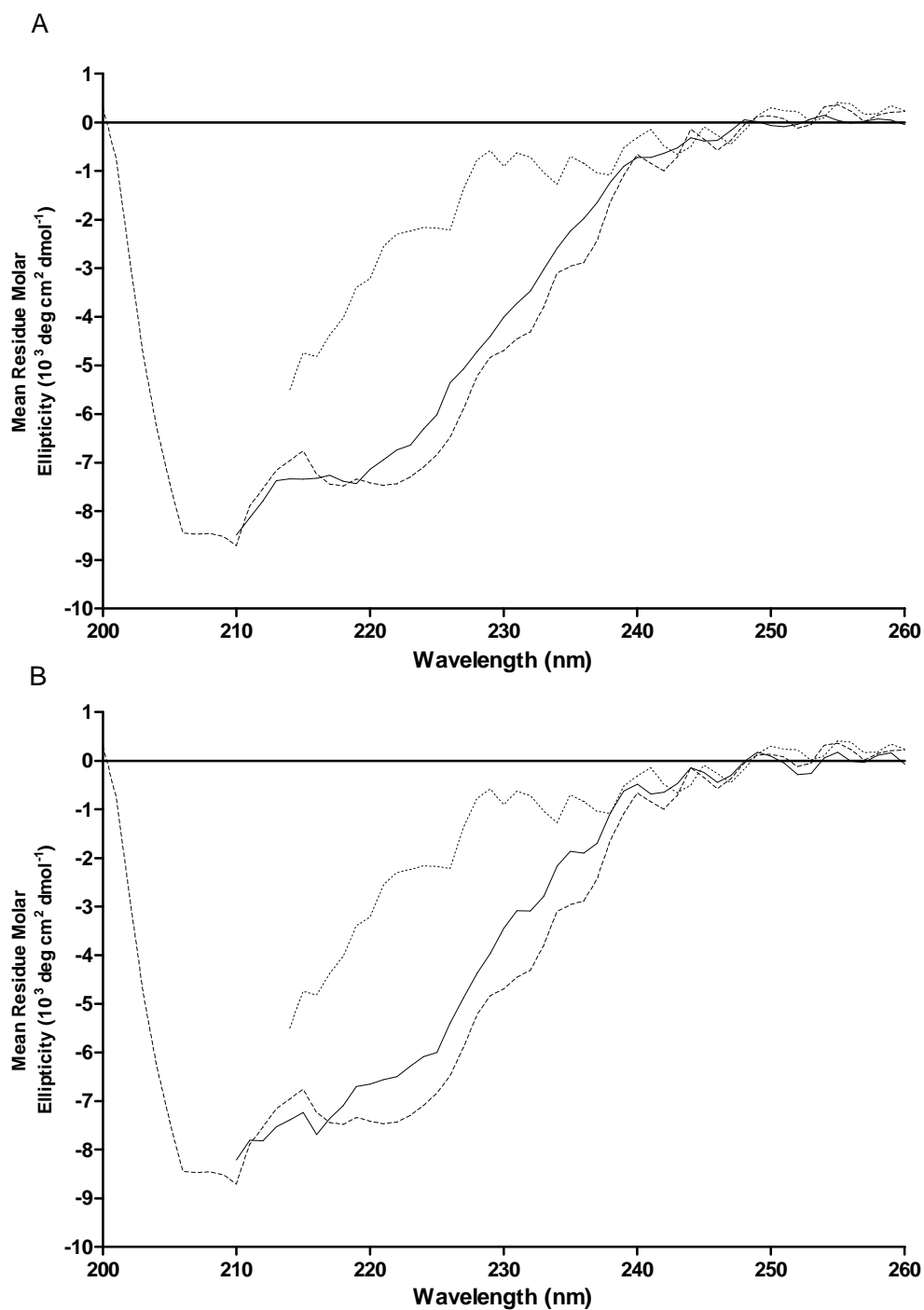
Information on the nature of the intermediate states discovered can be elucidated by comparing the folding as measured by far-UV CD and tryptophan fluorescence. Figure 5.6 shows these overlays. The intermediate formed upon unfolding of the protein occurs at ~1.5 M GdnHCl. At this denaturant concentration the far-UV CD spectra show that the secondary structure is almost completely formed. This indicates that a molten globule intermediate is being formed.

Upon refolding a different behaviour is observed. The far-UV CD monitoring of the refolding shows a recovery of secondary structure which is more coincident with the refolding observed by tryptophan fluorescence (Figure 5.6).

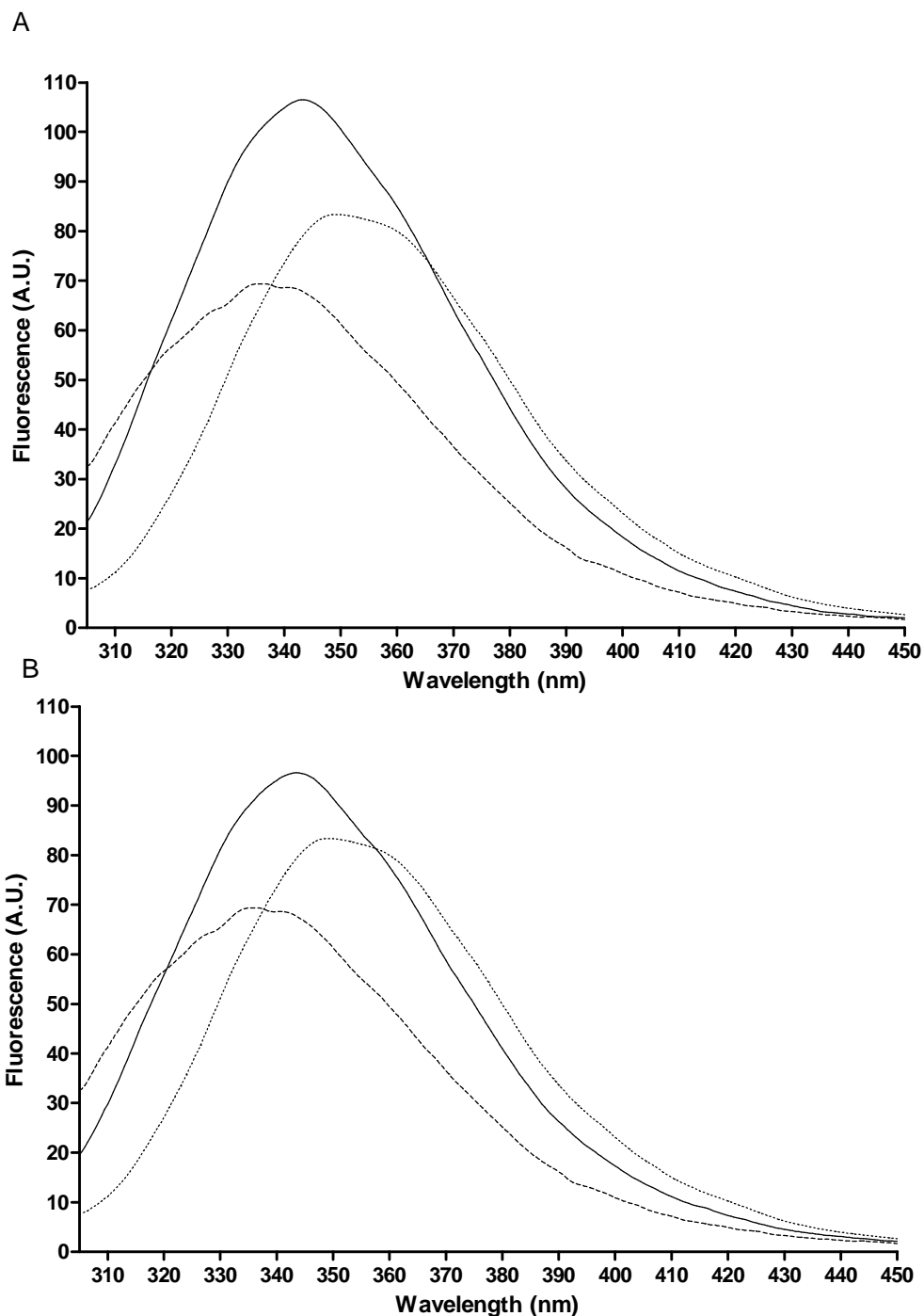


**Figure 5.6:** Overlay of far-UV CD and tryptophan fluorescence monitoring of the folding of wild type TTK. (A) Refolding of wild type TTK. (B) Unfolding of wild type TTK. CD data in blue, fluorescence in red. Solid lines show 3 state folding fits as described in Figures 5.3 and 5.5.

Further insights into the nature of the intermediate states identified can be obtained by an examination of the far-UV CD and tryptophan fluorescence spectrums of wild type TTK at the denaturant concentrations where the intermediate was noted to occur. The denaturant concentration to be examined was determined from the phase diagrams of the fluorescence monitored folding of wild type TTK (Figure 5.5).



**Figure 5.7:** Far-UV CD spectra of wild type TTK in native, intermediate and unfolded states. (A) wild type TTK under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon refolding of wild type TTK. GdnHCl concentration of intermediate 1.21 M. (B) wild type TTK under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon unfolding of wild type TTK. GdnHCl concentration of intermediate 1.43 M.



**Figure 5.8:** Tryptophan fluorescence spectra of wild type TTK in native, intermediate and unfolded states. (A) wild type TTK under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon refolding of wild type TTK. GdnHCl concentration of intermediate 1.21 M. (B) wild type TTK under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon unfolding of wild type TTK. GdnHCl concentration of intermediate 1.43 M.

The CD spectra of wild type TTK at the denaturant concentration corresponding to the folding intermediate identified by phase diagram analysis



(Figure 5.5A,B) show that both on refolding and upon unfolding the intermediate retains a high degree of secondary structure (Figure 5.7A,B). The degree of secondary structure present is similar upon refolding and unfolding of the protein.

The tryptophan fluorescence spectra of the intermediate formed in the folding of wild-type TTK is also similar upon unfolding (Figure 5.8B) and refolding (Figure 5.8A). The  $\lambda_{\text{max}}$  of the fluorescence spectra is blue shifted from the unfolded state, as shown in Figure 5.8. In addition the fluorescence intensity of the intermediate states is similar on both the unfolding and refolding transitions. This indicates that the intermediates may contain a similar degree of tertiary structure, which is in between the native and the unfolded states.

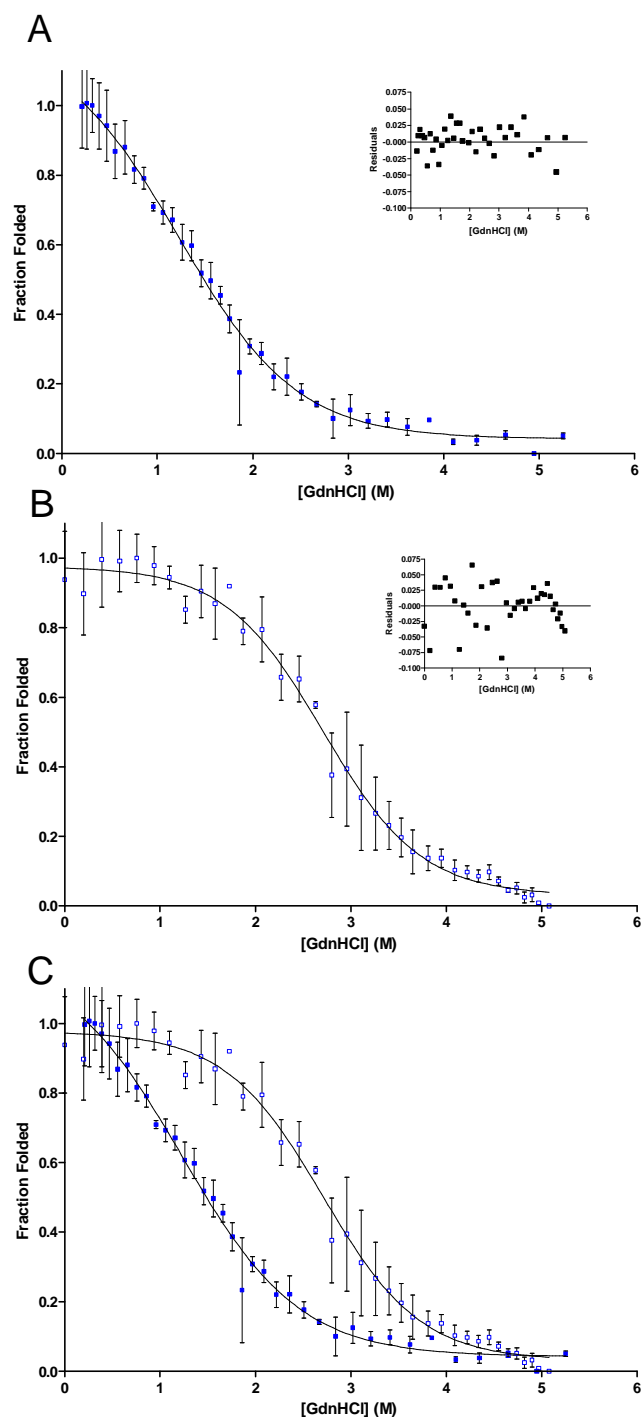
### 5.3.2 *Equilibrium Folding and Unfolding of TTK<sup>W612F W622F</sup>*

The crystal structure of the wild type TTK kinase domain indicates that tryptophans 612 and 622 are positioned such that they are exposed to solvent (Figure 4.2). These solvent exposed tryptophan residues were mutated to phenylalanine to allow the folding to be followed using the more specific probes of the two buried tryptophan residues in the C-terminal lobe of the kinase domain. The far-UV CD spectra (Figure 4.16) and the thermal unfolding (Figure 4.14A) indicate that no major structural perturbation has been induced by the replacement of W612 and W622 with phenylalanine.

The folding of TTK<sup>W612F W622F</sup> was followed by far-UV CD intensity at 222 nm (Figure 4.16). Similar to the spectra observed for wild type TTK kinase domain, upon the addition of GdnHCl the intensity of the 222 nm band decreases until the spectra is consistent with the protein adopting a random coil conformation. Upon the removal of the denaturant by dilution, a native like CD spectra was recovered, indicating complete refolding of the protein.

The far-UV CD intensity at 222 nm was converted to the fraction of protein folded (Equation 5.1). The refolding of TTK<sup>W612F W622F</sup> occurs via a two state transition (Figure 5.9A), with a mid point of 1.23 M GdnHCl. The total free energy change associated with the transition is  $-4.70 \pm 0.36 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $0.94 \pm 0.07 \text{ kcal mol}^{-1}\text{M}^{-1}$  (Table 5.1) Similarly, the unfolding of TTK<sup>W612F W622F</sup> also occurs via a two state transition (Figure 5.9B), with a mid point of 2.71 M GdnHCl. The total free energy change associated with the transition is

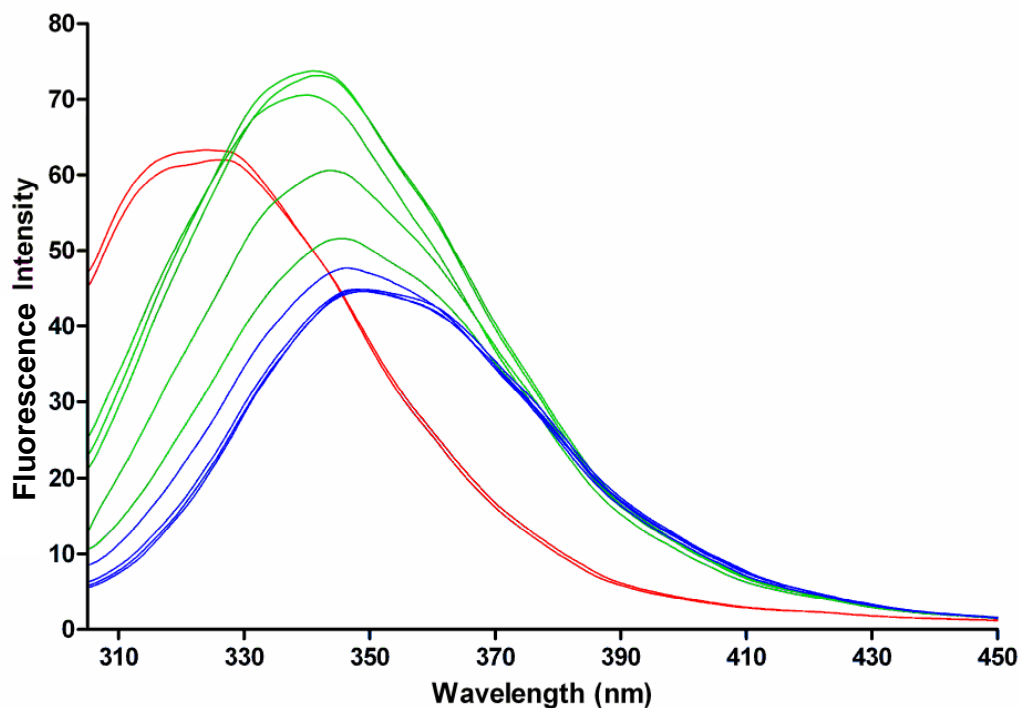
$5.93 \pm 0.44 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.13 \pm 0.08 \text{ kcal mol}^{-1}\text{M}^{-1}$   
(Table 5.1)



**Figure 5.9:** Unfolding and refolding of TTK<sup>W612F W622F</sup> induced by GdnHCl and monitored by far-UV CD. Fraction folded calculated by change in intensity at 222nm following equation 5.1. (A) Refolding of TTK<sup>W612F W622F</sup>, solid line shows fit to a two state folding model (Equation 5.7). Insert shows residuals of fit. (B) Unfolding of TTK<sup>W612F W622F</sup>, solid line shows fit to a two state folding model (Equation 5.7). Insert shows residuals of fit. (C) Overlay of refolding and unfolding and fits to two state folding models. Refolding shown as filled squares and unfolding as open squares.

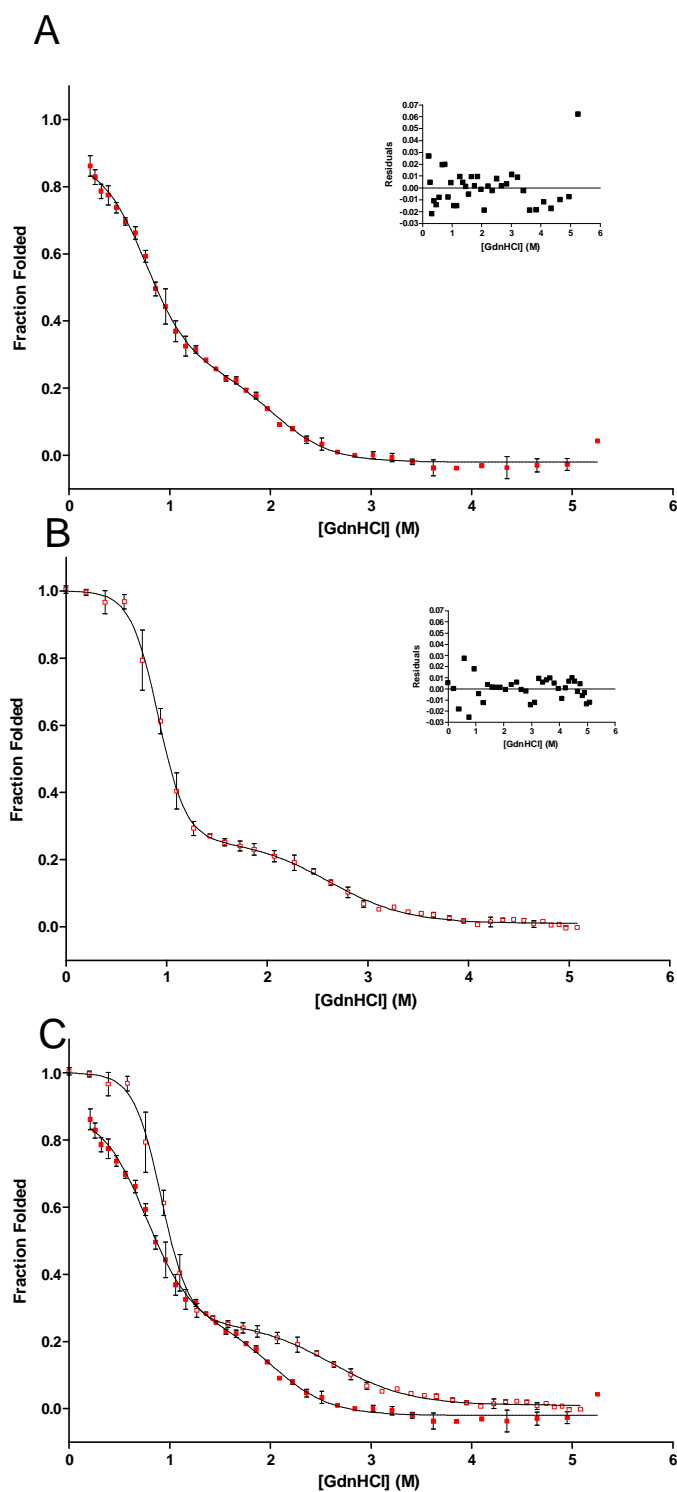
The refolding and unfolding curves observed by far-UV CD are not super-imposable. The recovery of secondary structure content on the refolding of TTK<sup>W612F W622F</sup> is shifted to higher GdnHCl concentrations compared to the loss of secondary structure on unfolding. The native like secondary structure is more stable upon unfolding, being retained up to a GdnHCl concentration of ~1 M. The full unfolding of the secondary structure does not occur until higher denaturant concentrations also and the recovery of secondary structure does not occur until lower GdnHCl concentrations of ~ 3.5 M upon refolding.

The folding of TTK<sup>W612F W622F</sup> was monitored by tryptophan fluorescence as a comparison to the folding monitored by far-UV CD spectroscopy. The native protein has a blue shifted  $\lambda_{\text{max}}$  of fluorescence compared to wild type TTK (Table 4.5). The  $\lambda_{\text{max}}$  of the W612F W622F mutant is 319 nm. When the protein is unfolded by the addition of 6 M GdnHCl the  $\lambda_{\text{max}}$  of fluorescence red shifts to 349 nm. Unlike wild type TTK there is no increase in the intensity of the tryptophan fluorescence upon full unfolding (Figure 4.18B). This is due to the replacement of the solvent exposed tryptophan residues with phenylalanine. Upon a stepwise unfolding of the native protein in GdnHCl there is a red shift in the  $\lambda_{\text{max}}$  towards the unfolded  $\lambda_{\text{max}}$  as well as an initial increase in the fluorescence intensity towards an apparent intermediate state. Following this there is a decrease in the fluorescence intensity and further red shift in the  $\lambda_{\text{max}}$  to the unfolded state (Figure 5.10).



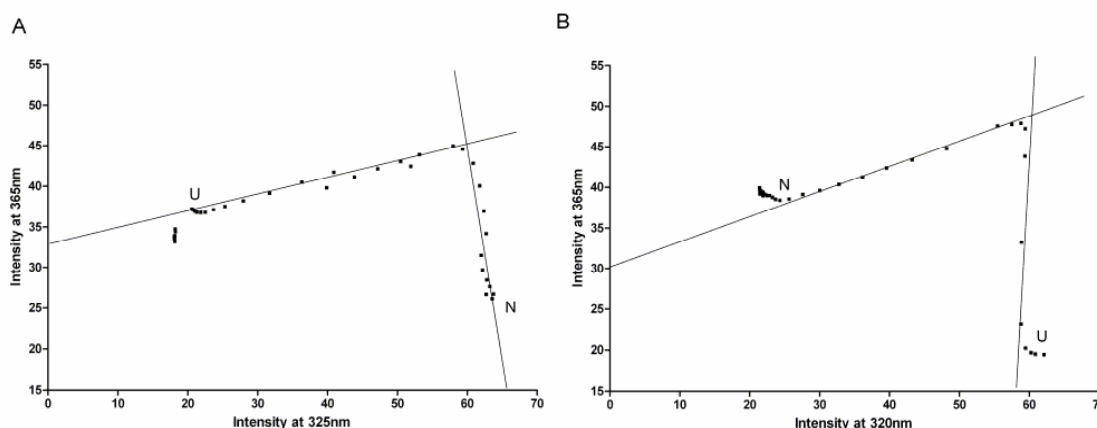
**Figure 5.10:** Fluorescence spectra of TTK<sup>W612F W622F</sup> under conditions of various denaturant concentration. Concentrations of denaturant for spectra in blue – 2.96 M, 3.65 M, 4.09 M and 4.95 M; in green – 1.1 M, 1.27 M, 1.58 M, 2.07 M and 2.63 M; and in red – 0.58 M and 0.2 M.

The refolding and unfolding transitions observed by tryptophan fluorescence were not well fitted to a two state folding model. The transitions were better fitted to a three state folding model, which describes an equilibrium between the unfolded, native and an intermediate state. Since the fluorescence intensity at the native  $\lambda_{\text{max}}$  rises to a value greater than the native intensity (Figure 5.10) the  $\lambda_{\text{max}}$  at each denaturant concentration used was recorded, and the folding transitions plotted using this measure. The fraction folded was calculated by equation 5.1 and the resultant data fitted to equation 5.10.



**Figure 5.11:** Refolding and unfolding of TTK<sup>W612F</sup> W622F induced by GdnHCl and monitored by change in  $\lambda_{\text{max}}$  of tryptophan fluorescence. (A) Refolding of TTK<sup>W612F</sup> W622F, solid line shows fit to three state folding model (Equation 5.10). Insert shows residuals. (B) Unfolding of TTK<sup>W612F</sup> W622F, solid line shows fit to three state folding model (Equation 5.10). Insert shows residuals. (C) Overlay of refolding and unfolding transitions. Solid lines show three state folding model fits. Refolding shown as filled squares and unfolding as open squares.

Phase diagram analysis was again used to support the fitting of the folding transitions to a three state folding model. Phase diagram analysis indicates that both the refolding (Figure 5.12A) and unfolding (Figure 5.12B) proceed via an equilibrium folding intermediate. The phase diagrams do not indicate that there are more than three states in the folding transitions observed.



**Figure 5.12:** Folding phase diagrams for TTK<sup>W612F W622F</sup>. (A) Phase diagram for the refolding of TTK<sup>W612F W622F</sup>. (B) Phase diagram for the unfolding of TTK<sup>W612F W622F</sup>. Positions of the unfolded and native states have been indicated. Phase diagrams constructed using fluorescence intensity at 320 nm and 365 nm.

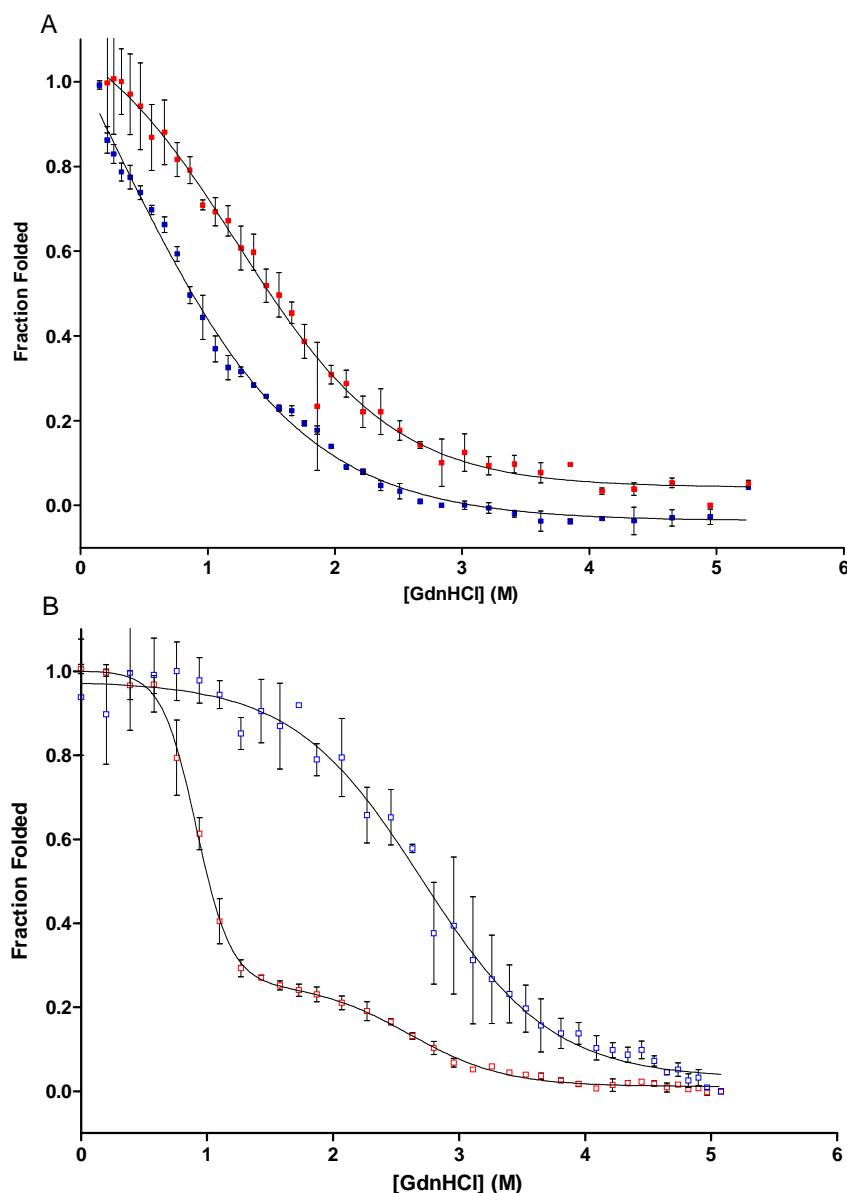
The refolding and unfolding transitions were fitted to three state models as indicated by phase diagram analysis (Figure 5.12). The refolding of TTK<sup>W612F W622F</sup> occurs in two phases. The first transition occurs between ~3 M GdnHCl and ~1 M GdnHCl (Figure 5.11A). The mid point of this transition is at ~2 M GdnHCl. The free energy change,  $\Delta G_{I \rightarrow U}$ , associated with this transition is  $-4.15 \pm 1.10 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $2.04 \pm 0.45 \text{ kcal mol}^{-1} \text{M}^{-1}$  (Table 5.2). The second transition occurs between ~1 M GdnHCl and ~0.2 M GdnHCl with a mid-point of ~0.5 M GdnHCl. The free energy change,  $\Delta G_{N \rightarrow I}$ , associated with this transition is  $-1.97 \pm 0.31 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $2.60 \pm 0.40 \text{ kcal mol}^{-1} \text{M}^{-1}$  (Table 5.2).

The unfolding transition of TTK<sup>W612F W622F</sup> occurs in two very distinct phases with a clear accumulation of an intermediate state (Figure 5.11B). The first transition occurs between native conditions and ~1.5 M GdnHCl, with a mid point of ~0.9 M. The free energy change,  $\Delta G_{N \rightarrow I}$ , associated with this transition is  $3.86 \pm 0.20 \text{ kcal}$

$\text{mol}^{-1}$  with a corresponding  $m$  value of  $4.22 \pm 0.22 \text{ kcal mol}^{-1}\text{M}^{-1}$  (Table 5.2). The second transition occurs between  $\sim 1.5 \text{ M GdnHCl}$  and  $\sim 4 \text{ M GdnHCl}$  with a midpoint of  $\sim 2.5 \text{ M GdnHCl}$ . The free energy change,  $\Delta G_{\text{I} \rightarrow \text{U}}$ , associated with this transition is  $3.90 \pm 0.52 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.50 \pm 0.18 \text{ kcal mol}^{-1}\text{M}^{-1}$  (Table 5.2).

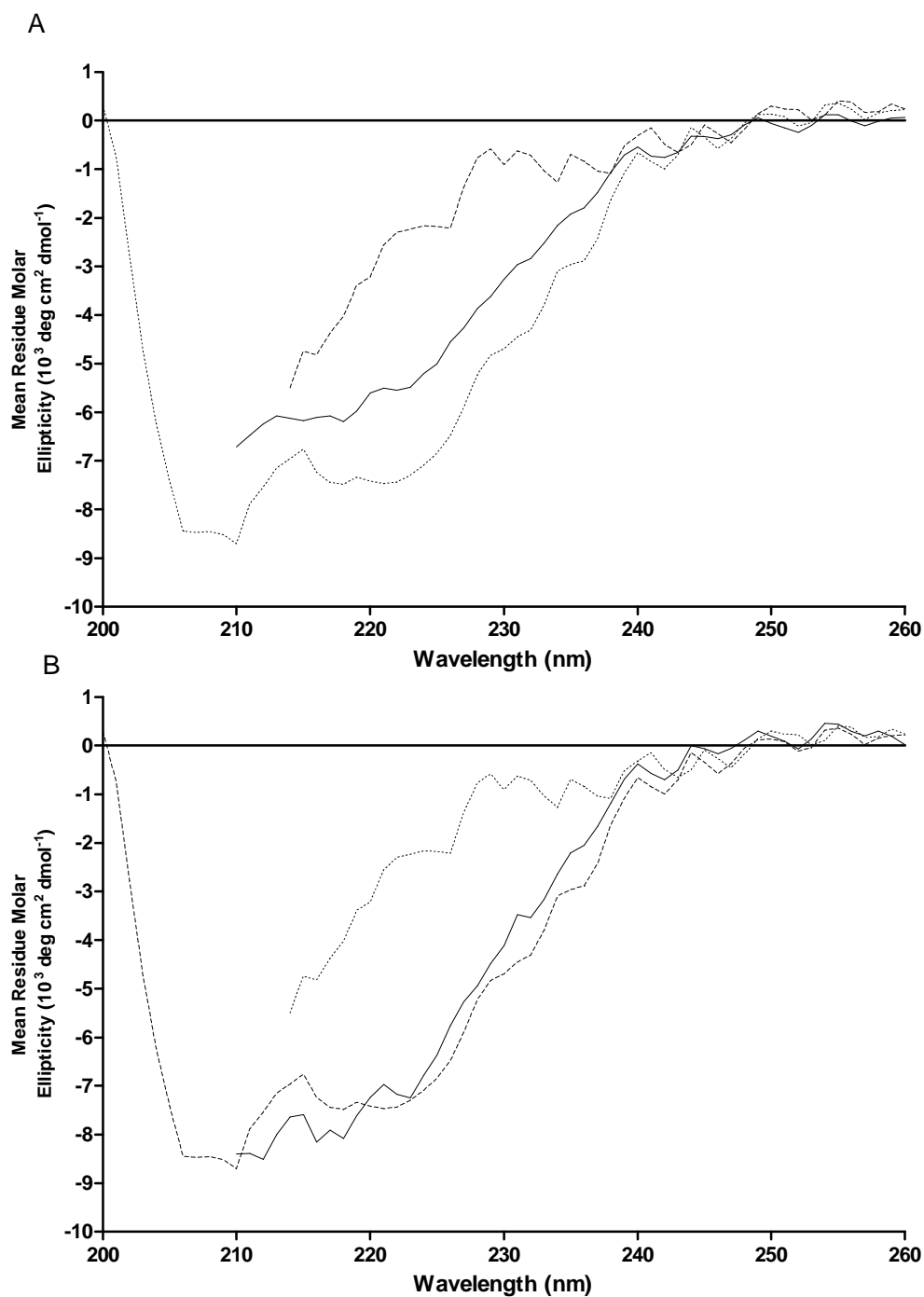
Information on the nature of the intermediate states discovered can be elucidated by comparing the folding as measured by far-UV CD and tryptophan fluorescence. Figure 5.13 shows these overlays. The intermediate formed upon unfolding of the protein occurs at  $\sim 1.5 \text{ M GdnHCl}$ . At this denaturant concentration the far-UV CD spectra show that the secondary structure is almost completely retained (Figure 5.13B). The tertiary structure of the protein is only partially retained. This indicates the presence of a molten globule folding intermediate. The relationship of the recovery of secondary and tertiary structure on the refolding of the W612F W622F mutant is different to the relationship observed upon unfolding (Figure 5.13A). The intermediate formed on the refolding pathway is not a molten globule type.



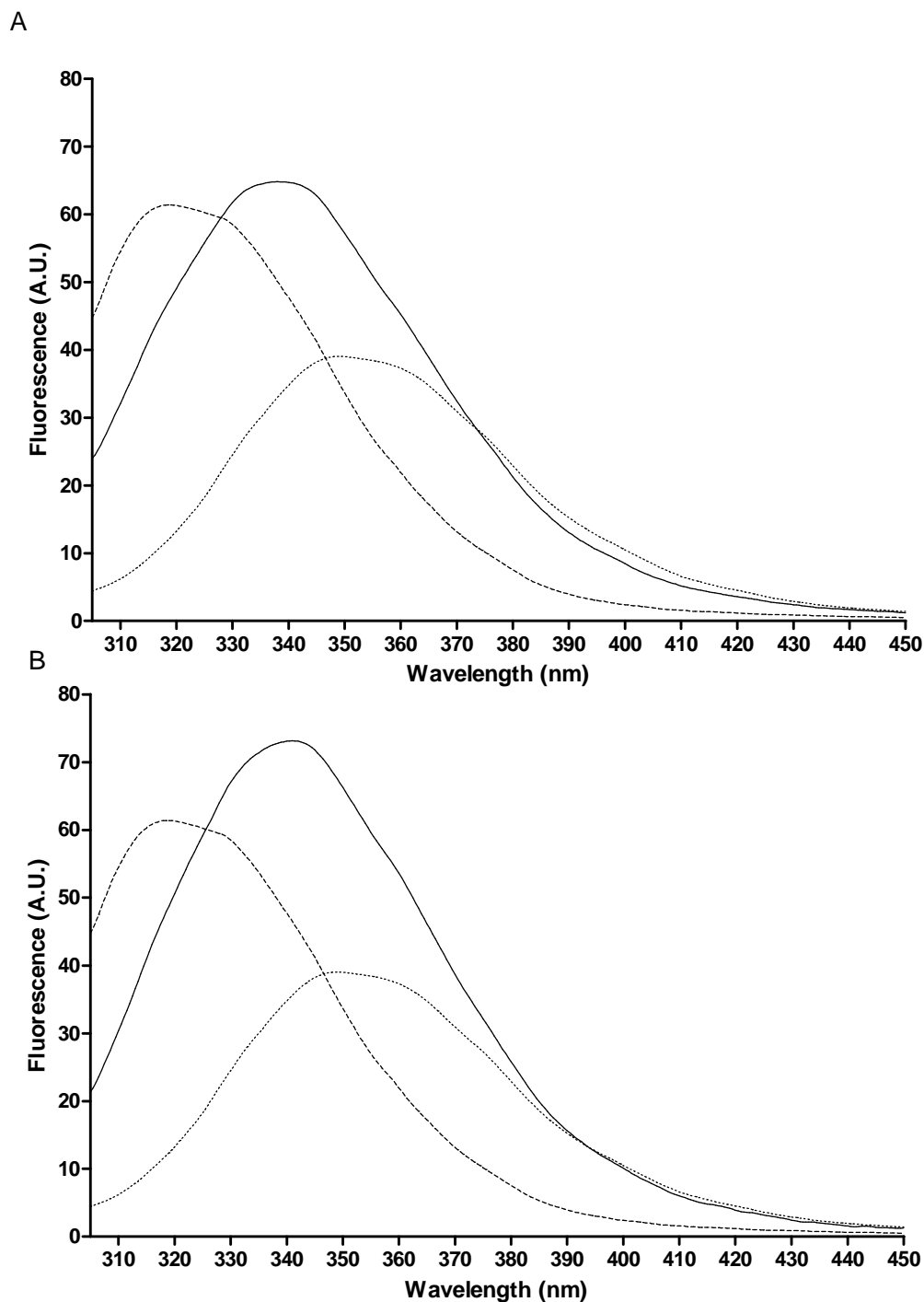


**Figure 5.13:** Overlay of far-UV CD and tryptophan fluorescence monitoring of the folding of TTK<sup>W612F W622F</sup>. (A) Refolding of TTK<sup>W612F W622F</sup>. (B) Unfolding of TTK<sup>W612F W622F</sup>. CD data in blue, fluorescence in red. Solid lines show three state or two state folding fits as described in Figures 5.10 and 5.12.

Further insights into the nature of the intermediate states identified can be obtained by an examination of the far-UV CD and tryptophan fluorescence spectrums of TTK<sup>W612F W622F</sup> at the denaturant concentrations where the intermediate was noted to occur. The denaturant concentration to be examined was determined from the phase diagrams of the fluorescence monitored folding of TTK<sup>W612F W622F</sup> (Figure 5.12).



**Figure 5.14:** Far-UV CD spectra of TTK<sup>W612F W622F</sup> in native, intermediate and unfolded states. (A) TTK<sup>W612F W622F</sup> under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon refolding of TTK<sup>W612F W622F</sup>. GdnHCl concentration of intermediate 1.06 M. (B) TTK<sup>W612F W622F</sup> under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon unfolding of TTK<sup>W612F W622F</sup>. GdnHCl concentration of intermediate 1.43 M.



**Figure 5.15:** Tryptophan fluorescence spectra of  $\text{TTK}^{\text{W612F W622F}}$  in native, intermediate and unfolded states. (A)  $\text{TTK}^{\text{W612F W622F}}$  under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon refolding of  $\text{TTK}^{\text{W612F W622F}}$ . GdnHCl concentration of intermediate 1.06 M. (B)  $\text{TTK}^{\text{W612F W622F}}$  under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon unfolding of  $\text{TTK}^{\text{W612F W622F}}$ . GdnHCl concentration of intermediate 1.43 M.

The CD spectra of TTK<sup>W612F W622F</sup> at the denaturant concentration corresponding to the folding intermediate identified by phase diagram analysis (Figure 5.13A,B) show that on unfolding of the protein the intermediate retains a high degree of secondary structure (Figure 5.14B). However upon refolding the intermediate formed contains a reduced amount of secondary structure content. The CD spectra shows a CD intensity at 222 nm of ~60 % of the difference between the native and the unfolded state.

The tryptophan fluorescence spectrum of the intermediate formed in the folding of wild-type TTK is similar upon unfolding (Figure 5.15B) and refolding (Figure 5.15A). The  $\lambda_{\text{max}}$  of the fluorescence spectra is blue shifted from the unfolded state, as shown in Figure 5.15. In addition the fluorescence intensity of the intermediate states is similar on both the unfolding and refolding transitions. This indicates that the intermediates may contain a similar degree of tertiary structure, which is in between the native and unfolded states.

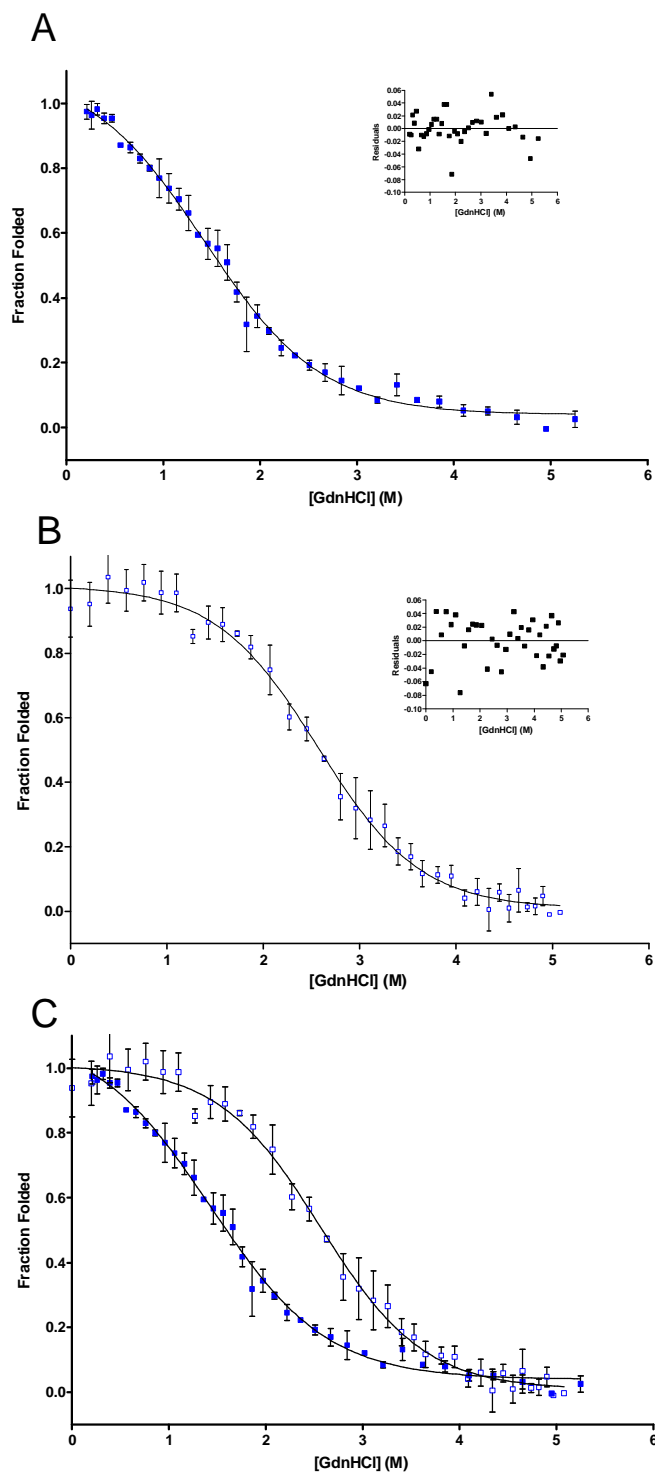
The CD and fluorescence spectra support the identification of the intermediate on the unfolding pathway as a molten globule intermediate. The differences in the CD spectra of the intermediate formed on the refolding pathway demonstrate that this intermediate is probably not a classical molten globule type intermediate.

### **5.3.3 Equilibrium Folding and Unfolding of TTK<sup>W612F W622F W628F</sup>**

The study of the folding of proteins by tryptophan fluorescence is best performed using proteins containing a single tryptophan residue, since multiple tryptophans in multiple environments are difficult to interpret with many possible changes occurring in the folding transition. The multiple tryptophans may also mask folding events reported on by particular tryptophans. Single tryptophan mutants of proteins allow the study of the folding of proteins to be better observed. Previous work on the folding of p38 $\alpha$  had identified a single tryptophan, W207, which was essential for the folding of p38 $\alpha$  and was found to be absolutely conserved throughout the kinase. For comparison with p38 $\alpha$  and to allow examination of the folding of the C-terminal lobe, the single tryptophan mutant, TTK<sup>W612F W622F W628F</sup> was created. The replacement of the single tryptophan W718 alone results in a construct which is insoluble. This prevented the study of other mutants of TTK containing one tryptophan residue, since W718 could not be substituted.

The thermal melting analysis (Figure 4.15B) and the native CD spectra (Figure 4.13) indicates that the mutations performed do not have a detrimental effect on the stability of the kinase domain fold of TTK<sup>W612F W622F W628F</sup>.

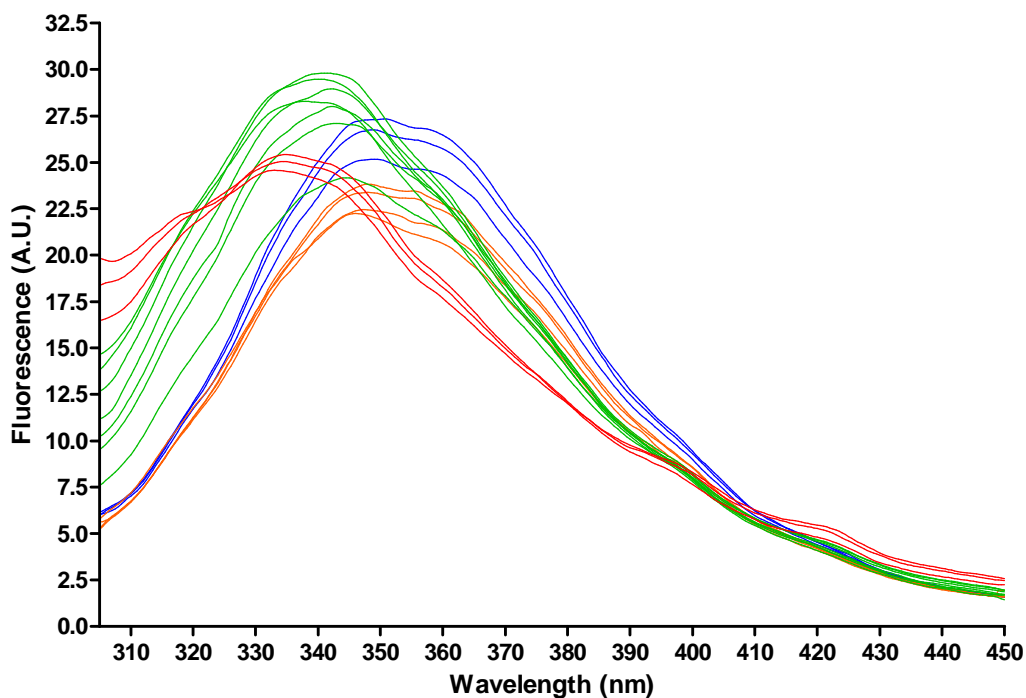
The folding of TTK<sup>W612F W622F W628F</sup> was examined following the far-UV CD intensity at 222 nm (Figure 5.16). Similar to the spectra observed for wild type TTK kinase domain and TTK<sup>W612F W622F</sup>, upon the addition of GdnHCl the intensity of the 222 nm band decreases until the spectrum is consistent with the protein adopting a random coil conformation. Upon the dilution of the denaturant, a native like CD spectra was recovered, indicating complete refolding of the protein.



**Figure 5.16:** Unfolding and refolding of TTK<sup>W612F W622F W628F</sup> induced by GdnHCl and monitored by far-UV CD. Fraction folded calculated by change in intensity at 222 nm following equation 5.1. (A) Refolding of TTK<sup>W612F W622F W628F</sup>, solid line shows fit to a two state folding model (Equation 5.7). Insert shows residuals of fit. (B) Unfolding of TTK<sup>W612F W622F W628F</sup>, solid line shows fit to a two state folding model (Equation 5.7). Insert show residuals of fit. (C) Overlay of refolding and unfolding and fits to two state folding models. Refolding shown as filled squares and unfolding as open squares.

The far-UV CD intensity at 222nm was converted to the fraction of protein folded (equation 5.1). The refolding of TTK<sup>W612F W622F W628F</sup> occurs via a two state transition (Figure 5.16A), with a mid point of 1.43 M GdnHCl. The total free energy change associated with the transition is  $-4.81 \pm 0.30 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $0.96 \pm 0.04 \text{ kcal mol}^{-1}\text{M}^{-1}$  (Table 5.1). Similarly, the unfolding of TTK<sup>W612F W622F W628F</sup> also occurs via a two state transition (Figure 5.16B), with a mid point of 2.57 M GdnHCl. The total free energy change associated with the transition is  $5.71 \pm 0.37 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.09 \pm 0.07 \text{ kcal mol}^{-1}\text{M}^{-1}$  (Table 5.1).

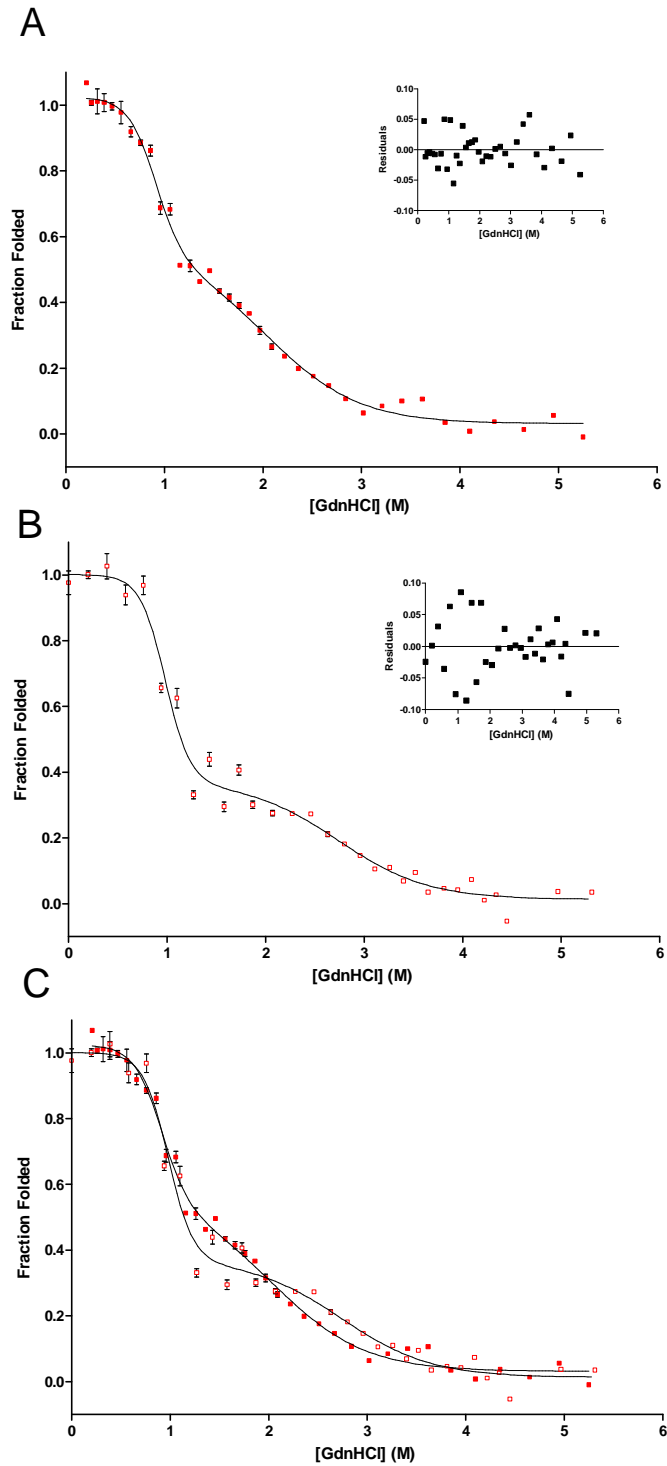
The folding of TTK<sup>W612F W622F W628F</sup> was monitored by tryptophan fluorescence as a comparison to the folding monitored by far-UV CD spectroscopy. The native protein has a slightly blue shifted  $\lambda_{\text{max}}$  of fluorescence compared to wild type TTK (Table 4.5). The  $\lambda_{\text{max}}$  of the W612F W622F W628F mutant is 332.5 nm. The  $\lambda_{\text{max}}$  of fluorescence is red shifted compared to the W612F W622F mutant. When the protein is unfolded by the addition of 6 M GdnHCl the  $\lambda_{\text{max}}$  of fluorescence red shifts to 349 nm. Unlike wild type TTK there is no increase in the intensity of the tryptophan fluorescence upon this unfolding (Figure 4.18C). Upon a stepwise unfolding of the native protein in GdnHCl there is a red shift in the  $\lambda_{\text{max}}$  towards the unfolded  $\lambda_{\text{max}}$  as well as an initial increase in the fluorescence intensity towards an apparent intermediate state. Following this there is a decrease in the fluorescence intensity and further red shift in the  $\lambda_{\text{max}}$  to the  $\lambda_{\text{max}}$  of the unfolded state (Figure 5.17).



**Figure 5.17:** Fluorescence spectra of TTK<sup>W612F W622F W628F</sup> under conditions of various denaturant concentration. Concentrations of denaturant for spectra in blue – 4.1 M, 4.65 M and 4.95 M; in green – 1.06 M, 1.16 M, 1.36 M, 1.56 M, 1.76 M, 1.86 M and 2.09 M; in orange – 2.51 M, 3.02 M, 3.41 M and 3.62 M; and in red – 0.39 M, 0.56 M and 0.76 M.

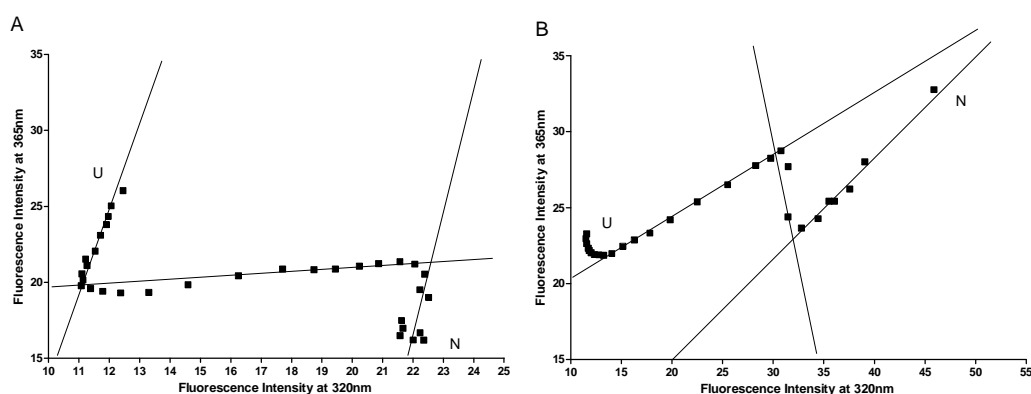
The refolding and unfolding transitions observed by tryptophan fluorescence were not well fitted to a two state folding model as was the case with wild type TTK and the W612F W622F mutant. The transitions were better fitted to a three state folding model, which describes an equilibrium between the unfolded, native and an intermediate state. Since the fluorescence intensity at the native  $\lambda_{\text{max}}$  rises to a value greater than the native intensity and the unfolded intensity (Figure 5.17) the  $\lambda_{\text{max}}$  at each denaturant concentration used was recorded, and the folding transitions plotted using this measure. The fraction folded was calculated by equation 5.1 and the resultant data fitted to equation 5.10.





**Figure 5.18:** Refolding and unfolding of TTK<sup>W612F W622F W628F</sup> induced by GdnHCl and monitored by change in  $\lambda_{\max}$  of tryptophan fluorescence. (A) Refolding of TTK<sup>W612F W622F W628F</sup>, Solid line shows fit to three state folding model (Equation 5.10). Insert shows residuals. (B) Unfolding of TTK<sup>W612F W622F W628F</sup>. Solid line shows fit to three state folding model (Equation 5.10). Insert shows residuals. (C) Overlay of refolding and unfolding transitions. Solid lines show three state folding model fits. Refolding shown as filled squares and unfolding as open squares.

Phase diagram analysis was used to support the use of three state models in describing the observed folding transitions. The phase diagram analysis of the refolding and unfolding of W612F W622F W628F appears to indicate that the folding transitions contain two intermediates in the folding transition (Figure 5.19). However, a closer examination of the fluorescence spectra indicates that one of these changes occurs only in the intensity of fluorescence, with no change in the  $\lambda_{\text{max}}$  occurring concurrently. Therefore, the correct model to use for the folding transitions observed by changes in the  $\lambda_{\text{max}}$  of fluorescence would be a three state folding model.



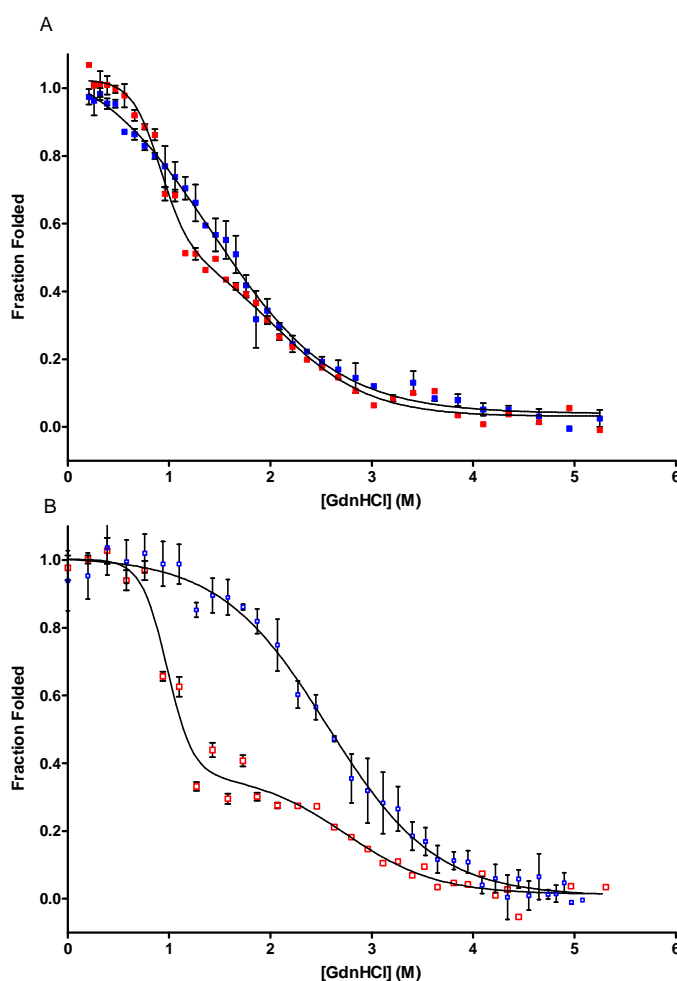
**Figure 5.19:** Folding phase diagrams for TTK<sup>W612F W622F W628F</sup>. (A) Phase diagram for the refolding of TTK<sup>W612F W622F W628F</sup>. (B) Phase diagram for the unfolding of TTK<sup>W612F W622F W628F</sup>. Positions of the unfolded and native states have been indicated. Phase diagrams constructed using fluorescence intensity at 320 nm and 365 nm.

The refolding of TTK<sup>W612F W622F W628F</sup> was successfully fitted to a three state folding model (Figure 5.18A). The first transition occurs between ~4 M GdnHCl and ~1.3 M GdnHCl. The mid point of this transition is at ~2.1 M GdnHCl. The free energy change,  $\Delta G_{U \rightarrow I}$ , associated with this transition is  $-2.52 \pm 0.83 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.24 \pm 0.29 \text{ kcal mol}^{-1} \text{ M}^{-1}$  (Table 5.2). The second transition occurs between ~1.3 M GdnHCl and ~0.4 M GdnHCl with a mid-point of ~0.9 M GdnHCl. The free energy change,  $\Delta G_{I \rightarrow N}$ , associated with this transition is  $-3.67 \pm 0.7 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $4.02 \pm 0.83 \text{ kcal mol}^{-1} \text{ M}^{-1}$  (Table 5.2)

The unfolding transition of TTK<sup>W612F W622F W628F</sup> shows the clear accumulation of an intermediate state in line with the folding transitions seen for the other TTK constructs (Figure 5.18B). The first transition occurs between native conditions and ~1.4 M GdnHCl, with a mid point of ~1 M. The free energy change,

$\Delta G_{N \rightarrow I}$ , associated with this transition is  $4.46 \pm 1.12 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $4.54 \pm 1.15 \text{ kcal mol}^{-1}\text{M}^{-1}$  (Table 5.2). The second transition occurs between  $\sim 1.4 \text{ M GdnHCl}$  and  $\sim 5 \text{ M GdnHCl}$  with a mid-point of  $\sim 2.8 \text{ M GdnHCl}$ . The free energy change,  $\Delta G_{I \rightarrow U}$ , associated with this transition is  $3.51 \pm 1.54 \text{ kcal mol}^{-1}$  with a corresponding  $m$  value of  $1.28 \pm 0.51 \text{ kcal mol}^{-1}\text{M}^{-1}$  (Table 5.2).

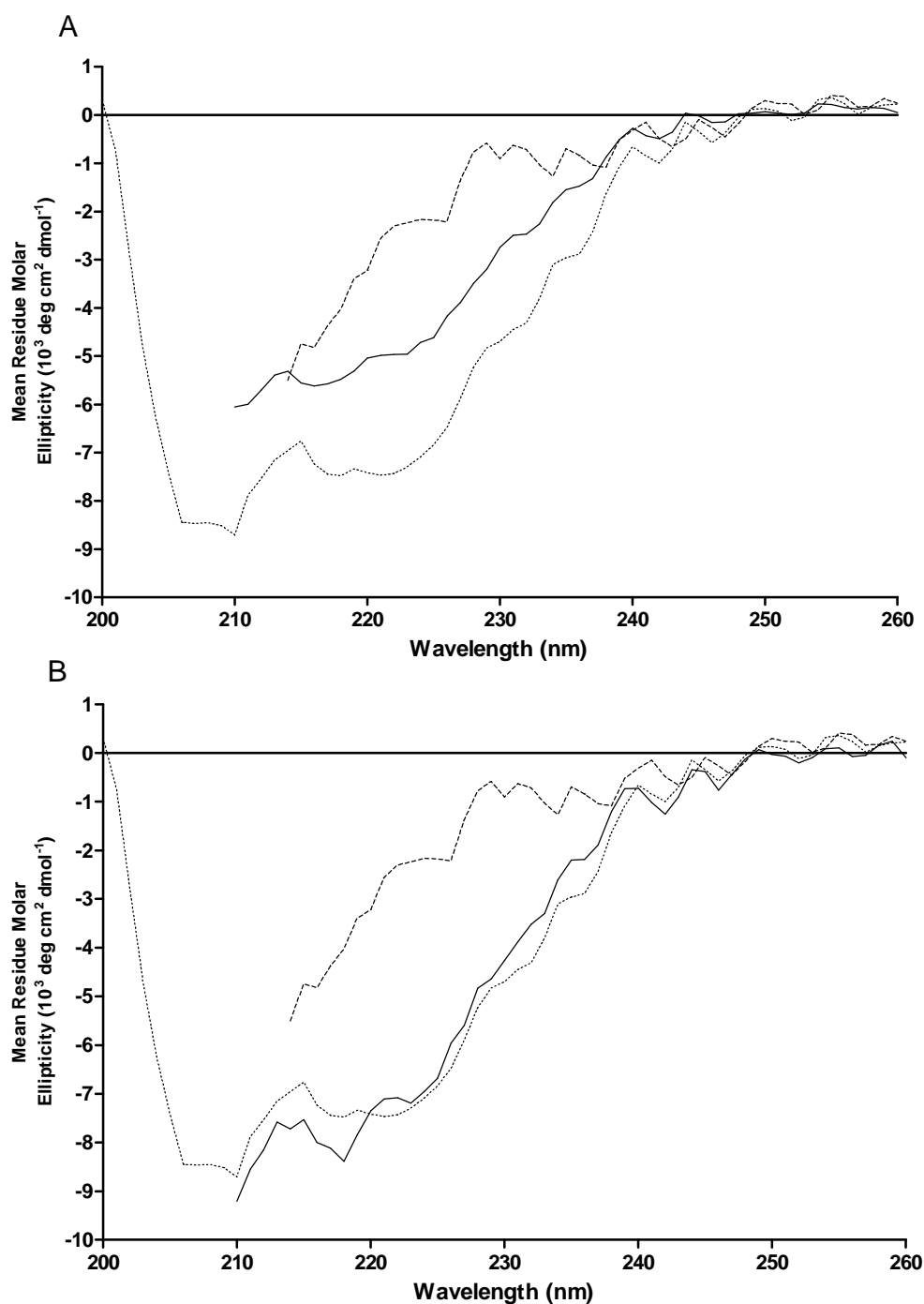
The comparison of the folding transitions observed by far-UV CD and tryptophan fluorescence sheds important light on the mechanisms of the folding of TTK<sup>W612F W622F W628F</sup>. This comparison is shown in Figure 5.20.



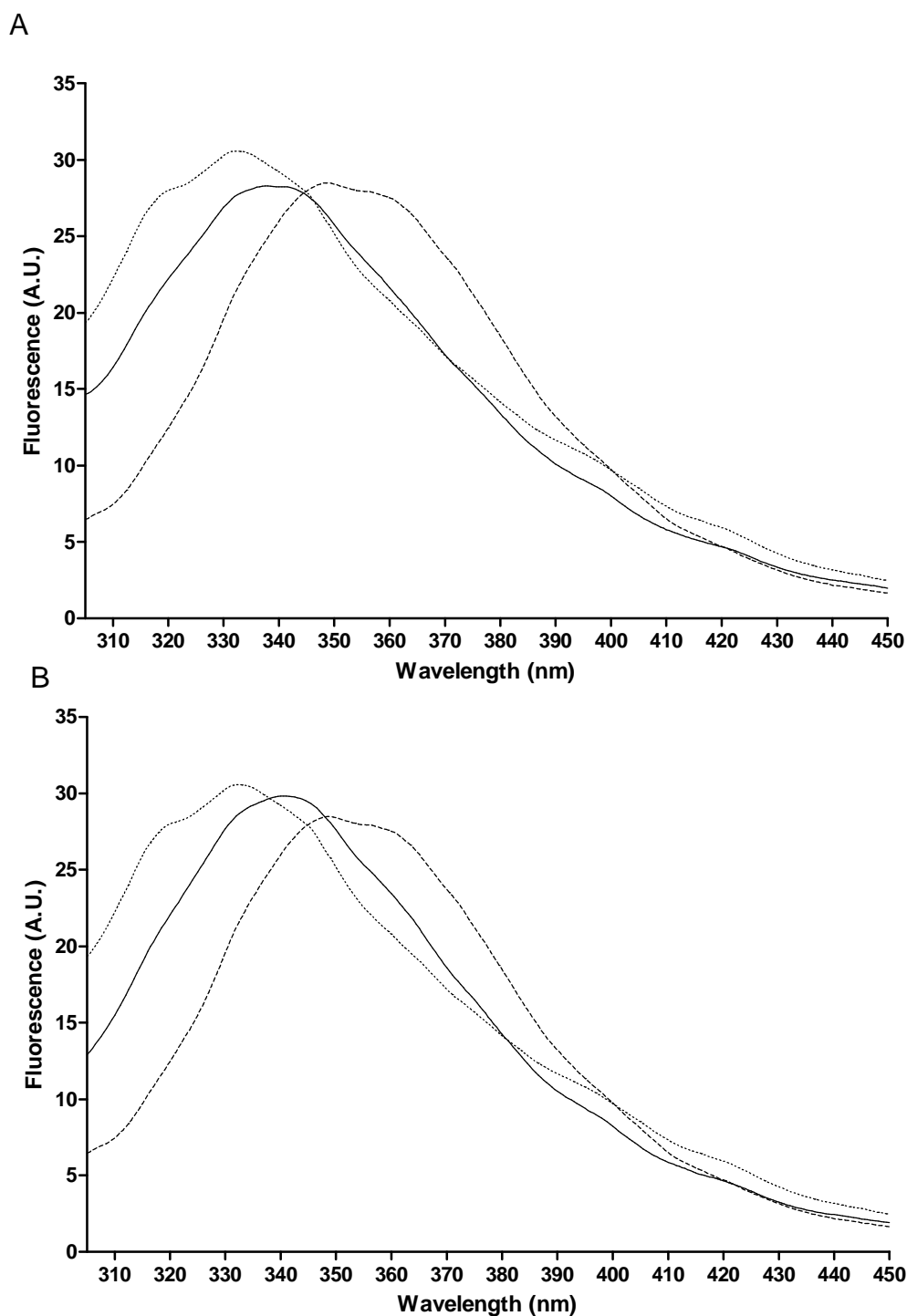
**Figure 5.20:** Overlay of Far-UV CD and tryptophan fluorescence monitoring of the folding of TTK<sup>W612F W622F W628F</sup>. (A) Refolding of TTK<sup>W612F W622F W628F</sup>. (B) Unfolding of TTK<sup>W612F W622F W628F</sup>. CD data in blue, fluorescence in red. Solid lines show 3 state or two state folding fits as described in Figures 5.17 and 5.19.

The comparison of the folding transitions observed by far-UV CD and tryptophan fluorescence shows that the recovery of tertiary structure is coincident

with the recovery of secondary structure upon the refolding of the W612F W622F W628F mutant of TTK. The folding intermediate observed by tryptophan fluorescence contains ~ 60 % of the native secondary structure. The overlay for the unfolding of TTK<sup>W612F W622F W628F</sup> indicates that the tertiary structure of the kinase domain is initially stable up to ~ 0.8 M GdnHCl. There is then a rapid loss of tertiary structure to the intermediate state. The secondary structure of the domain is more stable than the tertiary structure and is retained up to ~1 M GdnHCl. The high content of secondary structure in the partially unfolded intermediate indicates that a molten globule intermediate has been formed. Support for this characterisation of the intermediates formed can be gained by an examination of the CD spectra and tryptophan fluorescence spectra of the conditions under which the intermediate accumulates to its highest concentration. These points were identified using the phase diagrams previously constructed (Figure 5.19).



**Figure 5.21:** Far-UV CD spectra of TTK<sup>W612F W622F W628F</sup> in native, intermediate and unfolded states. (A) TTK<sup>W612F W622F W628F</sup> under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon refolding of TTK<sup>W612F W622F W628F</sup>. GdnHCl concentration of intermediate 1.26 M. (B) TTK<sup>W612F W622F W628F</sup> under native (dashed line), unfolded (dotted line) and intermediate conditions (solid line) upon unfolding of TTK<sup>W612F W622F W628F</sup>. GdnHCl concentration of intermediate 1.43 M.



**Figure 5.22:** Tryptophan fluorescence spectra of TTK<sup>W612F W622F W628F</sup> in native, intermediate and unfolded states. (A) TTK<sup>W612F W622F W628F</sup> under native (dotted line), unfolded (dashed line) and intermediate conditions (solid line) upon refolding of TTK<sup>W612F W622F W628F</sup>. GdnHCl concentration of intermediate 1.26 M. (B) TTK<sup>W612F W622F W628F</sup> under native (dotted line), unfolded (dashed line) and intermediate conditions (solid line) upon unfolding of TTK<sup>W612F W622F W628F</sup>. GdnHCl concentration of intermediate 1.43 M.

The CD spectra of TTK<sup>W612F W622F W628F</sup> at the denaturant concentration corresponding to the folding intermediate identified by phase diagram analysis (Figure 5.19A,B) show that on unfolding of the protein the intermediate retains a high degree of secondary structure (Figure 5.21B). However upon refolding the intermediate formed contains a reduced amount of secondary structure content. The CD spectra shows a CD intensity at 222 nm of ~ 53 % of the difference between the native and the unfolded state.

The tryptophan fluorescence spectra of the intermediate formed in the folding of wild-type TTK is similar upon unfolding (Figure 5.23B) and refolding (Figure 5.22A). The  $\lambda_{\text{max}}$  of the fluorescence spectra is blue shifted from the unfolded state, as shown in Figure 5.22. In addition the fluorescence intensity of the intermediate states is similar on both the unfolding and refolding transitions. This indicates that the intermediates may contain a similar degree of tertiary structure. This of particular interest since the unfolding and refolding transitions observed for the W612F W622F W628F were super-imposable as opposed to the unfolding and refolding transitions for the wild type protein and the W612F W622F mutant which were not super-imposable. This indicates that the folding of the region which is reported by the W718 residue is fully reversible.

The CD and fluorescence spectra support the identification of the intermediate on the unfolding pathway as a molten globule intermediate. The differences in the CD spectra of the intermediate formed on the refolding pathway demonstrate that this intermediate is not of the classical molten globule type.

The folding models which were applied to the folding transitions observed with wild type TTK and the W612F W622F and W612F W622F W628F mutants were used to determine the thermodynamic parameters associated with the observed folding and unfolding transitions. These parameters are summarised in Table 5.1 and Table 5.2.

**Table 5.1:** Thermodynamic parameters determined for the refolding and unfolding of wild type TTK and TTK tryptophan mutants followed by far-UV CD analysis. Parameters determined by fitting folding curves to a two state folding model.

Protein	Refolding			Unfolding		
	$C_m$ (M)	$\Delta G$ (kcal mol <sup>-1</sup> )	$m$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$C_m$ (M)	$\Delta G$ (kcal mol <sup>-1</sup> )	$m$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )
Wild Type TTK	1.95±0.04	-5.50±0.38	0.92±0.06	2.45±0.06	6.71±0.69	1.12±0.11
TTK <sup>W612F W622F</sup>	1.23±0.07	-4.70±0.36	0.94±0.07	2.71±0.04	5.93±0.44	1.13±0.08
TTK <sup>W612F W622F W628F</sup>	1.43±0.05	-4.81±0.30	0.96±0.06	2.57±0.04	5.71±0.37	1.09±0.07

**Table 5.2:** Thermodynamic parameters determined for the refolding and unfolding of wild type TTK and TTK tryptophan mutants followed by tryptophan fluorescence. Parameters determined by fitting folding curves to a three state folding model.

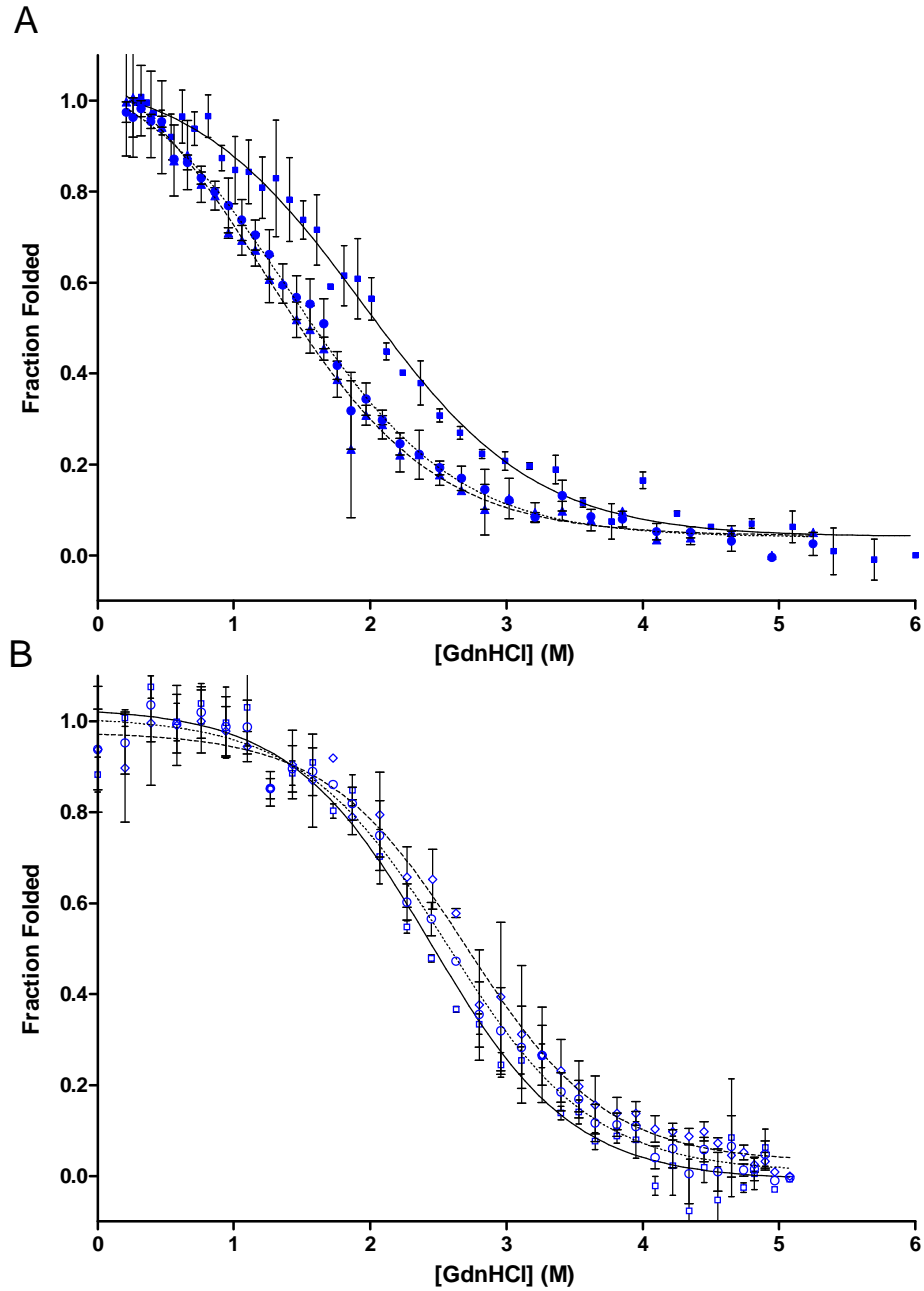
Protein	Refolding				Unfolding			
	$\Delta G_{I \rightarrow N}$ (kcal mol <sup>-1</sup> )	$m_{I \rightarrow N}$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta G_{U \rightarrow I}$ (kcal mol <sup>-1</sup> )	$m_{U \rightarrow I}$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta G_{N \rightarrow I}$ (kcal mol <sup>-1</sup> )	$m_{N \rightarrow I}$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )	$\Delta G_{I \rightarrow U}$ (kcal mol <sup>-1</sup> )	$m_{I \rightarrow U}$ (kcal mol <sup>-1</sup> M <sup>-1</sup> )
Wild Type TTK	-1.08±1.04	1.97±1.26	-3.24±0.78	1.46±0.27	5.38±0.94	6.15±1.09	3.33±0.54	1.41±0.20
TTK <sup>W612F W622F</sup>	-1.97±0.31	2.60±0.40	-4.15±1.10	2.04±0.45	3.86±0.20	4.22±0.22	3.90±0.52	1.50±0.18
TTK <sup>W612F W622F W628F</sup>	-3.67±0.7	4.02±0.83	-2.52±0.83	1.24±0.29	4.46±1.12	4.54±1.15	3.51±1.54	1.28±0.51



#### ***5.3.4 Comparison of Unfolding and Refolding of TTK***

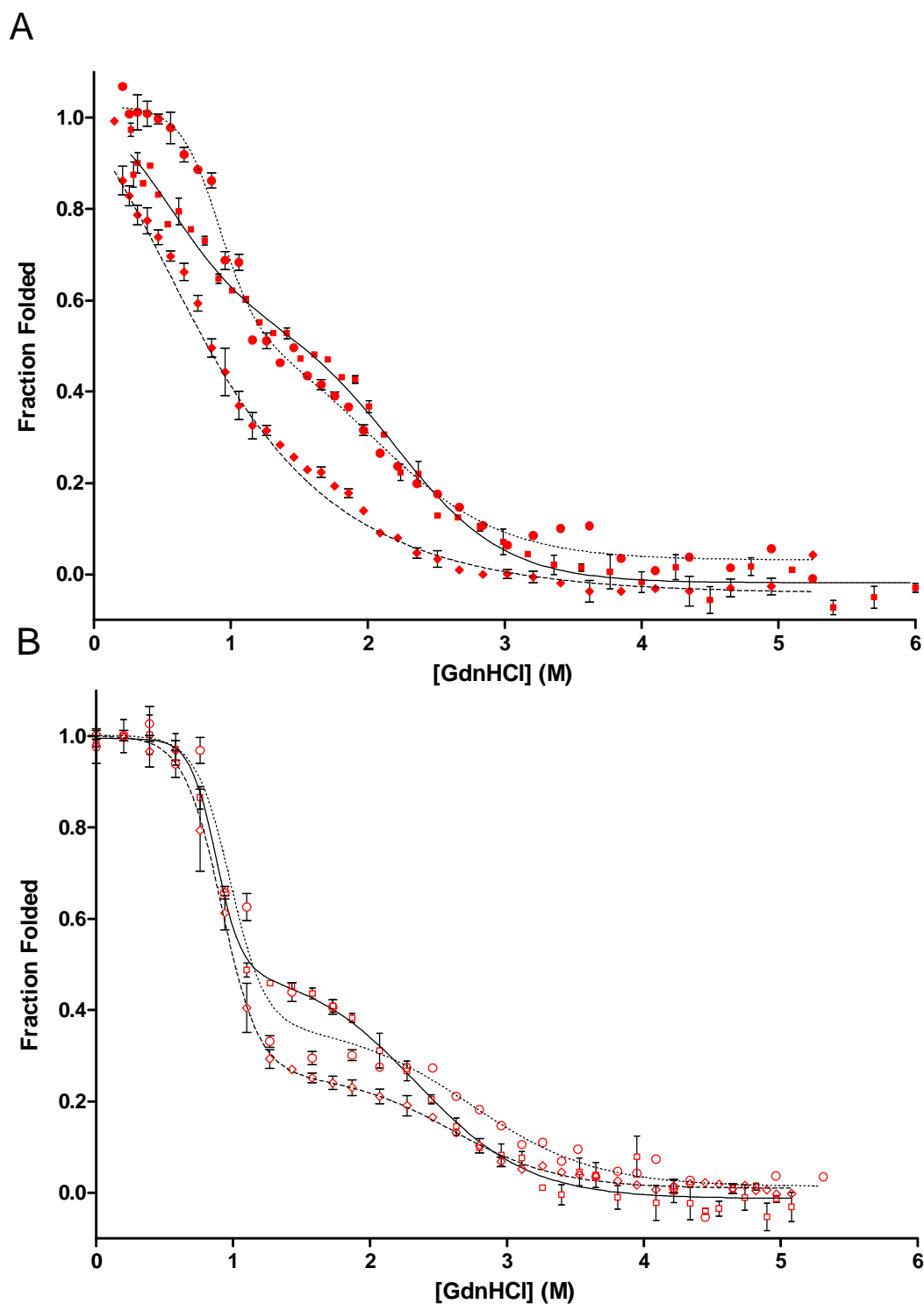
The folding transitions of wild type TTK and two tryptophan to phenylalanine substitution mutants have been observed by far-UV CD and tryptophan fluorescence. To determine if the mutations performed have affected the folding of the TTK kinase domain the folding transitions for the wild type and mutants can be compared. The comparison of the far-UV CD observed folding transitions is shown in Figure 5.23.

The unfolding transitions of TTK and the tryptophan mutants studied by far-UV CD are very similar (Figure 5.23B). The transitions all are well described by a two state folding model and show a broad transition between the native and the unfolded state. There is some variability in the fit of the folding transitions to the two state folding model, but this is within the experimental variation.



**Figure 5.23:** Overlay of the folding transitions of wild-type TTK, TTK<sup>W612F W622F</sup> and TTK<sup>W612F W622F W628F</sup> measured by far-UV CD. (A) Refolding transitions of ■ – wild type TTK, ▲ – TTK<sup>W612F W622F</sup>, ● – TTK<sup>W612F W622F W628F</sup>. Fits to two state folding models, as described previously shown by solid line for wild type, dashed line for W612F W622F and dotted line for W612F W622F W628F. Error bars show standard error of three experiments. (B) Unfolding transition of □ wild type TTK, ◇ - TTK<sup>W612F</sup>, ○ - TTK<sup>W612F W622F W628F</sup>. Lines as for panel A. Error bars show standard error of 3 experiments.

The refolding transitions show more variability, with the refolding transition of wild Type TTK being slightly different from the transitions of the W612F W622F and W612F W622F W628F mutants (Figure 5.23A). The refolding transitions of the tryptophan mutants of TTK closely overlay. The close overlay of the unfolding transitions of TTK and the TTK tryptophan mutants, and the similarity of the refolding transitions of these proteins indicates that the tryptophan to phenylalanine substitutions performed have not altered the folding properties of the TTK kinase domain.



**Figure 5.24:** Overlay of the folding transitions of wild-type TTK, TTK<sup>W612F W622F</sup> and TTK<sup>W612F W622F W628F</sup> measured by tryptophan fluorescence. (A) Refolding transitions of ■ – wild type TTK, ▲ – TTK<sup>W612F W622F</sup>, ● – TTK<sup>W612F W622F W628F</sup>. Fits to two state folding models, as described previously shown by solid line for wild type, dashed line for W612F W622F and dotted line for W612F W622F W628F. Error bars show standard error of three experiments. (B) Unfolding transition of □ wild type TTK, ◇ – TTK<sup>W612F W622F</sup>, ○ – TTK<sup>W612F W622F W628F</sup>. Lines as for panel A. Error bars show standard error of 3 experiments.

The comparison of the folding transitions observed by tryptophan fluorescence are expected to show difference between the three different constructs of TTK which were studied. The unfolding transitions of the studied constructs show very similar transitions (Figure 5.24B) despite the differences in the fluorescence spectra of the constructs (Figure 4.17). The native state is stable up to a GdnHCl concentration of ~ 0.8 M and then there is a rapid transition to an intermediate state. The difference in the apparent fraction folded of the native state can be traced to the changes in the  $\lambda_{\max}$  of fluorescence of the native state. The most blue shifted native state, that of the W612F W622F mutant gives rise to the lowest fraction folded value for the intermediate. Likewise, the most red shifted  $\lambda_{\max}$  of native protein, that of the wild type, gives rise to the highest fraction folded for the intermediate.

The refolding transitions of TTK are strikingly different for the different mutants studied (Figure 5.24A). This indicates that the different mutants are reporting on different folding. The mutants containing more than one tryptophan show broader transitions, with less clearly defined intermediate states than is seen with the single tryptophan mutant. The intermediates can be seen, however, to accumulate at approximately the same GdnHCl concentration. This indicates that the same type of intermediate is being reported on by the different mutants studied.

## ***5.4 Discussion***

The folding of wild type TTK was monitored by far-UV circular dichroism and tryptophan fluorescence. The folding transition, monitored by far-UV CD was not well fitted to a three state folding model, but was instead fitted to a two state folding model (Figure 5.2). The unfolding and refolding curves observed for wild type TTK were nearly super-imposable (Figure 5.2).

Phase diagram analysis performed on the folding of wild type TTK, as monitored by tryptophan fluorescence, indicated that the folding transitions were not two state transitions (Figure 5.6). The unfolding and refolding transitions were, therefore, fitted to three state folding models. The unfolding and refolding curves were not super-imposable however. The intermediate state recorded in both the refolding and unfolding accumulates significantly upon unfolding but does not

accumulate significantly upon refolding (Figure 5.4). In addition the native state is quite stable upon unfolding, but a similar stability of the native state is not seen upon refolding (Figure 5.4).

The intermediate formed on the unfolding of wild type TTK is a molten globule type intermediate (Figure 5.6B). The CD data indicate that upon refolding, the recovery of the full CD intensity is delayed, and significant amounts of the secondary structure are still to be formed in the intermediate.

The folding of TTK<sup>W612F W622F</sup>, when examined via far-UV CD was determined to proceed via a two state transition. The refolding and unfolding curves are however, not super-imposable (Figure 5.9). However the unfolding and refolding curves are similar to their wild type TTK counterparts. This indicates that the unfolding and refolding of the TTK kinase domain has not been altered by the tryptophan to phenylalanine substitutions performed.

The folding of TTK<sup>W612F W622F</sup> was also determined to follow a three state transition when examined by tryptophan fluorescence (Figure 5.12). The folding curves obtained are not super-imposable (Figure 5.11). The accumulation of the intermediate upon unfolding appears to occur at a different fraction folded than for wild type TTK. However, examination of the fluorescence spectra of the protein at the GdnHCl concentrations associated with the intermediate state reveals similar  $\lambda_{\max}$  to wild type TTK and TTK<sup>W612F W622F</sup>. The TTK<sup>W612F W622F</sup> mutant shows similar differences in unfolding and refolding that were seen with wild type TTK. The native state is stable upon unfolding to ~0.2 M GdnHCl (Figure 5.11).

As observed for wild type TTK the comparison of the analysis of the unfolding of TTK<sup>W612F W622F</sup> by far-UV CD and fluorescence indicates that when the intermediate is formed, as identified by tryptophan fluorescence, the CD spectra has been mostly recovered (Figure 5.13). This indicates that the intermediate formed is of the molten globule type. The intermediate formed upon refolding does not share this feature. At the point of the accumulation of the intermediate, approximately 53 % of the secondary structure has been formed. The recovery of the CD spectra and the shift in the  $\lambda_{\max}$  are nearly co-incident, indicating a different intermediate being formed (Figure 5.13).

The single tryptophan mutant, TTK<sup>W612F W622F W628F</sup> was created to study the conserved tryptophan residue, W718, allowing comparison with previous studies on the folding of p38 $\alpha$ . The unfolding and refolding of TTK<sup>W612F W622F W628F</sup> was

monitored by far-UV CD and was determined to undergo a two state transition (Figure 5.16). The unfolding and refolding were not super-imposable. However the unfolding transition was super-imposable with the unfolding transitions of wild type TTK and TTK<sup>W612F W622F</sup> (Figure 5.23). The refolding transition was also super-imposable with the other mutants. This indicates that the tryptophan to phenylalanine mutations did not alter the unfolding and refolding of the TTK kinase domain.

Phase diagram analysis of the folding of TTK<sup>W612F W622F W628F</sup> identified a four state transition (Figure 5.19); however inspection of the spectra identified that the final transition involved a change in fluorescence intensity only with no change in  $\lambda_{\text{max}}$ . Since the measure of folding was  $\lambda_{\text{max}}$  shift, the folding transitions were fitted to a three state folding model. The intermediate state recorded in both the refolding and unfolding accumulates significantly upon unfolding but does not accumulate significantly upon refolding. In addition the native state is quite stable upon unfolding, remaining structured up to ~0.4 M GdnHCl (Figure 5.18). The native state reached upon refolding is also stable below ~0.4 M GdnHCl, a feature not seen in the other mutants studied (Figure 5.24A).

The overlay of far-UV CD data and the tryptophan fluorescence of the single tryptophan mutant identifies that a molten globule intermediate forms on the unfolding of the single tryptophan mutant (Figure 5.20). However upon refolding a much lower amount of secondary structure is formed at the point at which the intermediate accumulates.

The folding studies performed on wild type TTK and the tryptophan mutants produced present a coherent story. The unfolding of the secondary structure of TTK, as measured by far-UV CD, proceeds via a two state transition in which no intermediate accumulates significantly (Figure 5.23). The tryptophan fluorescence data identifies an intermediate state which is formed when the secondary structure of the C-terminal lobe is present (Figure 5.24). This shows a classical molten globule intermediate. Techniques such as ANS fluorescence or H/D exchange NMR could be used to further define this intermediate state. The unfolding of this intermediate is completed as secondary structure of the protein unravels and the protein adopts a random coil conformation. The comparison of the spectra of the wild type TTK and the W612F W622F mutant shows that the intermediate has a similar spectrum in each mutant, with similar  $\lambda_{\text{max}}$  of emission, indicating that the exposure to solvent of the

tryptophan residues present is similar in the two proteins (Figure 5.8 and 5.15) . The structure of the intermediate state is likely to be similar in the two mutants.

A dissimilar intermediate is formed upon the refolding of the protein. The recovery of the native CD spectra and the recovery of the native  $\lambda_{\text{max}}$  are co-incident. Approximately 60% of the secondary structure is present in the intermediate state (Figure 5.23 and 5.24).

The hysteresis observed between the refolding and unfolding of wild type TTK and TTK tryptophan mutants is an interesting feature of the folding of the TTK kinase domain (Figure 5.2, 5.9 and 5.16). The loss and recovery of the CD signal upon the folding of wild type TTK is a simpler problem than the differences observed in the folding as measured by fluorescence. Engel and Bächinger (2000) observed the folding and unfolding of collagen III and attempted to explain the hysteresis between the unfolding and refolding of this protein. They observed that the unfolding transition maintained native structure to a higher denaturant concentration than the refolding spectra showed native structure, and that the unfolding transition occurred across a narrower range of denaturant. They attempted to fit these data to a simple kinetic model for the folding of collagen III but were unable to do so. Instead they posited that the high hysteresis and sharp transition were due to two factors; a cooperative equilibrium transition, and a slow annealing step. The slow annealing step is due to the requirement of cis-trans proline isomerisation in the folding of collagen.

The application of this result to the TTK kinase domain reveals a possible explanation for the hysteresis. The slow phase of folding was eliminated by an examination of the protein at various denaturant concentrations after 15 minute incubations and after overnight incubations. No difference was observed in the fluorescence or CD spectra collected after the incubations, indicating that no slow changes were occurring in the folding process. In addition, of the 14 proline residues in the construct of TTK crystallised to generate the structure shown in figure 4.1, 12 are visible in the structure and all adopt the trans configuration. No additional proline residues are present in the construct of TTK used for this study. Rather we must assume that the hysteresis is due to co-operative effects in the folding of the kinase domain of TTK. We cannot, presently however, speculate as to whether these co-operative effects are inter chain or intra chain effects. An insight into this could be gained by a repetition of the folding studies at a lower protein concentration to decrease inter chain effects. Further work would also need to be done to examine the



folding of the C-terminal lobe and the formation of the interface between the lobes that forms the substrate binding site. The same effect is likely to be responsible for the differences observed in the tryptophan fluorescence of the unfolded and folded proteins.

Tryptophan 718 shows a key importance in the folding of the TTK kinase domain. It is required for correct folding, adopting a structurally and sequence conserved position with a conserved structural environment. In addition the refolding of the single tryptophan mutant bears a resemblance to the refolding of the p38 $\alpha$  single tryptophan mutant. The folding of this mutant clearly proceeds via a folding intermediate which has the character of a molten globule intermediate upon unfolding, and a more mixed character upon refolding.

To fully understand the folding of the TTK kinase domain it would be necessary to examine the folding at a second, lower protein concentration to investigate the existence of inter-chain cooperativity. In addition FRET based studies would be vital to the examination of the folding of the N-terminal lobe, and the formation of the inter-lobe interface which forms the active site. Analysis of the kinetics of folding and unfolding would complete the picture of the folding of the TTK kinase domain.

## Chapter 6. General Discussion

### 6.1. *The Refolding Problem and Shared Protein Folds*

The vast increase in the number of solved protein structures available since the first high-resolution protein structure determined in 1959 (Bodo *et al.*, 1959), the number of folding studies performed on proteins to date and the importance of descriptors of these folds, such as contact order, in describing the folding of proteins, have raised questions concerning the importance of these common folds in determining the folding of proteins. The correlation of contact order with the rate of folding of small proteins suggests that the topology of the native state may be more important than the amino acid sequence in determining the folding of proteins (Plaxco *et al.*, 1998). From this observation some studies have been performed into whether proteins of similar folds follow similar pathways of folding and share similar structures for any transition states present in their folding. The results of these studies have been mixed, and no clear yes/no answer has been produced. These studies have largely been performed with small proteins and the folding of larger common folds is still an interesting area of study (Zarrine-Afsar *et al.*, 2005)

The kinase fold represents a valuable area of study to address these questions. The fold is larger and more complex than the simple folds commonly studied, with two lobes which interact to form the active site of the kinase domain. The kinase domain is of key pharmaceutical interest and the potential understanding of the folding of the kinase domain gained through studies on its folding could prove valuable in obtaining soluble recombinant kinase domains for drug development studies and structural studies on the kinase domain.

To address the problem of common folding pathways in the highly conserved kinase fold, and to explore the possibilities of common methods of refolding protein kinases from insoluble aggregates produced on overexpression of soluble protein a refolding screen for protein kinases was created using the protein kinase p38 $\alpha$  as a model protein kinase which was known to refold under equilibrium conditions (Davies, 2004). The screening system was designed and tested using this protein kinase, and the ability of the refolding screen to identify correctly refolded protein kinase (Figures 2.12 and 2.14) was demonstrated by examination of the secondary structure content of the refolded protein and the native protein, and by examination

of the inhibition of the binding of the refolded protein by a high affinity p38 $\alpha$  inhibitor, SB202190 (Figure 2.13). Following this development of the refolding screen, the screen was applied to additional protein kinases, to demonstrate the general suitability of the screen in refolding protein kinases, and to gather data on the commonalities in the refolding of the selected protein kinases.

## **6.2. The Refolding of Five Protein kinases Through a Broad Refolding Screen**

To allow the identification of conditions leading to the refolding of protein kinases, a refolding screen was designed for protein kinases, which consisted of a number of chemical refolding additives and a set of associated analytical methods which test for the extent of the refolding of the test protein kinase. p38 $\alpha$  was used initially as a model protein kinase to test the screen and to develop the readout methods for the screen. High numbers of conditions in the refolding screen resulted in a measurable recovery of refolded protein (Figure 2.9) with recoveries of refolded protein of 10% under the control conditions at high pH. Several interesting features of the refolding of p38 $\alpha$  were noted from the results of the refolding screens performed on p38 $\alpha$ . Firstly, the recoveries of refolded protein are substantially different depending on the source of the protein which was refolded. When the protein to be refolded was sourced from protein produced in a soluble, folded form in *E. coli* the recoveries of refolded protein were approximately two times greater than those achieved when refolding p38 $\alpha$  produced in inclusion bodies. It has been noted in studies on the scaling of tumbling speed with chain length and residual structure of denatured protein that denatured protein adopts a more compact ensemble of structures than was expected from initial studies on the tumbling of denatured protein performed by Tanford *et al.* (1967). In addition residual native like structure has been observed in the denatured state of Protein L (Yi *et al.*, 2000). The differences in the refolding of p38 $\alpha$  between these different sources of refolded protein are likely to be due to differences in the residual structure present in the denatured state of p38 $\alpha$ , as a measurement of the multimeric nature of the denatured state showed both sources of protein to contain only monomeric protein of similar distributions of hydrodynamic radius (Figure 2.2). The differences between the refolding yield could also be due to a requirement for cis-trans proline isomerisation in the folding of p38 $\alpha$ , however an examination of the crystal structure to p38 $\alpha$  does

not reveal any proline residues in a cis orientation, and the two preparations of protein were stored in a denatured state for at least 24 hours prior to use.

The refolding yields for p38 $\alpha$  were shown to be highly dependent on the pH under which refolding was performed (Figure 2.9). The stability of the native state of p38 $\alpha$  under different pH and buffer conditions was tested (Figure 2.11) and the thermal stability of p38 $\alpha$  was shown to be substantially lowered at low pH, with the mid point of the thermal melting transition decreasing by 12 °C between pH 7.0 and pH 5.8. This result suggests that a key determinant of the yield of refolding is the stability of the native state. An increase in the stability of the native state will deepen the folding funnel and will promote faster folding. In addition, the protein will populate any partially folded intermediate states less than if the difference between the stability of the native state and any intermediates is lower. This reduces the propensity and possibility for the partially folded intermediates to aggregate, leading to high recoveries of refolding.

The successful development of the screen using p38 $\alpha$ , and the demonstration of the screen readouts identifying correctly refolded p38 $\alpha$ , allowed the screen to be applied to protein kinases other than p38 $\alpha$ . The refolding of four additional protein kinases was tested using the refolding screen, and the results of the refolding of these kinase domains suggest that there are few commonalities in the refolding of protein kinases from inclusion bodies. The effect of pH upon the refolding of p38 $\alpha$  was not consistently observed in the refolding of the four additional kinases. In the case of KIS and PhK there was a significant difference observed between the average recoveries of refolded protein obtained at low pH and at high pH. With AKT2 and TTK however, no significant difference was observed in the average soluble protein recoveries between high pH and low pH conditions. The thermal melting of TTK was also tested in the same fashion as p38 $\alpha$  to further examine the role of the stability of the native state in the refolding yields obtained (Figure 3.22). The pattern of changes in the mid-point of the thermal melting transition across the different pH and buffer conditions tested was different from the pattern observed with p38 $\alpha$  (Table 2.2 and 3.2). The analysis showed little change in the  $T_m$  for the thermal melting across the pH and buffer range. This matches the result obtained with the average recoveries of soluble protein in the screen, which does not vary significantly between high and low pH conditions. However, if the melting temperatures are compared to the recoveries of monomeric protein obtained, there is not a clear

relationship. p38 $\alpha$  melts at a lower temperature than TTK, ~44 °C vs ~48 °C. However the pattern in the recovery of refolded protein is opposite to this pattern, with p38 $\alpha$  refolding yields being greater than those obtained with TTK, with p38 $\alpha$  giving recoveries of ~10 %, whereas TTK recoveries are in the range of 4 %. It can be seen that the stability of the native state is correlated with the relative refolding yields obtained between different pH and buffer conditions, but the absolute stability of the native state does not appear to be related to the yield of refolding when different proteins are compared.

The large number of different chemical refolding additives present in the refolding screen allows for other potential similarities in the refolding of the kinase domain to be identified. The conditions which give rise to high yields of refolded protein, when compared to the control conditions at each pH, should highlight similarities in the refolding pathways of the proteins, since the formation of similar intermediates and the existence of similar traps for the refolding of the different kinases should be highlighted by commonalities in the effective additives. The effect of the additives used on the refolding of the tested protein kinases was mixed. Some commonalities were observed between the additives that were effective in supporting the refolding of two or more protein kinases, however, no single additive supports the refolding of all five protein kinases. Instead the kinases which were refolded behave differently. The kinases which are most similar in their pattern of effective additives are KIS, p38 $\alpha$  and PhK; however, even among these three kinases there are important differences in the effectiveness of the refolding additives (Figures 3.13 and 3.17). The pattern of the effect of the tested additives on the refolding of the five tested protein kinases, and the pattern of the average soluble protein recovery at different pHs do not support a common refolding conditions for the protein kinases tested.

### ***6.3. The Refolding of Five Protein Kinases through a Fractional Factorial Screen***

The initial refolding screen design that had been developed with p38 $\alpha$  and tested with the four additional members of the kinase panel only tests the effect of single additives over control conditions at three different pHs. In addition, the design of the screen does not allow easy analysis of the screen using statistical tests. To

address this shortcoming in the screen, a new screen design was created that allowed the testing of the combination of additives and the robust statistical analysis of the screen results. The number of additives used was limited to six. The screen created was a fractional screen, since this allowed the size of the screen to be decreased while limiting the effect of the size reduction on the coverage of the combinations of the refolding additives used in the shortened screen.

When the five kinases were refolded using the fractional factorial screen designed, the refolding of protein kinases was improved by combining additives in the refolding buffers used. All five protein kinases showed an improvement in the refolding yields obtained in the screen format by combining different refolding additives (Table 3.3).

The monomeric protein recovery of the refolded proteins was analysed using ANOVA, and the significant, positive factors and combinations of factors from the screen identified. The analysis of the refolding yields obtained in the fractional factorial refolding screen shows a similar comparison between the refolding additives to that seen in the initial screen design. It was observed that no one refolding additive or combinations of additives was found to lead to a significant increase in the refolding yield of the tested protein kinases. The single additive, NV-10 was found to be a significant positive factor in the refolding of AKT2, KIS and TTK. However, it was not a positive factor in the refolding of p38 $\alpha$  and PhK (Table 3.9). Similarly some combinations of additives have a positive effect on the refolding of some kinases and a negative effect on the refolding of other protein kinases. Overall, from the analysis of the positive and negative factors on the refolding of the five tested protein kinases, it is clear that no single additive or combination of additives is effective in assisting the refolding of the five protein kinases. This undermines the concept of a common kinase folding pathway.

#### ***6.4. The Role of the Core Tryptophan of the Kinase Domain***

Commonalities in the folding of the kinase domain were further examined by a study of the equilibrium folding of a kinase domain. The kinase domain of the dual specificity kinase TTK was chosen for this study. To allow closer study of the folding of the kinase domain of TTK a number of mutants of TTK were created

which substituted tryptophans in the kinase domain by phenylalanines. This was performed to allow the study of the kinase domain with single tryptophan probes.

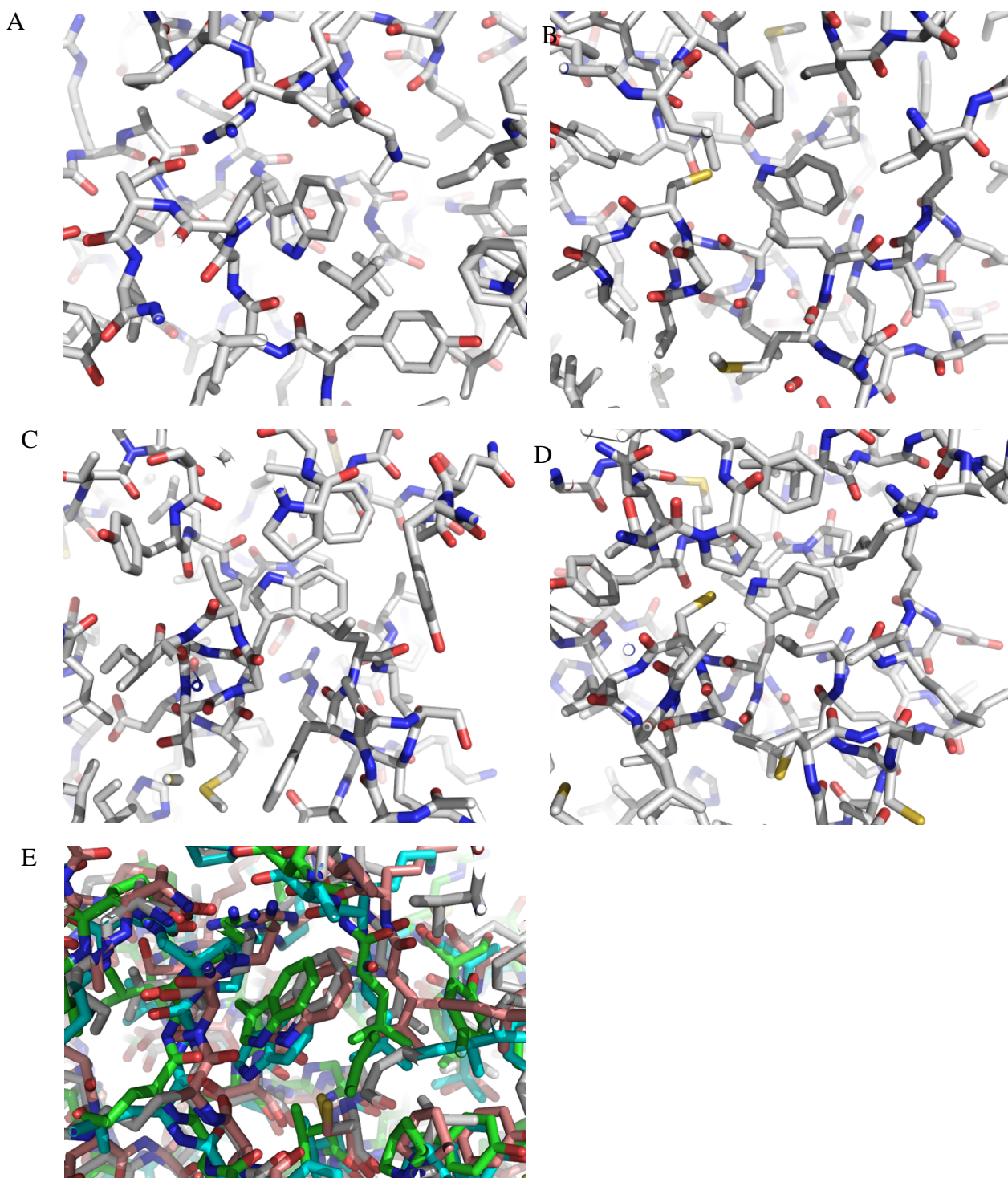
The mutants with tryptophan to phenylalanine substitutions were created by site directed mutagenesis, and expressed and purified from *E. coli* after co-expression with  $\lambda$ -phosphatase (Sections 4.3.2 through 4.3.4). The far-UV CD spectra (Figure 4.14) and the calculated secondary structure content (Table 4.3) of the mutants did not indicate that substantial changes in the fold had occurred, and the  $T_m$ s of wild type TTK and the mutants did not indicate a destabilisation of the tertiary structure (Table 4.4), which would be indicated by a decrease in the  $T_m$ . The incorporation of the desired mutations was demonstrated in the purified protein by ESI-MS analysis (Table 4.2).

The replacement of the core tryptophan of the kinase domain, W718 in TTK, was attempted in the background of the wild type TTK kinase domain. This mutant could not be produced in a soluble form in *E. coli*. This result is similar to that achieved with p38 $\alpha$ , in which the core tryptophan of p38 $\alpha$ , W207, could not be replaced without preventing the folding of p38 $\alpha$  (Davies, 2004). In the study of the folding of p38 $\alpha$ , the replacement of W207 with residues other than phenylalanine, specifically tyrosine, histidine and lysine, did not result in the production of soluble protein (Davies, 2004). The examination of the local environment of W207 in the crystal structure of p38 $\alpha$  indicates the importance of the tryptophan residue in the hydrophobic core of the protein, with the NH group of the indole ring of the tryptophan protected from the hydrophobic core by a hydrogen bond to an ordered water molecule (Davies, 2004).

The core tryptophan of the TTK kinase domain shows a very similar structure to that of p38 $\alpha$ . In the core of the C-terminal lobe of the TTK kinase domain the core tryptophan residue forms similar hydrogen bonds to the core tryptophan residue in p38 $\alpha$ . The local environment of the tryptophan is also very similar, with the local residues being very closely aligned in the two structures (Figure 4.3).

If the available structures of the other protein kinases that were studied using the refolding screens performed in chapters 3 and 4 are examined, it can be seen that the core tryptophan of the kinase domain, which is conserved in the kinome (Davies, 2004) is structurally conserved in the studied kinase. The local environment of the conserved residue is also surprisingly conserved, with the exception of AKT2, which

includes a second tryptophan residue in the core in the place of the varied residues present in other kinases (Figure 4.3, Figure 3.1).



**Figure 6.1:** Local environments of the core tryptophan of the kinase domain in the structures of the kinases of the kinase panel. The conserved core tryptophan is placed at the centre of each frame. (A) AKT2, (B) p38 $\alpha$ , (C) PhK, (D) TTK, (E) Alignment of the structures of the 4 kinase C-terminal lobe cores showing close alignment of kinase core residues, AKT2 – blue, PhK – green, p38 $\alpha$  – grey, TTK - pink.



The conservation of the core tryptophan residue of the kinase domain suggests that there may be similarities in the folding of the kinase domains of p38 $\alpha$  and TTK. Both kinase domains require the presence of the tryptophan residue for correct folding and the environment of the tryptophan is also conserved, suggesting that the role of this tryptophan previously identified in the folding of p38 $\alpha$  may also be found in TTK. Further analysis of the folding of the TTK kinase domain would examine this possibility.

### ***6.5. The Formation and Nature of Intermediate States on the Folding Pathway of TTK and p38 $\alpha$ .***

The equilibrium folding pathway of TTK has been characterised using the folding probes of far-UV circular dichroism and tryptophan fluorescence. The monitoring of the unfolding and refolding of TTK by far-UV CD showed similar folding transitions for wild type TTK and the tryptophan mutants studied, further indicating that the substitutions performed have not affected the fold or the folding of the kinase domain (Figure 5.23). The folding transitions observed for refolding and unfolding are not superimposable, indicating that the unfolding and refolding of TTK occurs via different pathways. Both transitions are broad and difficult to accurately fit to three state folding models, but do fit well to two state folding models. Visual inspection of the folding transitions does not indicate the presence of any folding intermediates (Figure 5.2B, 5.9B and 5.16B). A native like CD spectra is maintained in low concentrations of denaturant, up to a GdnHCl concentration of approximately 1 M upon the unfolding of TTK. However, when refolding TTK, a native like CD spectra is not regained until a GdnHCl concentration of approximately 0.5 M (Figure 5.2A, 5.9A, 5.16A). In addition, there is hysteresis between the unfolding and refolding curves observed (Figures 5.2C, 5.9C, 5.16C).

In the equilibrium folding of p38 $\alpha$ , a substantially different behaviour of the kinase domain on refolding and unfolding was observed. The folding transitions were broad and were determined to not be two state transitions by visual inspection, which indicated a plateau at ~2 M GdnHCl, suggesting that an intermediate was present. It was not possible, however to fit the transitions to three state folding models, as the intermediate did not accumulate. Visual inspection of the suspected intermediate suggested that it contained ~70 % of the native secondary structure, or

that it contained ~23%  $\alpha$ -helix (Davies, 2004). This behaviour of p38 $\alpha$  presents a clear difference from the folding observed with TTK.

The folding of the kinase domain of TTK was also monitored by tryptophan fluorescence. The four tryptophan residues of the TTK kinase domain are found in the C-terminal lobe (Figure 4.1), no tryptophans are found in the inter-lobe interface of the kinase domain of TTK nor in the N-terminal lobe. It was also observed that, according to the crystal structure, two of the native tryptophans appear to be solvent exposed in the native state (Figure 4.2). Two particular tryptophan to phenylalanine mutants were created to explore the folding of TTK. The first substituted the solvent exposed tryptophans, W612 and W622, by phenylalanine to allow the study of the folding of the hydrophobic core of TTK. A single tryptophan mutant, containing only the core tryptophan, W718, was also created.

The folding of the kinase domain of TTK monitored by tryptophan fluorescence shows clear three state transitions, with an intermediate accumulating at ~1.5 M GdnHCl. The folding transitions observed by tryptophan fluorescence do not coincide with the folding transitions observed by far-UV CD. This occurs because these folding probes are monitoring the formation of tertiary and secondary structure respectively, and the lack of coincidence of the two folding curves indicates the presence of folding intermediates (Grimsley *et al.*, 1997; Wang *et al.*, 1998; Sathish *et al.*, 2002). The tryptophan mutants which were created showed similar folding transitions to those observed with wild type TTK. The position of the intermediate formed on unfolding changes on the axis of fraction folded, although the  $\lambda_{\text{max}}$  of the intermediate is similar, the lower fraction folded of the intermediate being due to the blue shift in the  $\lambda_{\text{max}}$  of the native state of the W612F W622F mutant over the wild type and W612F W622F W628F mutant (Figure 4.17). The thermodynamic parameters determined for the folding transitions are similar for the entire transition between the three constructs studied, with an average total free energy change on folding of -8.15 kcal mol<sup>-1</sup> (Table 5.2). The intermediate present in the folding of TTK was much better resolved in the unfolding of TTK than in the refolding. The refolding and unfolding transitions of TTK as measured by tryptophan fluorescence are not super imposable. The native state is stable up to a GdnHCl concentration of ~0.6 M on the unfolding of TTK, but is not similarly stable in this range on the refolding of TTK. The behaviour of the W612F W622F mutant shows that the spectra observed for the native state of wild type TTK is dominated by the buried

tryptophans. The single tryptophan mutant of TTK shows that the intermediate which was observed is described well by the core tryptophan residue; however the differences observed between the refolding and unfolding of TTK were absent in the fluorescence probed folding of the single tryptophan mutant..

The comparison of the folding of TTK monitored by far-UV CD and tryptophan fluorescence identifies a folding intermediate in the equilibrium folding of TTK. The intermediate that accumulates on the unfolding pathway of TTK shows the features of a classical molten globule intermediate. The intermediate accumulates at ~ 1.5 M GdnHCl in the wild type and at similar GdnHCl concentrations in the tryptophan mutants. At this GdnHCl concentration the CD spectra show that at least 90% of the  $\alpha$ -helical content of the protein is still present. Upon the refolding of TTK an intermediate is formed at approximately 1.5 M GdnHCl. At this concentration of denaturant the proportion the secondary structure which is formed is closer to 70%, indicating that a significant amount of secondary structure has not yet formed in this intermediate. The intermediate which is formed could still be considered to be a molten globule intermediate, as a large proportion of the secondary structure is present in the intermediate, but it is clear that the intermediate states are different, since ~30 % of the secondary structure which is formed in the unfolding intermediate remains to be formed in the refolding intermediate. An examination of tryptophan emission spectra of the intermediates formed on the unfolding and refolding of TTK reveals that the intermediates have similar spectra, indicating a similar arrangement of tertiary structure in the two intermediates. Further study, possibly using a C-terminal lobe only construct in H/D exchange NMR studies would be valuable to identify the residues which remain formed into secondary structure upon unfolding but are not formed into secondary structure on refolding.

The refolding of p38 $\alpha$  showed that the core tryptophan of the kinase domain was key in the formation of the observed folding intermediate. Comparison of the folding of the tryptophan mutants showed that W207 identified the intermediate state, and the intermediate is therefore likely to be formed in the local environment of the W207 residue. The comparison of the folding studies performed on the W612F W622F W628F mutant of TTK show the I $\rightarrow$ N transition and the accumulation of the intermediate is more clearly defined upon refolding compared to

wild type TTK and W612F W622F mutant. This suggests that this intermediate forms in the area reported on by the W718 residue.

#### ***6.6. Super-imposable and Non Super-imposable Folding Curves***

The folding curves obtained for TTK via both far-UV CD and tryptophan fluorescence are different between the refolding of the protein and the unfolding of the protein, with the exception of the tryptophan fluorescence of the W612F W622F W628F mutant. The folding transitions are not super-imposable, indicating that the unfolding and refolding of TTK proceeds via different pathways (Figure 5.4C, 5.11C). The super-imposability of the unfolding and refolding transitions for the single tryptophan mutant of TTK indicates that the unfolding and refolding of this mutant proceed via the same pathway (Figure 5.18C). This indicates that the differences observed in the folding of TTK occur in the C-terminal lobe, but not in the core of the C-terminal lobe.

The folding of p38 $\alpha$  has been previously observed to produce super-imposable folding transitions with the folding probes of both far-UV CD and tryptophan fluorescence of various tryptophan mutants. All of the folding transitions observed with p38 $\alpha$  produced super-imposable transitions, indicating that the protein was folding and unfolding through the same equilibrium pathway. This is a key difference between the folding of the two protein kinases.

#### ***6.7. Simple Folding and Co-operative Folding.***

The far-UV CD observation of the folding of TTK showed a difference in the folding and the unfolding of TTK under equilibrium conditions. The native state of the protein was stable to a higher GdnHCl concentration on unfolding and on refolding. Similarly, the protein is fully unfolded at lower concentrations of GdnHCl upon refolding than upon unfolding (Figure 5.23). This results in hysteresis in the folding curves, indicating that the protein is unfolding and refolding through different pathways. A similar difference is observed in the unfolding and the refolding of TTK observed by tryptophan fluorescence. The unfolding of the kinase domain proceeds via a clearly defined equilibrium intermediate, and the native state is stable to approximately 0.6 M GdnHCl on unfolding (Figure 5.24). The refolding

of TTK does not show this stability of the native state at low GdnHCl concentrations, and the intermediate does not accumulate in the fashion seen upon unfolding. The refolding and unfolding of the single tryptophan mutant however is similar, indicating that the differences in the folding and unfolding occur in areas not monitored by the single tryptophan probe (Figure 5.24).

This is supported by the difference in the far-UV CD signal observed at the GdnHCl concentration where the intermediate observed by tryptophan fluorescence is formed. The CD signal is ~30 % lower upon refolding of the protein compared to the signal upon unfolding (Figure 5.23). This indicates that more secondary structure is retained in the intermediate. The hysteresis in the folding of TTK is similar to hysteresis observed in the folding of creatine kinase (Zhu *et al.*, 2001) and collagen III (Engel and Bächinger, 2000), which was shown to be caused by a cooperatively in the unfolding of the protein. TTK therefore unfolds via a cooperative mechanism which may be mediated through the inter-lobe interface. The refolding of TTK does not display the cooperatively observed on unfolding. This may be due to the lack of the stabilising influence of the inter-lobe interface present in the native state.

The folding of p38 $\alpha$  was observed to be fully reversible by both far-UV CD and tryptophan fluorescence. The folding transitions observed by both far-UV CD and tryptophan fluorescence were super-imposable, indicating that the unfolding and refolding of p38 $\alpha$  was proceeding via the same pathway (Davies, 2004). The folding experiments performed on p38 $\alpha$  do not indicate a cooperative unfolding of the native state of p38 $\alpha$ . This result points to a significant difference in the folding pathways of p38 $\alpha$  and TTK. However both kinases fold via an intermediate state which shares similar properties.

#### **6.8. Does the Kinase Domain Fold Via a Common Pathway?**

The folding of the kinase domain has been explored through the refolding of five protein kinases through two different refolding screens, and through the study of the equilibrium folding of the kinase domain of TTK. Previous work on the folding of p38 $\alpha$  has also been used to examine the existence of a common protein kinase folding mechanism. The refolding of AKT2, KIS, p38 $\alpha$ , PhK and TTK through the two refolding screens performed does not present a picture of a common folding mechanism. The patterns of refolding of the tested protein kinases with respect to the

pH at which refolding was performed, and the chemical additives which were effective in supporting the refolding are not similar across the five kinases tested, with the response to pH and the effective additives being different between the proteins. The analysis of the fractional factorial screen does not show the additives and combinations of additives selected in the screen to be effective for all the protein kinases tested; indeed some additives were both positive and negative in their effect on the folding of different kinases.

The more detailed study of the equilibrium folding of TTK does raise a commonality identified between the folding of p38 $\alpha$  and TTK. In both kinases the core tryptophan of the kinase domain, W207 in p38 $\alpha$  and W718 in TTK, was found to be essential for the folding of the kinase domain. This residue is both sequence and structurally conserved in both kinases and in other kinases also. This tryptophan found in the hydrophobic core of the kinase domain forms a key component of the folding of the kinases tested, and is important in the formation of the equilibrium folding intermediates observed for both TTK and p38 $\alpha$ . Further studies on the structure of this intermediate state would be needed to confirm that the conformation of the intermediate is similar; however the amount of formed secondary structure, when the differences in secondary structure content between the two kinases is considered, is similar.

Despite the formation of a similar folding intermediate in the folding of TTK and p38 $\alpha$ , the two kinases fold via dissimilar mechanisms. In the folding of p38 $\alpha$  the recovery of tertiary structure is not coincident with the recovery of secondary structure. In TTK the recovery of the tertiary structure is nearly coincident with the recovery of secondary structure on refolding. In addition, the recovery of structure on the refolding and the loss of structure on the unfolding of p38 $\alpha$  are superimposable; however this is not the case with TTK. This is the key difference between the two protein kinases. The folding of p38 $\alpha$  is fully reversible and follows the same pathways. The folding of TTK does not follow the same pathway, and instead unfolds through a cooperative mechanism, which is not followed on refolding.

The study of the folding of protein kinases presented here was restricted to serine/ threonine and dual specificity protein kinases. However, the limited study of the folding of FGFR1 by Davies (2004) supports the extension of the results

presented here to cover the tyrosine protein kinases as well since this protein does not refold from denatured protein in equilibrium experiments.

The combination of the refolding studies performed on the five protein kinases presented in chapters 2 and 3, and the folding studies performed on TTK in chapter 5, when compared with the studies of Davies (2004) on p38 $\alpha$  do not show a common mechanism of kinase folding. The unfolding of TTK is a cooperative process and the folding of p38 $\alpha$  is not similarly cooperative. The refolding screens on AKT2, KIS, p38 $\alpha$ , PhK and TTK have not identified common conditions beneficial to the refolding of all five of the screened protein kinases. It is likely, therefore, that the kinase domain folds through different pathways, although the study of the folding of more protein kinases would be required to support this conclusion. Therefore, the production of soluble protein kinases from *E. coli* will require a refolding screen for those kinases which do not fold in the host cell but must instead be refolded *in vitro*.

## **6.9. Future Work**

The characterisation of the folding of the kinase domain performed in detailed experiments so far has focused on the folding of the C-terminal lobe of the kinase domain due to the requirement of the core tryptophan for the correct folding of the kinase domain. The key questions that remain to be answered in the folding of p38 $\alpha$  and TTK cover the folding of the N-terminal lobe and the formation of the inter-lobe interface. To allow the study of the folding of these areas, in the context of the whole protein would require the creation of mutants exploiting FRET techniques to probe the folding of these regions. Following the creation of these mutants, and the study of the folding of these mutants through equilibrium techniques, the study of the kinetics of folding for both p38 $\alpha$  and TTK could be performed to complete the comparison of the folding of these two protein kinases and to identify if the difference seen between the two protein kinases are due to the formation and stability of the inter-lobe interface.

The study of the refolding of protein kinases was also limited to the study of serine/threonine and dual specificity protein kinase domains. It would be interesting to apply the screens to a number of tyrosine kinase domains to follow the folding behaviour of these domains, and to identify how different the folding of these domains is.

## References

**Abrieu, A., Magbaghi-Jaulin, L., Kahana, J.A., Peter, M., Castro, A., Vegneron, S., Larca, T., Cleveland, D.W. and Labbé, J.C.** (2001) Mps1 is a kinetochore-associated kinase essential for the vertebrate mitotic checkpoint. *Cell* **106**, 83-93.

**Adams, J.A.** (2001). Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* **101**, 2271-2290.

**Ami, D., Natalello, A., Taylor, G., Tonon, G. and Doglia, S.M.** (2006) Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. *Biochimica et biophysica acta* **1764** 793-799.

**Andrade, M. A., Chacón, P., Merelo, J. J. and Morán, F.** (1993). Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network. *Prot. Eng.* **6**, 383–390.

**Anfinsen, C. B.** (1973). Principles that govern the folding of protein chains. *Science* **181**, 223–230.

**Arakawa, T., Ejima, D., Tsumoto, K., Obeyama, N., Tanaka, Y., Kita, Y. and Timasheff, S.N.** (2007) Suppression of protein interactions by arginine: A proposed mechanism of the arginine effects, *Biophys. Chem.* **127**, 1-8.

**Arakawa, T. and Timasheff, S.N.** (1982) Stabilization of protein structure by sugars. *Biochemistry* **21**, 6536-6544.

**Arakawa, T. and Timasheff, S.N.** (1985) The stabilization of proteins by osmolytes. *Biophys. J.* **47**, 411-414.

**Bae, S.S., Cho, H., Mu, J. and Birbaum, M.J.** (2003) Isoform-specific Regulation of Insulin-dependent Glucose Uptake by Akt/Protein Kinase B. *J. Biol. Chem.* **278**, 49530-49536.



**Baker, D.** (2000). A surprising simplicity to protein folding. *Nature* **405**, 39–44.

**Batas, B., Schiraldi, C. and Chaudhuri, J.B.** (1999) Inclusion body purification and protein refolding using microfiltration and size exclusion chromatography, *Journal of Biotechnology* **68** 149-158

**Baum, J., Dobson, C. M., Evans, P. A. and Hanley, C.** (1989). Characterization of a partly folded protein by NMR methods: studies on the molten globule state of guinea pig alpha-lactalbumin. *Biochemistry* **28**, 7–13.

**Baynes, B.B., Wang, D.I. and Trout, D.L.** (2005) Role of arginine in the stabilization of proteins against aggregation. *Biochemistry* **44**, 4919-4925.

**Berg, J. M., Tymoczko, J. L. and Stryer, L.** (2002). *Biochemistry*, 5<sup>th</sup> edn. W. H. Freeman and Company, New York.

**Bjellqvist, B., Hughes, G.J., Pasquali, Ch., Paquet, N., Ravier, F., Sanchez, J.-Ch., Frutiger, S. and Hochstrasser, D.F.** (1993) The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis* **14** 1023-1031.

**Blackwell, J.R. and Horgan, R.** (1991) A novel strategy for production of a highly expressed recombinant protein in an active form. *FEBS Letters* **295**, 10-12.

**Bodo, G., Dintziz, H.M., Kendrew, J.C. and Wyckoff, H.W.** (1959) The crystal structure of myoglobin V. A low-resolution three dimensional fourier synthesis of sperm-whale myoglobin crystals. *Proc. Royal Society London* **253** 70-102.

**Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J. and Greenberg, M.E.** (1999) Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857-868.

**Brushia, R.J. and Walsh, D.A.** (1999) Phosphorylase kinase: the complexity of its regulation is reflected in the complexity of its structure. *Front. Biosci.* **15**, D618-D641.

**Bueno, M., Ayuso-Tejedor, S. and Sancho, S.** (2006) Do Proteins with Similar Folds Have Similar Transition State Structures? A Diffuse Transition State of the 169 Residue Apoflavodoxin. *J. Mol. Bio.* **359**, 813-824.

**Burstein, E. A., Abornev, S. M. and Reshetnyak, Y. K.** (2001). Decomposition of protein tryptophan fluorescence spectra into log-normal components. 1. Decomposition algorithms. *Biophys. J.* **81**, 1699–1709.

**Cao, P., Mei, J.J., Diao, Z.Y. and Zhang, S.** (2005) Expression, refolding, and characterization of human soluble BAFF synthesized in *Escherichia coli*. *Protein Expr. Purif.* **41**, 199-206.

**Carrió, M.M., Corcheroa, J.L. and Villaverde, A.** (1998) Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. *FEMS Microbiology Letters* **169** 9-15.

**Carrió, M.M. and Villaverde, A.** (2001). Protein Aggregation as bacterial inclusion bodies is reversible. *FEBS Letters* **489** 29-33.

**Chan, H. S., Bromberg, S., Dill, K. A.** (1995). Models of cooperativity in protein folding. *Phil. Trans. R. Soc. Lond.* **348**, 61–70.

**Cheung, M. S., Garcia, A. E. and Onuchic, J. N.** (2002). Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. USA* **99**, 685–690.

**Chong, Y. and Cheng, H.** (2000) Preparation of functional recombinant protein from *E. coli* using a nondetergent sulfobetaine, *Biotechniques* **29**, 1166-1167.

**Chow, M.K., Amin, A.A., Fulton, K.F., Fernando, T., Kamau, L., Batty, C., Louca, M., Ho, S., Whisstock, J.C., Bottomley, S.P. and Buckle, A.M.** (2006) The REFOLD database: a tool for the optimization of protein expression and refolding. *Nucleic Acids Res.* **34** 207-212

**Chyan, C. L., Wormald, C., Dobson, C. M., Evans, P. A. and Baum, J.** (1993). Structure and stability of the molten globule state of guinea-pig alpha-lactalbumin: a hydrogen exchange study. *Biochemistry* **32**, 5681-5691.

**Cowan, R.H., Davies, R.A. and Pinheiro, T.J.T.** (2008) A screening system for the identification of refolding conditions for a model protein kinase, p38 $\alpha$ . *Analytical Biochemistry* **376**, 25-38.

**Creighton, T. E.** (1990). Protein folding. *Biochem. J.* **270**, 1–16.

**Damaschun, G., Damaschun, H., Gast, K. and Zirwer, D.** (1999). Proteins can adopt totally different folded conformations. *J. Mol. Biol.* **291**, 715–725.

**Davies, G.** (2004) The folding of p38 mitogen activated protein kinase, PhD Thesis, *University of Warwick*.

**Deniz, A.A., Laurence, T.A. Beligere, G.S., Dahan, M., Martin, A.B., Chemla, D.S., Dawson, P.E., Schultz, P.G. and Weiss, S.** (2000) Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* **97** 5179-5184.

**Dill, K.A. and Shortle, D.** (1991) Denatured states of proteins. *Annual Review of Biochemistry* **60**, 795-825.

**Dobson, C. M.** (2004). Principles of protein folding, misfolding and aggregation. *Sem. Cell. Dev. Biol.* **15**, 3–16.

**Dobson, C. M., Sali, A. and Karplus, M.** (1998). Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* **37**, 868–893.

**Dolgikh, D. A., Gilmanshin, R. I., Brazhnikov, E. V., Bychkova, V. E., Semisotnov, G. V., Veniaminov, E. V. and Ptitsyn, O. B.** (1981). Alpha-Lactalbumin: compact state with fluctuating tertiary structure? *FEBS Lett.* **136**, 311-315.

**Dill, K. A. and Chan, H. S.** (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19.

**Eisenmesser, E.Z., Millet, O., Labeikovsky, W., Kozhnev, D.M., Wolf-Watz, M., Bosco, D.A., Skalicky, J.J., Kay, L.E. and Kern, D.** (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117-121.

**Engel, J. and Bächinger, H.P.** (2000) Cooperative equilibrium transitions coupled with a slow annealing step explain the sharpness and hysteresis of collagen folding. *Matrix Biology* **19** 235-244.

**Enslen, H., Raingeaud, J. and Davis, R. J.** (1998). Selective activation of p38 Mitogen-Activated Protein (MAP) kinase isoforms by the MAP kinase kinases and MKK3 and MKK6. *J. Biol. Chem.* **273**, 1741–1748.

**Fersht, A. R.** (1997). Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.* **7**, 3– 9.

**Fersht, A. R.** (1999). Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding. W. H. Freeman and Company, New York.

**Fersht, A. R.** (2000). Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. USA* **97**, 1525–1529.

**Font, J., Torrent, J., Ribo, M., Laurents, D.V., Balny, C., Vilanova, M. and Lange, R.** (2006) Pressure-Jump-Induced Kinetics Reveals a Hydration Dependent

Folding/Unfolding Mechanism of Ribonuclease A. *Biophysical Journal* **91** 2264-2274.

**Frantz, B., Klatt, T., Pang, M., Parsons, J., Rolando, A., Williams, H., Tocci, M.J., O'Keefe, S.J. and O'Neill, E.A.** (1998) The Activation State of p38 Mitogen-Activated Protein Kinase Determines the Efficiency of ATP Competition for Pyridinylimidazole Inhibitor Binding, *Biochemistry* **37** 13846-13853

**Georgiou, G. and Valax, P.** (1999). Isolating inclusion bodies from bacteria. *Methods in Enzymology*. **309**, 48-58.

**Gloss, L.** (2007) Tying the Knot that Binds. *Structure* **15** 2-4.

**Godzik, A.** (2002) Counting and Classifying Possible Protein Folds, *Trends in Biotechnology* **15** 147-151.

**Goldberg, M.E., Eeret-Bezancon, N., Vuillard, L. and Rabilloud, T.** (1996) Non-detergent sulfobetaines: A new class of molecules that facilitate *in vitro* protein renaturation, *Fold. Des. I*, 21-27.

**Greene, R. F. Jr. and Pace, C. N.** (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme,  $\alpha$ -chymotrypsin, and  $\beta$ -lactoglobulin. *J. Biol. Chem.* **249**, 5388–5393.

**Grimsley, J. K., Scholtz, J. M., Pace, C. N. and Wild, J. R.** (1997). Organophosphorus hydrolase is a remarkably stable enzyme that unfolds through a homodimeric intermediate. *Biochemistry* **36**, 14366–14374.

**Gruebele, M.** (2002). Protein folding: the free energy surface. *Curr. Opin. Struct. Biol.* **12**, 161–168.

**Hajdуч, E. Litherland, G.J. and Hundal, H.S.** ( 2001) Protein kinase B (PKB/Akt)--a key regulator of glucose transport? *FEBS Letters* **492**, 199-203.

**Hanahan, D.** (1983) Studies on transformation of *Escherichia coli* with plasmids, *J. Mol Bio.* **166** 557-80

**Harrowing, S.R. and Chaudhuri, J.B.** (2003) Effect of column dimensions and flow rates on size-exclusion refolding of beta-lactamase. *J. Biochem. Biophys. Methods* **56**, 177-188.

**Heukeshoven, J. and Dernick, R.** (1985) Simplified method for silver staining of proteins in polyacrylamide gels and the mechanism of silver staining. *Electrophoresis* **6**, 103-112.

**Hibino, T., Misawa, S., Wakiyama, M., Maeda, S., Yazaki, K., Kumagai, I., Ooi, T. and Mirua, K.** (1994) High-level expression of porcine muscle adenylate kinase in *Escherichia coli*: effects of the copy number of the gene and the translational initiation signals. *J. Biotechnology* **32**, 139-148.

**Huang, F., Settanni, G. and Fersht, A.R.** (2008) Fluorescence resonance energy transfer analysis of the folding pathway of Engrailed Homeodomain. *Protein Engineering Design and Selection* **21** 131-146.

**Hughson, F. M., Wright, P. E. and Baldwin, R. L.** (1990). Structural characterisation of a partly folded apomyoglobin intermediate. *Science* **249**, 1544-1548.

**Ikeguchi, ., Kuwajima, K., Mitani, M. And Sugai, S.** (1986) Evidence for the Identity between the Equilibrium Unfolding Intermediate and Transient Folding Intermediate: A Comparative Study of the Folding Reactions of  $\alpha$ -Lactalbumin and Lysozyme, *Biochemistry* **25** 6969-6972

**Invernizzi, G. and Grandori, R.** (2007) Detection of the equilibrium folding intermediate of  $\beta$ -lactoglobulin in the presence of trifluoroethanol by mass spectrometry. *Rapid Comms. In Mass Spec.* **21** 1049-1052

**Jenkins, D.C., Pearson, D.S., Harvey, H., Sylvester, I.D., Geeves, M.A. and Pinhiero, T.J.T.** (2009) Rapid folding of the prion protein captured by pressure-jump. *European Biophysics Journal* **38** 625-635.

**Kadi, N.El., Taulier, N., Le Huerou, J.Y., Gindre, M., Urbach, W., Nwigwe, I., Kahn, P.C. and Waks, M.** (2006) Unfolding and refolding of bovine serum albumin at acid pH: ultrasound and structural studies, *Biophys. J.* **91**, 3397-3404.

**Kang, J., Chen, Y., Zhao, Y. and Yu, H.** (2007) Autophosphorylation-dependent activation of human Mps1 is required for the spindle checkpoint. *PNAS* **104**, 20232-20237.

**Kannan, N. and Taylor, S.S.** (2008) Rethinking Pseudokinases. *Cell* **133**, 204-205.

**Katta, V. and Chait, B.T.** (1993) Hydrogen/Deuterium Exchange Electrospray Ionization Mass Spectrometry: A Method for Probing Protein Conformational Changes in Solution. *J. Am. Chem. Soc.* **115** 6317-6321.

**Kelly, S.M., Jess, T.J. and Price, N.C.** (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751** 119-139.

**Kim, P. S. and Baldwin, R. L.** (1990). Intermediates in the folding reactions of small proteins. *Ann. Rev. Biochem.* **59**, 631–660.

**Kim, A.H., Khursigara, G., Sun, X., Franke, T.F. and Chao, M.V.** (2001) Akt Phosphorylates and Negatively Regulates Apoptosis Signal-Regulating Kinase 1. *Mol. Cell. Bio.* **21**, 893-901.

**Klein-Seetharaman, J., Oikawa, M., Grimshaw. S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. and Schwalbe, H.** (2002). Long-range interactions within a non-native protein. *Science* **295**, 1719–1722.

**Kuznetsova, I.M., Turoverov, K.T. and Uversky, V.N.** (2004) Use of the diagram method to analyse the protein unfolding-refolding reactions: fishing out the “invisible” intermediates. *J. Proteome Research* **3**, 485-494.

**Ladokhin, A. S., Jayasinghe, S. and White, S. H.** (2000). How to measure and analyse tryptophan fluorescence in membranes properly, and why bother? *Anal. Biochem.* **285**, 235–245.

**Laemmli, U. K.** (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.

**Lakowicz, J.R.** (2006) Principles of fluorescence spectroscopy, Springer, New York.

**Lauzé, E., Stoelcker, B., Luca, F.C., Weiss, E., Chutz, A.R. and Winey, M.** (1995) Yeast spindle pole body duplication gene MPS1 encodes an essential dual specificity protein kinase. *EMBO Journal* **14**, 1655-1663.

**Levinthal, C.** (1968). Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45.

**Li, L., Wetxl, S., Plückthun, A. and Fernandex, J.M.** (2006) Stepwise unfolding of ankyrin repeats in a single protein revealed by atomic force microscopy. *Biophysical Journal* **90** L30-L32.

**Li, Y., Dowbenko, D. and Lasky, L.A.** (2001) AKT/PKB Phosphorylation of p21<sup>Cip/WAF1</sup> Enhances Protein Stability of p21<sup>Cip/WAF1</sup> and Promotes Cell Survival. *J. Biol. Chem.* **277**, 11352-11361.

**Lui, S.T., Chan, G.K.T., Hittle, J.C., Fujii, G., Lees, E. and Yen, T.Y.** (2003) Human MPS1 Kinase Is Required for Mitotic Arrest Induced by the Loss of CENP-E from Kinetochores. *Mol. Biol. Of the Cell* **14**, 1638-1651.



- Machida, S., Ogawa, S., Xiaohua, S., Takaha, T., Fujii, K. and Hayashi, K.** (2000) Cycloamylose as an efficient artificial chaperone for protein folding, *FEBS Letters* **486**, 131-135.
- Mailer, C.S. and Deinzer, M.L.** (2005) Protein Conformations, Interactions and H/D Exchange, *Methods in Enzymology* **402** 312-360
- Mallam, L. and Jackson, S.E.** (2007) A Comparison of the Folding of two Knotted Proteins: YbeA and YibK. *J. Mol. Bio.* **356** 650-665.
- Mannall, G.J., Tichener-Hooker, N.J. and Dalby, P.A.** (2007) Factors affecting protein refolding yields in a fed-batch and batch-refolding system. *Biotechnology and Bioengineering* **97**, 1523-1534.
- Manning, G., Whyte, D.B., Martinez, R., Hunter, T. and Sudarsanam, S.** (2002) The Protein Kinase Complement of the Human Genome. *Science* **298**, 1912-1934.
- Mattison, C., Old, W.M., Steiner, E., Huneycutt, B.J., Resing, K.A., Ahn, N.G. and Winey, M.** (2007) Mps1 Activation Loop Autophosphorylation Enhances Kinase Activity. *J. Biol. Chem.* **282**, 30553-30561.
- Maucer, A., Camonis, J.H. and Sobel, A.** (1995) Stathmin interaction with a putative kinase and coiled-coil-forming protein domains. *PNAS* **92**, 3100-3104.
- Maucer, A., Ozon, S., Manceeau, V., Gavet, O., Lawler, S., Curmi, P. and Sobel, A.** (1997) KIS is a Protein Kinase with an RNA Recognition Motif. *J. Biol. Chem.* **272**, 23151-23156.
- Matouschek, A.** (2003). Protein unfolding – an important process *in vivo*? *Curr. Opin. Struct. Biol.* **13**, 98–109.
- Mello, C.C. and Barrick, D.** (2003) Measuring the stability of partially folded proteins using TMAO. *Protein Science* **12**, 1522-1529.

**Morjana, N. A., McKeone, B. J. and Gilbert, H. F.** (1993). Guanidine hydrochloride stabilization of a partially unfolded intermediate during the reversible denaturation of protein disulfide isomerase. *Proc. Natl. Acad. Sci. USA* **90**, 2107–2111.

**Müller, C. and Rinas, U.** (1999) Renaturation of heterodimeric platelet-derived growth factor from inclusion bodies of recombinant *Escherichia coli* using size-exclusion chromatography. *Journal of Chromatography A* **855**, 203-213.

**Nanda, V., Liang, S-M. And Brand, L.** (2000) Hydrophobic Clustering in Acid-Denatured IL-2 and Fluorescence of a Trp NH $\cdots\pi$  H-bond, *Biochemical and Biophysical Research Communications* **279** 770-778

**Navon, A., Ittah, V., Laity, J. H., Scheraga, H. A., Haas, E. and Gussakovsky, E.** (2001). Local and long-range interactions in the thermal unfolding transition of bovine pancreatic ribonuclease A. *Biochemistry* **40**, 93–104.

**Nienhaus, G.U.** (2009) Single-Molecule fluorescence studies of protein folding. *Methods Mol. Biol.* **490** 311-337.

**Nishimura, C., Uverski, V.N. and Fink, A.L.** (2001) Effect of salts on the stability and folding of staphylococcal nuclease, *Biochemistry* **40**, 4113-4128.

**Ohgushi, M. and Wada, A.** (1983). ‘Molten globule state’: a compact form of globular proteins with mobile side-chains. *FEBS Lett.* **164**, 21–24.

**Owen, D.J., Papageorgiou, A.C. Garman, E.F., Noble, M.E.M. and Johnson, L.N.** (1985) Expression, Purification and Crystallisation of Phosphorylase Kinase Catalytic Domain. *J. Mol. Bio.* **246** 374-381.

**Pace, C. N.** (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266-280.

**Park, H.S., Kim, M.S., Huh, S.H., Park, J., Chung, J., Kand, S.S. and Choi, E.J.** (2002) Akt (protein kinase B) negatively regulates SEK1 by means of protein phosphorylation. *J. Biol. Chem.* **277**, 2573-2578.

**Pinheiro, T.J.T.** (1994) The interaction of horse heart cytochrome c with phospholipids bilayers. Structural and dynamic effects. *Biochimie* **76** 489-500

**Plaxco, K. W., Simons, K. T. and Baker, D.** (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **227**, 985–994.

**Ptitsyn, O. B.** (1987). Protein folding: hypotheses and experiments. *J. Prot. Chem.* **6**, 273–293.

**Ptitsyn, O. B.** (1992). The molten globule state. In: Protein folding (Ed. T. E. Creighton), pp243-300. New York: W. H. Freeman and Co.

**Ptitsyn, O. B., Bychkova, V. E. and Uversky, V. N.** (1995). Kinetic and equilibrium folding intermediates. *Phil. Trans. R. Soc. Lond. B* **348**, 35–41.

**Ptitsyn, O. B., Pain, R. H., Semisotnov, G. V., Zerovnik, E. and Razgulyaev, O. I.** (1990). Evidence for a molten globule state as a general intermediate in protein folding. *FEBS Lett.* **262**, 20–24.

**Ptitsyn, O. B. and Rashin, A. A.** (1975). A model of myoglobin self-organization. *Biophys. Chem.* **3**, 1–20.

**Radford, S. E. and Dobson, C. M.** (1995). Insights into protein folding using physical techniques: studies of lysozyme and  $\alpha$ -lactalbumin. *Phil. Trans. R. Soc. Lond. B* **348**, 17–25.

**Ren, J., Kachel, K., Kim, H., Malenbaum, S. E. and Collier, R. J.** (1999). Interaction of diphtheria toxin T domain with molten globule-like proteins and its implications for translocation. *Science* **284**, 955–957.

**Raingeaud, J., Gupta, S., Rogers, J. S., Dickens, M., Han, J., Ulevitch, R.J. and Davis, R. J.** (1995). Pro-inflammatory cytokines and environmental stress cause p38 mitogen-activated protein kinase activation by dual phosphorylation on tyrosine and threonine. *J. Biol. Chem.* **270**, 7420-7426.

**Raingeaud, J., Whitmarsh, A. J., Barrett, T., Derijard, B. and Davis, R. J.** (1996). MKK3- and MKK6-regulated gene expression is mediated by the p38 mitogen-activated protein kinase signal transduction pathway. *Mol. Cell. Biol.* **16**, 1247–1255.

**Rinas, U., Hoffmann, F., Betiku, E., Estape, D. and Marten, S.** (2007) Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in Escherichia coli. *J. Biotechnology* **127** 244-257/

**Robinson, C. V., Gross, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M. and Radford, S. E.** (1994). Conformation of GroEL-bound alpha-lactalbumin probed by mass spectrometry. *Nature* **372**, 646–651.

**Rouse, J., Cohen, P., Trigon, S., Morange, M., Alonso-Llamazares, A., Zamanillo, D., Hunt, T. and Nebreda, A. R.** (1994). A novel kinase cascade triggered by stress and heat shock that stimulates MAPKAP kinase-2 and phosphorylation of the small heat shock proteins. *Cell* **78**, 1027–1037.

**Ruan, K., Xu, C., Li, T., Li, J., Lange, R. and Balny, C.** (2003) The thermodynamic analysis of protein stabilization by sucrose and glycerol against pressure-induced unfolding., *Eur. J. Biochem.* **270**, 1654-1661.

**Ryu, K., Kim, Y.H., Kim, Y., Lee, J.S., Jeon, B. and Kim, T.Y.** (2008) Increased yield of high-purity and active tetrameric recombinant human EC-SOD by solid phase refolding. *J. Microbiol. Biotechnol.* **18**, 1648-1654.

**Sale, E.M. and Sale, G.J.** (2007) Protein kinase B: signalling roles and therapeutic targeting. *Cell Mol. Life Sci.* **65**, 113-127.

**Santoro, M. M. and Bolen, D. W.** (1988). Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of Phenylmethanesulfonyl  $\alpha$ -Chymotrypsin using different denaturants. *Biochemistry* **27**, 8063–8068.

**Sathish, H. A., Cusan, M., Aisenbrey, C. and Bechinger, B.** (2002). Guanidine hydrochloride induced equilibrium unfolding studies of colicin B and its ion channel-forming fragment. *Biochemistry* **41**, 5340–5347.

**Sato, S. and Raleigh, D.P.** (2002) pH-dependent Stability and Folding Kinetics of a Protein with an Unusual  $\alpha$ - $\beta$  Topology: The C-terminal Domain of the Ribosomal Protein L9. *J. Mol. Bio.* **318**, 571-582.

**Scott, K.A. and Daggett, V.** (2007) Folding Mechanisms of Proteins with High Sequence Identity but Different Folds. *Biochemistry* **46** 1545-1556.

**Sherr, C.J. and Roberts, J.M.** (2004) Living with or without cyclins and cyclin-dependent kinases. *Genes and Development* **18**, 2699-2711.

**Shin, H. C., Merutka, G., Waltho, J. P., Wright, P. E. and Dyson, H. J.** (1993a). Peptide models of protein folding initiation sites. 2. The G-H turn region of myoglobin acts as a helix stop signal. *Biochemistry* **32**, 6348–6355.

**Shin, H. C., Merutka, G., Waltho, J. P., Tennant, L.L., Dyson, H.J. and Wright, P.E.** (1993b). Peptide models of protein folding initiation sites. 3. The G-H helical hairpin of myoglobin. *Biochemistry* **32**, 6356–6364.

**Shortle, D. and Ackerman. M.S.** (2001) Persistence of native-like topology in a denatured protein in 8M urea. *Science* **293**, 487-489.

**Silow, M. and Oliveberg, M.** (2003) High concentrations of viscogens decrease the protein folding rate constant by prematurely collapsing the coil, *J. Mol. Bio* **326**, 263-271.

**Sobel, A.** (1991) Stathmin: a relay phosphoprotein for multiple signal transduction? *Trends. Biochem. Sci.* **16**, 301-305.

**Song, G., Ouyang, G. and Bao, S.** (2005) The activation of Akt/PKB signaling pathway and cell survival. *J. Cell. Mol. Med.* **9**, 59-71.

**Stryer, L. And Haugland, R.P.** (1967) Energy Transfer: A Spectroscopic Ruler, *Proc. Natl. Acad. Sci. USA* **58** 719-726

**Stucke, V.M., Silljé, H.H., Arnaud, L. and Nigg, E.A.** (2002) Human Mps1 kinase is required for the spindle assembly checkpoint but not for centrosome duplication. *EMBO Journal* **21**, 1723-1732.

**Sullivan, J.E., Holdgate, G.A., Campbell, D., Timms, D., Gerhardt, S., Breed, J., Breeze, A.L., Bermingham, A., Pauptit, R.A., Norman, R.A., Embrey, K.J., Read, J., VanScyoc, W.H. and Ward, W.H.** (2005) Prevention of MKK6-dependant activation by binding to p38 alpha MAP kinase. *Biochemistry* **44**, 16475-16490.

**Tanford, C.** (1970). Protein denaturation part C. theoretical models for the mechanism of denaturation. *Adv. Prot. Chem.* **24**, 1–95.

**Tanford, C., Kawahara, K., Lapanje, S., Hooker Jr., T.M., Zarlengo, M.H., Salahuddin, A., Aune, K.C. and Takagi, T.** (1967) Proteins as random coils. III. Optical rotatory dispersion in 6 M guanidine hydrochloride. *J. Am. Chem. Soc.* **89**, 5023-5029.

**Tokatlidis, K., Dhurjati, P., Millet, J., Beguin, P. and Aubert, J.P.** (1991) High activity of inclusion bodies produced in *Escherichia coli* overproducing *Clostridium thermocellum* endoglucanase D. *FEBS letters* **282** 205-208.

**Tsumoto, K., Ejima, D., Kita, Y. and Arakawa, T.** (2005) Review: Why is arginine effective in suppressing aggregation? *Protein Peptide Letters* **12**, 613-619.

**Vendruscolo, M., Zurdo, J., MacPhee, C. E. and Dobson, C. M.** (2003). Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Phil. Trans. R. Soc. Lond. A.* **361**, 1205–1222.

**Ventura, S. and Villaverde, A.** (2006) Protein quality in bacterial inclusion bodies. *Trends in Biotechnology* **24** 179-185.

**Vigneron, S., Prieto, S., Bernis, C., Labbé, J.C., Castro, A. and Lorca, T.** (2004) Kinetochores localization of spindle checkpoint proteins: who controls whom? *Mol. Biol. Cell.* **15**, 4584-4596.

**Villaverde, A. and Carrió, M.M.** (2003) Protein aggregation in bacteria: biological role of inclusion bodies. *Biotechnology Letters* **25** 1385-1395.

**Waltho, J. P., Feher, V. A., Merutka, G., Dyson, H. J. and Wright, P. E.** (1993). Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry* **32**, 6337–6347.

**Wang, Y., Alexandrescu, A. T. and Shortle, D.** (1995). Initial studies of the equilibrium folding pathway of staphylococcal nuclease. *Phil. Trans. R. Soc. Lond. A.* **348**, 27–34.

**Wang, Q., Buckle, A.M., Foster, N.W., Johnson, C.M. and Fersht, A.R.** (1999) Design of highly stable functional GroEL minichaperones. *Protein Science* **8** 2186-2193.

**Wang, C., Lascu, I. and Giartosio, A.** (1998). Bovine serum fetuin is unfolded through a molten globule state. *Biochemistry* **37**, 8457–8464.

- Wang, Z.X.** (1998) A re-estimation for the total numbers of protein folds and superfamilies, *Protein Engineering* **11** 621-626
- Wedemeyer, W.J., Welker, E. and Scheraga, H.A.** (2002) Proline cis-trans isomerization and protein folding. *Biochemistry* **41** 14637-14644.
- Wei, J.H., Chou, Y.F., Ou, Y.H., Yeh, Y.H., Tyan, S.W., Sun, T.P., Shen, C.Y. and Shieh, S.Y.** (2005) TTK/hMps1 Participates in the Regulation of DNA Damage Checkpoint Response by Phosphorylating CHK2 on Threonine 68. *J. Biol. Chem.* **280**, 7748-7757.
- Weiss, E. and Winey, M.** (1996) The *Saccharomyces cerevisiae* spindle pole body duplication gene MPS1 is part of a mitotic checkpoint. *J. Cell. Bio.* **132**, 111-123.
- Willis, M.S., Hogan, J.K., Prabhakar, P., Liu, X., Tsai, K., Wei, Y. and Fox, T.** (2005) Investigation of protein refolding using fractional factorial screen: A study of reagent effects and interactions, *Protein Science* **14**, 1818-1826.
- Wilson, K.P., McCaffrey, P.G., Hsiao, K., Pazhanisamy, S., Galullo, V., Bemis, G.W., Fitzgibbon, M.J., Caron, P.R., Murko, M.A. and Su, M.S.S.** (1997) The structural basis for the specificity of pyridinylimidazole inhibitors of p38 MAP kinase, *Chemistry and Biology* **4** 423-431
- Winey, M., Goetsch, L. Baum, P. and Byers, B.** (1991) MPS1 and MPS2: novel yeast genes defining distinct steps of spindle pole body duplication. *J. Cell. Bio.* **114**, 745-754.
- Wetlaufer, D. B.** (1973). Nucleation, rapid folding and globular intrachain regions in proteins. *Proc. Natl. Acad. Sci. USA* **70**, 697-701.
- Wolf, Y.I., Grishin, N.V. and Koonin, E.V.** (2000) Estimating the number of protein folds and families from complete genome data, *J. Mol. Bio.* **299** 897-905



**Xinsheng, L., Fan, K. and Wang, W.** (2003) The number of protein folds and their distribution over families in nature. *Proteins: Structure, Function and Bioinformatics* **54** 491-499

**Yasuda, M., Murakami, Y., Sowa, A., Ogino, H. and Ishikawa, H.** (1998) Effect of additives on refolding of a denatured protein, *Biotechnol. Prog.* **41**, 601-606.

**Yi, Q., Scalley-Kim, M.L., Alm, E.J. and Baker, D.** (2000) NMR characterization of residual structure in the denatured state of protein L. *J. Mol. Bio.* **299**, 1341-1351.

**Zardeneta, G. and Horowitz, P.M.** (1992) Micelle-assisted protein folding. Denatured rhodanese binding to cardiolipin-containing lauryl maltoside micelles results in slower refolding kinetics but greater enzyme reactivation, *J. Biol. Chem.* **267**, 5811-5816.

**Zarrine-Asfar, A., Larson, S.M. and Davidson, A.R.** (2005) The family feud: do proteins with similar structures fold via the same pathway? *Curr. Opp. In Struc. Biol.* **15**, 42-49.

**Zhu, L., Fan, Y.X. and Zhou, J.M.** (2001) Identification of equilibrium and kinetic intermediates involved in folding of urea-denatured creatine kinase. *Biochimica et Biophysica Acta* **1544**, 320-332.

# A screening system for the identification of refolding conditions for a model protein kinase, p38 $\alpha$

Richard H. Cowan<sup>a</sup>, Rick A. Davies<sup>b</sup>, Teresa T.J. Pinheiro<sup>a,\*</sup>

<sup>a</sup> Department of Biological Sciences, Gibbet Hill Road, University of Warwick, Coventry CV4 7AL, UK

<sup>b</sup> Discovery Enabling Capabilities & Sciences, AstraZeneca, Mereside, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

Received 23 October 2007

Available online 2 February 2008

## Abstract

Protein kinases are key drug targets involved in the regulation of a wide variety of cellular processes. To aid the development of drugs targeting these kinases, it is necessary to express recombinant protein in large amounts. The expression of these kinases in *Escherichia coli* often leads to the accumulation of the expressed protein as insoluble inclusion bodies. The refolding of these inclusion bodies could provide a route to soluble protein, but there is little reported success in this area. We set out to develop a system for the screening of refolding conditions for a model protein kinase, p38 $\alpha$ , and applied this system to denatured p38 $\alpha$  derived from natively folded and inclusion body protein. Clear differences were observed in the refolding yields obtained, suggesting differences in the folded state of these preparations. Using the screening system, we have established conditions under which soluble, folded p38 $\alpha$  can be produced from inclusion bodies. We have shown that the refolding yields obtained in this screen are suitable for the economic large-scale production of refolded p38 $\alpha$  protein kinase.

© 2008 Elsevier Inc. All rights reserved.

**Keywords:** Protein kinases; Protein folding; Folding additives; Inclusion bodies; Refolding; Recombinant protein expression

Protein kinases are key components of cell signaling pathways involved in the regulation of many critical cellular functions. Because of the key role that protein kinases have in signaling cascades and regulatory functions in cells, they are attractive drug targets. Many protein kinases have altered expression or activity profiles in a wide variety of diseases. Cancers are often associated with overactive or overexpressed kinases, such as cyclin-dependent kinases in which overexpression alters the cell cycle [1]. Other kinases are implicated in conditions such as rheumatoid arthritis [2].

P38 $\alpha$  is a member of the protein kinase family, which is viewed as a desirable target for pharmaceutical compounds. P38 $\alpha$  is a 41-kDa protein, that exhibits serine/threonine-specific phosphorylation activity. The targets of the

activity of p38 $\alpha$  are transcription factors such as ATF2<sup>1</sup> [3] and other protein kinases such as MAPKAP2 [4]. The signaling pathway that leads to the activation of p38 $\alpha$  is triggered by cell stresses, such as the presence of bacterial lipopolysaccharide or osmotic shock [5]. The presence of these signals causes the production of proinflammatory cytokines such as interleukin 1. The continued production of these cytokines further stimulates p38 $\alpha$  and leads to

<sup>1</sup> Abbreviations used: ATF2, activating transcription factor 2; Capso, 3-(cyclohexylamino)-2-hydroxy-1-propanesulfonic acid; CD, circular dichroism; Chaps, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; DTT, dithiothreitol; ESI-MS, electrospray ionization mass spectrometry; Hepes, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; IPTG, isopropyl  $\beta$ -D-1-thiogalactopyranoside; MAPKAPK2, MAP kinase activated protein kinase 2; NDSB, 3-(1-pyridinio)-1-propanesulfonate; PEG, polyethylene glycol; Mes, 2-(*N*-morpholino)ethanesulfonic acid; SDS-PAGE, sodium dodecyl sulfate polyacrilamide gel electrophoresis; SPR, surface plasmon resonance; UTDC, ureidoquinazoline target definition compound; NHS, *N*-hydroxysulfosuccinimide.

\* Corresponding author. Fax: +44 2476 523 701.

E-mail address: [t.pinheiro@warwick.ac.uk](mailto:t.pinheiro@warwick.ac.uk) (T.T.J. Pinheiro).

inflammatory diseases, such as rheumatoid arthritis [2]. The key role of the p38 $\alpha$ -signaling cascade in inflammatory responses underlies the pharmaceutical interest in p38 $\alpha$ . By inhibiting the activity of p38 $\alpha$  with small-molecule inhibitors, stimulation of p38 $\alpha$  would be blocked, leading to a down regulation of the inflammatory response.

The production of recombinant protein kinases is important for the development of specific inhibitory compounds for use as drugs. Purified, correctly folded, recombinant protein is essential to solve the structure of the protein kinase in complex with a putative inhibitor, allowing the binding mode and routes for increasing the specificity or potency of the inhibitor to be identified [6]. Additionally, large-scale provision of recombinant protein is required for the development of high-throughput screens aimed at identifying small-molecule inhibitors. Protein for these purposes is usually produced, by preference, in simple bacterial expression systems for reasons of cost, speed, and yield. However, several protein kinases are challenging to express in a soluble form in *Escherichia coli*, commonly aggregating into insoluble deposits of protein called inclusion bodies [7]. These inclusion bodies are ordered aggregates [8] deposited by cells when the folding machinery of the cell is overloaded by the high expression level of the recombinant protein [9]. Inclusion bodies form a convenient source of high-quality, easy to isolate protein [10]. However, it is necessary to solubilize the protein from the inclusion bodies and then refold the protein before it can be used for structural studies. The refolding of protein kinases is regarded a challenging prospect, with only a single success in the refolding of the catalytic subunit of phosphorylase kinase reported in the literature [11].

The solubilization of protein from inclusion bodies is commonly performed by the addition of high concentrations of chaotropic denaturants such as guanidine or urea to the purified inclusion bodies. These chaotropic denaturants solubilize and denature inclusion bodies by combinations of direct and indirect actions. First, the structure of water is perturbed, so that hydrophobic molecules are more soluble in water. Second, denaturants can also bind to polar groups on the protein molecule, disrupting electrostatic and hydrogen bonding interactions key for the maintenance of the protein structure [12]. Addition of high concentrations of these denaturants to proteins is generally considered to result in the complete disruption of permanent secondary structure within the protein [13].

The folding of proteins from the unfolded state is a complex procedure and is thought to proceed along a large number of different pathways [14]. This process of folding to the native state is usually depicted by the protein traveling through a landscape of conformations of differing energies toward an energy minimum. However, this landscape (also known as the folding funnel) contains many local energy minima in which a polypeptide chain may become trapped as it attempts to fold to the native state. These energy minima, from which the polypeptide chain may not be able to escape, are termed misfolded states of the

protein and are prone to aggregation. This aggregation commonly occurs through exposed hydrophobic patches that in the folded protein would be protected and, when the protein folds *in vivo*, could be shielded by chaperone proteins, such as Hsp90 [15].

To avoid the misfolding of proteins that often occurs during refolding, certain chemical additives can be used. These additives function in a variety of ways to promote the refolding of the protein to the native state. Some additives, such as detergents, bind to exposed hydrophobic patches on partially refolded protein to protect the protein from aggregation. Low concentrations of denaturant can also be used; these stabilize several intermediate states, increasing the potential refolding yield obtained once the denaturant is reduced. Osmolytes, such as betaine, have also been shown to promote the refolding of certain proteins to their native states [16].

There is no rational method for predicting conditions, that will promote the refolding of proteins. Consequently, to rapidly identify conditions under which a protein may refold to the native form from a denatured state, a wide variety of conditions may need to be examined. To do this, it is desirable to format this search for refolding conditions into a screen. Several generic refolding screens, such as the iFold screen (Novagen), are available commercially. However, these screens do not specify the methods that are used to identify the conditions under which refolding has occurred. Commonly, the use of a specific assay for the protein to be refolded is suggested. This approach is not suitable for protein kinases, because to measure kinase activity, activation of the protein kinase is required. This generally occurs via a phosphorylation mechanism involving an activating protein kinase, a process possibly inhibited by the refolding additives. It would also be necessary to purify the activated protein kinase from the activating protein kinase. Therefore, to screen for the refolding of protein kinases it is desirable to identify and validate other methods to measure the yields of refolding and the folded state of the refolded protein. These methods must be carefully selected, because the additives that are often used to promote the refolding of proteins can interfere with the methods that are used to quantify the yields of refolded protein. Additionally, these additives are often used at high concentrations, further complicating the identification of suitable methods for the assessment of folded state of the refolded protein.

To allow the development of a screen for the refolding of protein kinases and the selection and development of methods for the identification of the yields of refolded protein kinases, a model protein kinase was selected. We used the mitogen-activated protein kinase p38 $\alpha$ . This protein kinase was selected because it is possible to produce p38 $\alpha$  in *E. coli* either as a soluble folded protein or as inclusion bodies, allowing the easy production of material for refolding and for the validation of the analytical methods proposed. This also made possible the comparison of the refolding of protein kinases from inclusion bodies and from native protein that has been denatured.

In this report we demonstrate a high-throughput screen able to identify conditions to promote the refolding of a model protein kinase, p38 $\alpha$ . We show a series of analytical methods able to deliver specific information on the yields of refolded protein kinase and that could be applied to protein kinases other than p38 $\alpha$ . The refolded protein identified by a number of analytical methods was shown to be correctly folded. Increasing the scale at which refolding is performed does not lead to large changes in the yields of refolded protein. Additionally, the refolding of p38 $\alpha$  results in the production of amounts of protein suitable for the structural study of the kinase. We also show that the refolding yields obtained are strongly dependent on the pH at which the refolding occurs.

## Materials and methods

### Materials

NV-10 was purchased from Novexin Ltd (Cambridge, UK). 3-(1-Pyridino)-1-propanesulfonate and  $\beta$ -cyclodextrin were purchased from Fluka (Buchs SG, CH). Tris was purchased from Acros Organics (Geel, BE); P20 surfactant was supplied by Biacore and dimethyl sulfoxide by Fisons (Ipswich, UK). All other chemicals were supplied by Sigma–Aldrich (Poole, UK).

### Production of human p38 $\alpha$ inclusion bodies

Recombinant human p38 $\alpha$  was produced as insoluble protein in *E. coli* strain BL21 (DE3), transformed with an expression plasmid encoding full-length human p38 $\alpha$ . Cells were grown to a OD<sub>600</sub> of 0.6 at 37 °C, induced for 5 h with 0.4 mM IPTG, and harvested by centrifugation at 8000g for 20 min. Inclusion bodies were prepared using a modification of the protocol described by Georgio and Valax [17]. Briefly, isolated cells were resuspended in ice-cold 50 mM Tris, 150 mM NaCl, 2 mM DTT, pH 9.0, and lysed by sonication. Insoluble material was separated by centrifugation (35,000g for 40 min) and resuspended in the same buffer. The insoluble fraction was washed with 2% Triton X-100 and 2 M urea to remove membrane fragments and loosely bound proteins. Washed inclusion bodies were solubilized in 50 mM Tris, 150 mM NaCl, 8 M Urea, 10 mM DTT, pH 9.0 for 1 h at 30 °C and stored at 4 °C. Protein concentration was calculated from absorbance at 280 nm, ( $A_{280}$ ) using  $\epsilon_{280}$  of 48,130 M<sup>-1</sup> cm<sup>-1</sup>. The molecular weight of the protein and the absence of covalent modifications were confirmed by ESI-MS.

### Production of native p38 $\alpha$

Recombinant human p38 $\alpha$  was produced according to Davies [18]. Native, soluble p38 $\alpha$  for comparing refolding with inclusion body protein was denatured with 8 M urea, 10 mM DTT for 1 h at 30 °C and stored at 4 °C until use. The protein mass was verified by ESI-MS.

### Design of a refolding screen for kinases

A series of additives known to be effective in facilitating the refolding of proteins were selected to be included in a screen for the refolding of p38 $\alpha$ . In total 31 additives were chosen and grouped according to their chemical features. Arginine, glycine, L-proline, sarcosine, and an arginine/glutamate mix were grouped together into the “amino acids” group [16,19–34]. Glucose, betaine sorbitol, and trimethylamine *N*-oxide were grouped into the “osmolytes” group [16,19–34]. Sodium chloride, sodium sulfate, and ammonium sulfate were grouped into the “simple salts” group [16,19–34]. Guanidine (2, 1, and 0.5 M) and urea (2, 1, and 0.5 M) were grouped as “denaturants” [16,19–34]. Lauryl maltoside, Chaps, and Triton X-100 were grouped as “detergents” [16,19–34]. Cyclohexanol, 1-pentanol, ethanol, glycerol,  $\beta$ -cyclodextrin, ethylene glycol, and PEG 3440 were grouped as “alcohols and polyols” [16,19–34]. 3-(1-Pyridino)-1-propanesulfonate, formamide, and NV-10 were grouped as “other additives” [16,19–34]. These 31 additives and a control lacking a specific refolding additive were formatted into a 96-well refolding screen, utilizing three different buffers (Table 1). The three buffers used were 0.1 M Mes (pH 5.8), 50 mM Tris (pH 8.0), and 0.1 M Capso (pH 9.5).

### Refolding of human p38 $\alpha$

Refolding of p38 $\alpha$  was initiated by rapid dilution of denatured protein into various renaturation buffers in a 96-well screen, formatted in four 24-deep-well plates. A volume of 5 ml of each renaturation buffer was aliquoted into each well. Under rapid agitation at 4 °C, 100  $\mu$ l of denatured protein solution, at 5 mg/ml in 8 M urea, was added in a single step to each renaturation buffer, for a final protein concentration of 0.1 mg/ml and a final urea concentration of 160 mM. Refolding was allowed to occur under gentle agitation overnight at 4 °C. After refolding, samples identified as containing soluble protein by SDS–PAGE were concentrated 10-fold using Centricon concentrators and then dialyzed against 10 mM Hepes, 150 mM NaCl, pH 7.4.

The percentage recovery of refolded protein, as identified by the analytical techniques described below, was calculated by comparing the protein concentration identified by the analytical techniques to the theoretical maximum possible recovery of protein.

### SDS–PAGE of refolded protein

Refolded protein was taken directly from the screen and added to E-PAGE sample buffer (Invitrogen) and 1%  $\beta$ -mercaptoethanol. Samples were heated to 70 °C for 10 min in a PCR thermal cycler block, according to the manufacturers' instructions. Samples were run using 96-lane E-PAGE gels (Invitrogen; Cat. No. EP096-06) and

Table 1  
Refolding screen

0.9 M Arginine	0.5 M Glycine	1 M Proline	2 M Sarcosine	50 mM Arginine/ 50 mM Glutamate	1 M Glucose	10% Betaine	20% Sorbitol	1 M TMAO	0.5 M NaCl	0.5 M Sodium Sulphate	1 M Ammonium Sulphate 10% Ethanol
2 M Guanidine	1 M	0.5 M	2 M Urea	1 M Urea	0.5 M Urea	0.006% Lauryl maltoside NV-10	20 mM Chaps	20 mM Triton X-100	1 mM	5 mM	2 M Sarcosine
50% Glycerol	60 mM	Guanidine	0.04% PEG 3440	1 M NDSB	1 M	Formamide	No additive	0.9 M Arginine	0.5 M	1 M Proline	2 M Sarcosine
50 mM Arginine/ 50 mM Glutamate	1 M Glucose	10% Betaine	20% Sorbitol	1 M TMAO	0.5 M NaCl	0.5 M Sodium sulphate	1 M	2 M Guanidine	1 M	0.5 M	2 M Urea
1 M Urea	0.5 M Urea	0.006% Lauryl Maltoside	20 mM Chaps	20 mM Triton X-100	1 mM	Cyclohexanol	10% Ethanol	50% Glycerol	60 mM	10% Ethylene glycol	0.04% PEG 3440
1 M NDSB	1 M	NV-10	No Additive	0.9 M Arginine	0.5 M	Glycine	2 M Sarcosine	50 mM Arginine/ 50 mM Glutamate	1 M Glucose	10% Betaine	20% Sorbitol
1 M TMAO	0.5 M NaCl	0.5 M Sodium sulphate	1 M Ammonium sulphate	2 M Guanidine	1 M	Guanidine	2 M Urea	1 M Urea	0.5 M Urea	0.006% Lauryl maltoside	20 mM Chaps
20 mM Triton X-100	1 mM	5 mM	10% Ethanol	50% Glycerol	60 mM	Cyclodextrin	0.04% PEG 3440	1 M NDSB	1 M	NV-10	No Additive
0.1 M Mes, pH 5.8	0.1 M Capso, pH 9.5										

silver-stained to visualize bands. Samples for which no band was visible were not further analyzed.

#### Denaturing capillary electrophoresis

Denaturing capillary electrophoresis was performed on refolded protein samples subsequent to concentration and dialysis using an Agilent ALP-5100 instrument; 4  $\mu$ l of concentrated, dialyzed refolded protein solution was added to 2  $\mu$ l of reducing sample buffer and heated to 95 °C for 5 min. The heated solution was diluted with 24  $\mu$ l of water and analyzed using the protein analysis program on the ALP-5100 instrument on a HT-2 preprepared chip. Calibration of the elution time of various molecular weight proteins was carried out automatically using the supplied protein ladder.

#### Analytical size exclusion chromatography

Analytical size exclusion chromatography was performed using an Ettan LC system (GE Healthcare). Proteins were eluted from a prepacked analytical scale Superdex 75 column in 50 mM Tris, 150 mM NaCl, pH 9.0, at a flow rate of 50  $\mu$ l/min. Samples were diluted 1:1 with 0.2 mg/ml myoglobin in 10 mM Hepes before analysis and 25  $\mu$ l was loaded onto the column. Absorbance was monitored at 280 nm for p38 $\alpha$  and 410 nm for myoglobin. Peak areas were compared to a standard curve to calculate the p38 $\alpha$  protein concentration.

Analysis of the multimeric state of denatured inclusion body protein and denatured soluble protein was carried out as above, with the exception that a Superdex 200 column was used in place of the Superdex 75 and the column was equilibrated in 8 M urea, 50 mM Tris, 150 mM NaCl, pH 9.0, and eluted in this buffer.

#### Binding activity

The binding activity of refolded p38 $\alpha$  was assessed using a Biacore 3000 instrument (Biacore AS). An ureidoquinazoline target definition compound (UTDC) [19] was immobilized in a single lane of a CM5 chip (Biacore) using standard amine coupling kit (Biacore). The surface of the chip was activated by a solution of 1-ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride and *N*-hydroxysulfosuccinimide (NHS). This resulted in the creation of semistable amine-reactive NHS-ester on the chip surface. UTDC at 400  $\mu$ M was then flowed over the surface, with the free amine group reacting with the ester, resulting in the release of NHS and the formation of a stable amide bond between the UTDC and the chip surface. Unreacted NHS-esters were removed by treatment of the surface with ethanolamine. Control lanes were prepared in a similar manner, except that TDC was not added to the surface subsequent to activation. p38 $\alpha$  was flowed over the prepared surface in 10 mM Hepes, 150 mM NaCl, 0.05% P20 surfactant, 0.5% dimethyl sulfoxide. Response units



due to protein binding to immobilized UTDC were compared to response units due to binding to the control flow lane. To eliminate nonspecific binding, protein was injected in the presence of 10  $\mu$ M UTDC to abolish specific binding. Samples were maintained at 8 °C until analysis. Response units due to specific p38 $\alpha$  binding were compared to a standard curve prepared by measurement of binding of soluble, correctly folded p38 $\alpha$ . Each condition was analyzed in triplicate.

### Large-scale refolding

Larger scale refolding of p38 was carried out by drop-wise addition of solubilized inclusion bodies, at a protein concentration of 5 mg/ml, to renaturation buffer until the protein concentration reached 0.1 mg/ml and the urea concentration 160 mM. The buffer was vigorously stirred. Refolding was allowed to occur overnight at 4 °C under gentle stirring. The refolding solution was then concentrated 10-fold using an Amicon stirred cell with a 10-kDa cut-off ultrafiltration membrane (Millipore) and subsequently dialyzed against 10 mM Hepes, 150 mM NaCl, pH 7.4. A portion of the resulting solution was used for analysis via analytical size exclusion chromatography and assayed for binding activity by a surface plasmon resonance (SPR) method. The remaining protein was further concentrated to approximately 1 ml using an Amicon cell and Millipore membrane as previously described. The sample was then applied to a Superdex 75 12-mm  $\times$  60-cm column equilibrated in 50 mM Tris, 150 mM NaCl, pH 8.0, and was eluted from the column in the same buffer at a flow rate of 1 ml/min. Fractions identified as containing monomeric p38 $\alpha$ , by comparison of peak position with analytical size exclusion chromatography, were pooled and concentrated using a Centricon spin concentrator with a 10-kDa cut-off.

### Circular dichroism

Natively folded p38 $\alpha$  and refolded p38 $\alpha$  to be analyzed by circular dichroism were extensively dialyzed against 10 mM sodium phosphate buffer, pH 7.4, before analysis. Circular dichroism analysis of p38 $\alpha$  samples was performed on a Jasco J-810 spectropolarimeter, using a 1-mm-path-length quartz cuvette. Spectra were measured at a temperature of 20 °C, with a resolution of 0.5 nm. A scanning rate of 100 nm/min and a bandwidth of 1 nm were used, and eight scans were averaged per spectrum. Buffer blanks were subtracted from all spectra. The concentration of samples for CD analysis was confirmed by A<sub>280</sub> measurement before analysis was performed.

### Thermal melting analysis

Thermal melting analysis of p38 $\alpha$  was performed using a PCR-based thermal cycler (iCycler). Analyses were performed in a 96-well format (Bio-Rad) using sypro-orange

dye (Molecular Probes) at a 1:1000 final dilution. Concentrated p38 $\alpha$  was diluted to a concentration of 0.4 mg/ml and mixed in a 1:1 ratio with two-fold concentrated buffers to a final volume of 25  $\mu$ l. The plate was sealed with optical tape (Bio-Rad) and centrifuged at 1000 rpm for 1 min to remove air bubbles. A temperature gradient of 12 to 90 °C in 0.2 °C steps was applied to the sample. Sample fluorescence was recorded for 12 s at each temperature step. The midpoint of the transition, T<sub>m</sub>, was calculated through fitting the data obtained to a six-parameter curve describing thermal unfolding.

## Results

### Screen development

Solubilized inclusion body preparation was found to be of high purity (Fig. 1). The yield of purified, solubilized p38 $\alpha$  from inclusion bodies was 375 mg protein per litre of expression culture. Inclusion bodies and native folded protein were denatured in 8 M urea for 1 h at 30 °C in accordance with general practice expected to produce fully denatured protein. The denatured states of inclusion body p38 $\alpha$  and native denatured p38 $\alpha$  were compared by denaturing analytical size exclusion chromatography to show that the unfolded state of both preparations were similar. Both preparations produced comparable elution profiles, with a central peak at retention time  $\sim$ 0.7 CV (Fig. 2), which corresponds to the retention time measured for monomeric control p38 $\alpha$  (data not shown).

The screening methods used were rapid dilution of concentrated denatured protein into solutions containing various additives and at different pH values, followed by concentration of the protein and dialysis into a neutral buffer. Several analytical methods were developed to provide

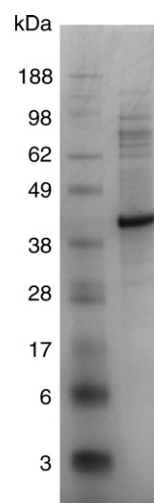


Fig. 1. Purity of p38 $\alpha$  inclusion body preparation assessed by reducing SDS-PAGE. 10% Bis-Tris gel run with NuPAGE Mes-SDS running buffer. Lane 1, molecular weight markers (SeeBlue Plus 2; Invitrogen); lane 2, p38 $\alpha$  inclusion body preparation in 8 M urea.

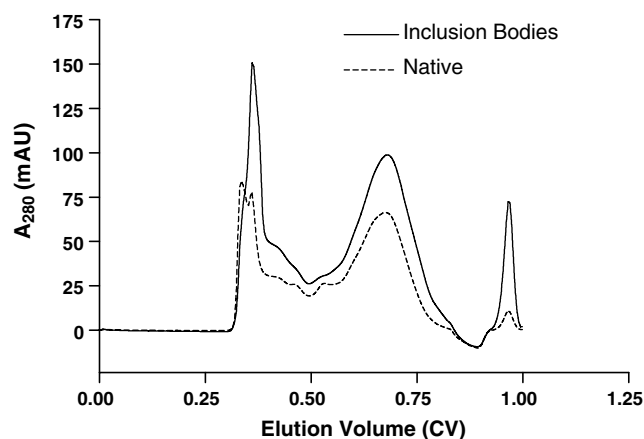


Fig. 2. Elution profile of native folded protein and inclusion body protein denatured with 8 M urea from analytical scale Superdex 200 column (volume 2.4 ml). Elution buffer 8 M urea, 50 mM Tris, 150 mM NaCl, pH 8.0; 125  $\mu$ g protein was injected.

an assessment of the yield and folded state of the protein. Purified folded p38 $\alpha$  was used to identify the limits of detection for each method and to demonstrate the repeatability of the methods. An overview of the refolding screen is outlined in Fig. 3. An analytical method was required to identify the effectiveness of the screen additives at maintaining p38 $\alpha$  in a soluble form after dilution of the denatured p38 $\alpha$  into the various conditions. SDS-PAGE, using a 96-well E-PAGE gel, was chosen because of its relative insensitivity to refolding additives compared to that of other analytical methods and the low limit of detection which was critical for detecting soluble protein at the low concentration present in the screen (maximum of 0.1 mg/ml). A series of dilutions of p38 $\alpha$  was run in quadruplicate on a single gel and stained. Using this dilution series, the lowest detected band contained  $\sim$ 16  $\mu$ g protein, which corresponded to a screen recovery of 1.25% (Fig. 4). We regarded this recovery a minimum for the refolding of protein at a large scale to allow production of sufficient protein for structural studies. This analytical method was used to restrict the number of conditions that were further analyzed. If a particular condition did not give a detectable band on the gel, i.e., the recovery of soluble protein was less than 1.25%, then the condition was not further analyzed. Control p38 $\alpha$  was used to test the ability of this analytical method to identify soluble p38 $\alpha$  in the presence of the various refolding additives (data not shown). Refolding conditions containing guanidine were found to precipitate the SDS in the sample buffer, preventing the analysis of these samples. All samples containing guanidine were therefore analyzed by further analytical methods, despite the lack of a visible band. The conditions identified as containing potentially useful yields of refolded protein by SDS-PAGE were concentrated 10-fold to give a maximal theoretical concentration of 1 mg/ml, which would be expected to be compatible with the use of less sensitive analytical methods with higher information content. Samples were subsequently dialyzed into 10 mM HEPES, 150 mM

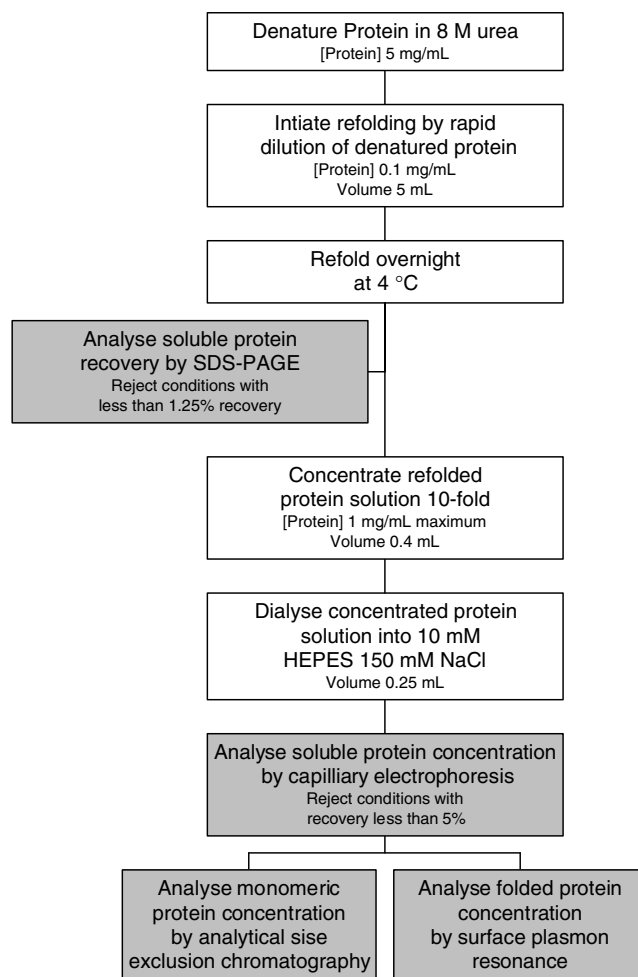


Fig. 3. Overview of the process of screening for conditions that enhance the refolding of p38 $\alpha$ . The process was designed to rapidly identify the yield of refolding in a large number of conditions. Analytical methods are shown in shaded boxes. Volume per condition at each stage and maximum protein concentration are also shown.

NaCl to remove additives that might interfere with these techniques.

Capillary electrophoresis was chosen as a second analytical technique to be applied to refolded protein produced by the screen to quantify the concentration of soluble protein after concentration and dialysis. This analytical method was chosen because it offers a high-throughput, quantitative method for analyzing the total protein remaining in solution. The earlier analytical method, SDS-PAGE, only allowed a semiquantitative assessment of protein concentration. The mean recovery was 237  $\mu$ g/ml with a standard error of 2%, demonstrating that the assay is reproducible. Example data from this analytical method are shown in Fig. 5. This technique was used as a high-throughput method to screen for the recovery of soluble protein after concentration and dialysis. The recovery of soluble protein identified was used as an additional restriction on the number of conditions analyzed by further techniques. Conditions that did not show a recovery of at least 5% were not analyzed further.

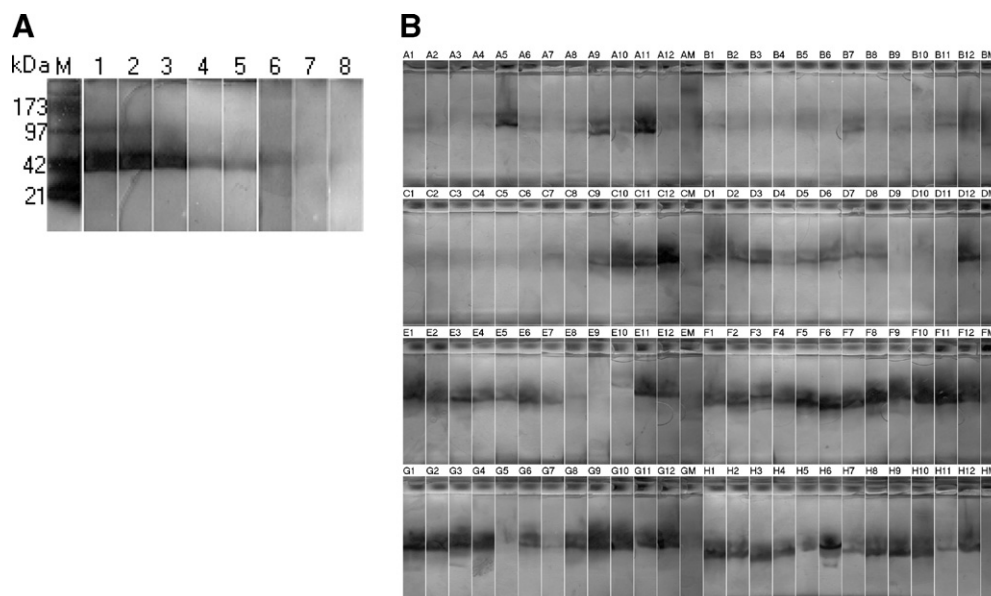


Fig. 4. (A) Dilution series of native folded p38 $\alpha$  run on 6% E-PAGE gel under reducing conditions and stained using silver staining. Lane M, molecular weight markers; lane 1, 1040  $\mu$ g; lane 2, 520  $\mu$ g; lane 3, 260  $\mu$ g; lane 4, 130  $\mu$ g; lane 5, 65  $\mu$ g; lane 6, 32.5  $\mu$ g; lane 7, 16  $\mu$ g; lane 8, 8  $\mu$ g. (B) Example of E-PAGE gel from screen; numbered lanes correspond to single conditions from screen (Table 1). Gel images were processed using E-Editor (Invitrogen) to align lanes.

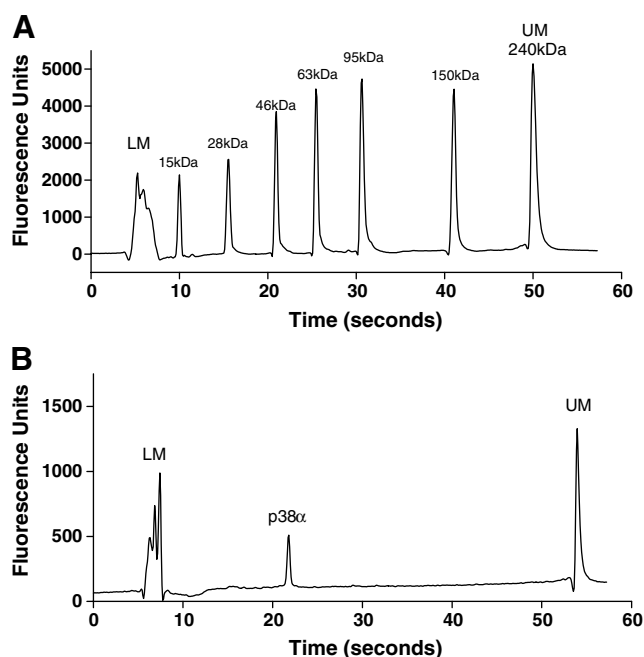


Fig. 5. Capillary electrophoresis analysis of refolded p38 $\alpha$ . (A) Protein ladder; LM, lower marker; UM, upper marker; molecular weight of markers indicated above peaks. (B) Example screen readout electropherogram. Typical protein analyzed was 0.02–1  $\mu$ g per condition.

In vitro refolding of a protein may result in intermediate species prone to intermolecular interactions with other p38 $\alpha$  molecules, resulting in the formation of soluble oligomers, that are not present in correctly folded preparations of native protein. Therefore, to determine whether the soluble protein is in the correct monomeric form, an analytical method that separates and quantifies the monomeric protein recovered from a refolding condition was

required. For these purposes, analytical size exclusion chromatography was used. To account for injection-to-injection variability due to the autosampler used on the system, it was necessary to include an internal standard. Myoglobin was selected as its elution peak can be identified by heme absorbance at 410 nm away from p38 $\alpha$  absorbance at 280 nm. A sample elution profile from a standard curve is shown in Fig. 6A. A standard curve of p38 $\alpha$  was analyzed via this method in triplicate to allow the concentration of monomeric protein to be calculated and to demonstrate the consistency of the assay. The assay was reproducible but proved to have a lowest limit of detection of 50  $\mu$ g/ml (Fig. 6B). A typical chromatogram obtained from a screen condition is shown in Fig. 6C.

The refolding of a protein may also result in a soluble form of the protein that is monomeric but not correctly folded. To identify the recovery of correctly folded p38 $\alpha$ , the binding of refolded protein to a UTDC known to bind to unphosphorylated correctly folded p38 $\alpha$  was quantified using SPR [35]. Conditions in the refolding screen are likely to lead to the formation of nonnative p38 species, which may interact nonspecifically with the compound at the surface of the chip. To control for this nonspecific binding, the binding of p38 $\alpha$  to the prepared surface was analyzed in the absence and presence of an excess of the UTDC in solution. The excess of free UTDC at 1700-fold above the affinity in solution [35] prevented the specific binding of p38 $\alpha$  to the prepared surface and so allowed the amount of nonspecific binding to be identified. A wide range of protein concentration was tested using solubly expressed purified p38 (Fig. 7A). The linear range of response was 10–90  $\mu$ g/mL of protein and a standard curve was calculated from a trip-



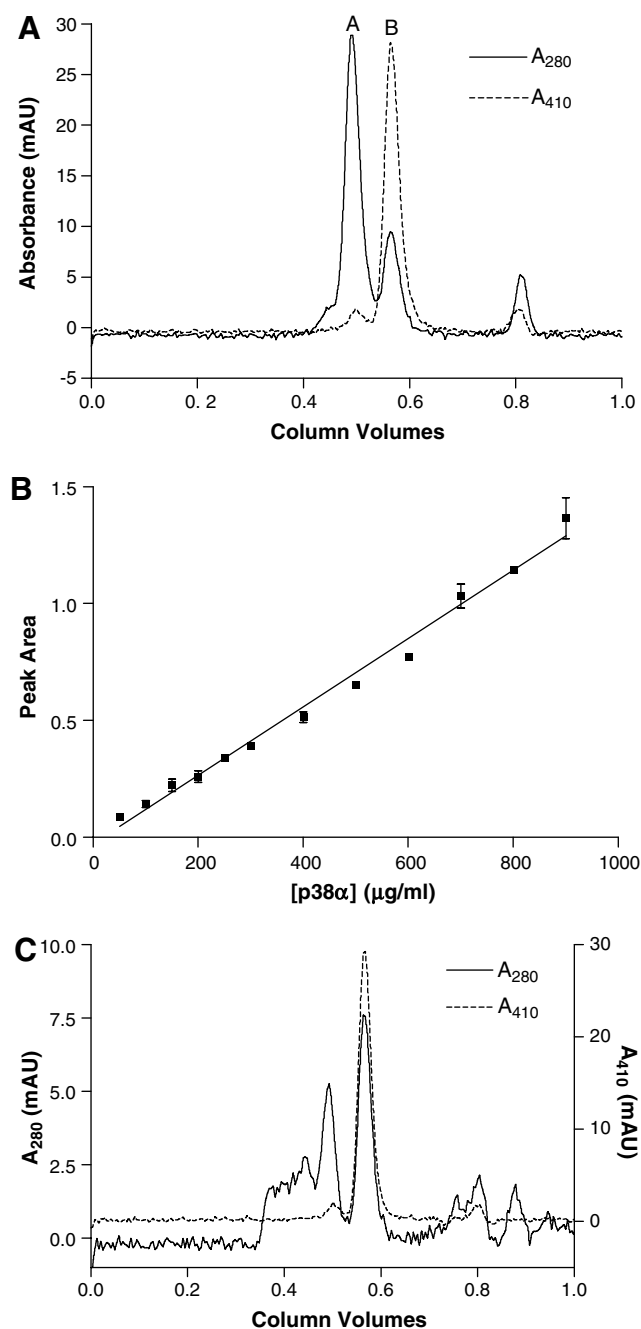


Fig. 6. Analytical size exclusion chromatography analysis of native folded and refolded p38 $\alpha$ . (A) Elution profile of 0.5 mg/ml p38 $\alpha$  and 0.1 mg/ml myoglobin. Peak A, p38 $\alpha$ , MW 41.3 kDa; peak B, myoglobin, MW 16.2 kDa. Elution time recorded in column volumes. (B) Standard curve of p38 $\alpha$  analyzed by analytical size exclusion chromatography. (C) Example of an elution profile from a screen; recovery shown 18%. Typical protein analyzed was 0.25–12.5  $\mu$ g.

licate analysis of control p38 $\alpha$  (Fig. 7B), showing that the analytical method is reproducible. The relative response units used to quantify the binding of p38 $\alpha$  to the prepared surface was the difference in response units in the presence and absence of free UTDC in solution at 10  $\mu$ M. Typical response curves obtained when analyzing a screen condition are illustrated in Fig. 7C.

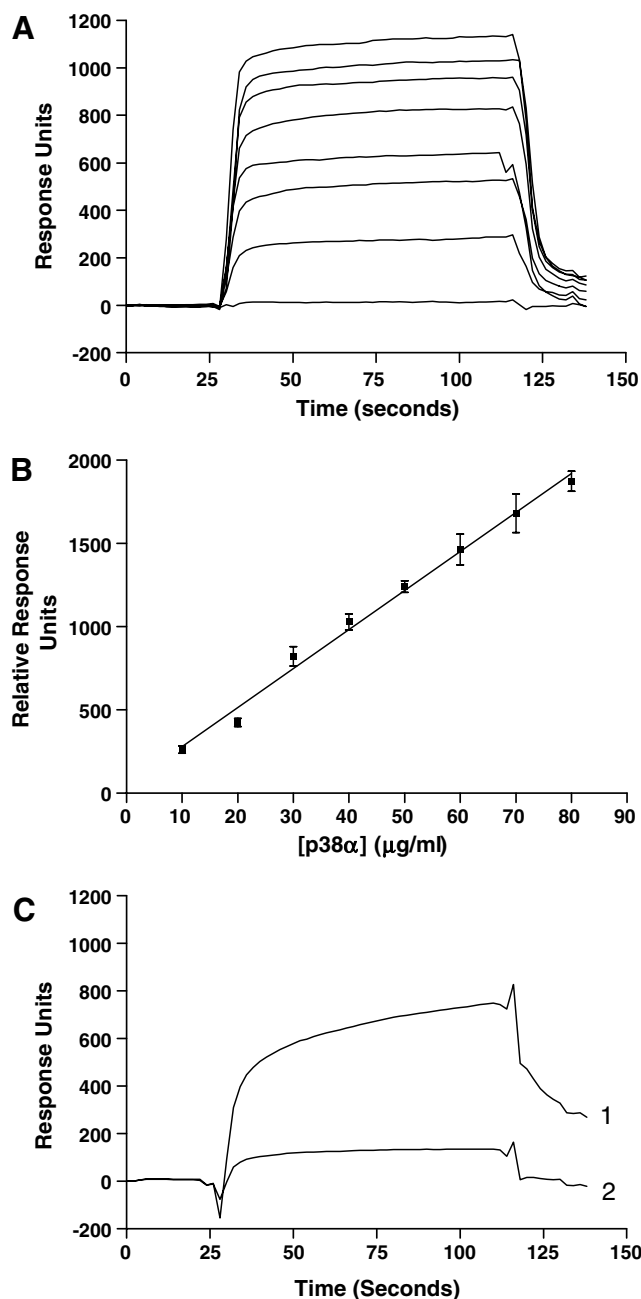


Fig. 7. Surface plasmon resonance analysis of refolded and native p38 $\alpha$  binding to immobilized TDC. (A) Series of injections of p38 $\alpha$  onto prepared chip showing increasing response with increasing p38 $\alpha$  concentration from 10 to 60  $\mu$ g/ml. (B) Standard curve for control binding experiments of soluble p38 $\alpha$ . Each point is the mean of three replicate analyses and error bars show the standard deviation. (C) Example analysis of screen condition post dialysis. Curve 1, refolded p38 $\alpha$ ; curve 2, refolded p38 $\alpha$  with 10  $\mu$ M TDC in solution. Typical protein concentration was 25–75  $\mu$ g/ml.

#### Refolding screen for p38 $\alpha$

Inclusion bodies and soluble p38 $\alpha$  were denatured with 8 M urea and tested in the screen. The refolding screen was performed in triplicate to provide reliable data on the refolding of p38 $\alpha$ . The recovery of soluble protein measured by capillary electrophoresis was generally higher than

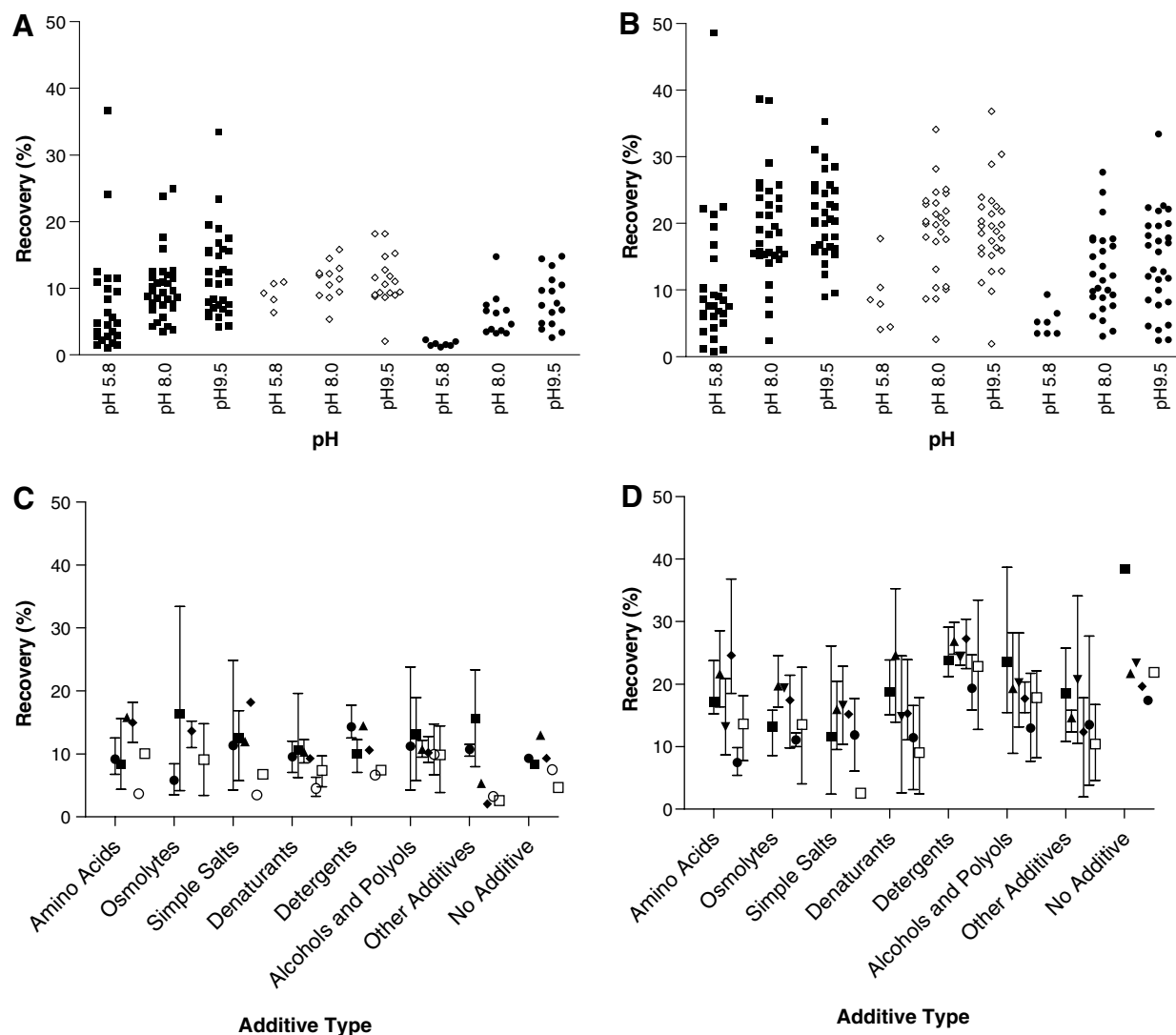


Fig. 8. Recoveries of soluble, monomeric, and folded protein obtained when refolding inclusion bodies (A, C) and soluble, denatured protein (B, D). For A and B results are for soluble protein recovery (■), monomeric protein recovery (◇), and folded protein recovery (●). Each data point represents a single refolding condition and is the mean of three replicates. For C and D results are for experiments at pH 8.0 for soluble protein recovery (■), pH 9.5 for soluble protein recovery (▲), pH 8.0 for monomeric protein recovery (▼), pH 9.5 for monomeric recovery (◆), pH 8.0 for folded protein recovery (○), and pH 9.5 for folded protein recovery (■). Each data point represents the average of the mean recoveries for all conditions at a particular pH and additive category (as described under Materials and Methods). The error bars indicate the range of values within a category.

the recovery of monomeric protein measured by analytical size exclusion chromatography and the recovery of folded protein measured by SPR was generally similar to the recovery of monomeric protein (Figs. 8A and 8B). The refolding yields obtained for each additive in the screen were reproducible across the three replicates, as indicated by the standard error of the mean shown in Fig. 9.

The refolding yields identified by the analytical methods used were compared for p38 $\alpha$  derived from inclusion body and for denatured native p38 $\alpha$ . Strong differences in the effect of the pH on the efficiency of the refolding were found. With both sources of refolded protein (inclusion body protein and denatured soluble protein), there were large drops in the recoveries of soluble protein, monomeric protein, and correctly folded protein at pH 5.8 compared

to the same recoveries at pH 8.0 and pH 9.5. This can be seen by comparing all of the recoveries obtained at a particular pH, as shown in Figs. 8A and 8B. The recoveries of soluble, monomeric, and folded protein, in the presence of refolding additives, cluster at a higher recovery at pH 8 and 9.5 compared to that at pH 5.8. The drop in the recovery of refolded protein was also generally consistent across the 31 refolding additives used in the screen (Figs. 8A, 8B, and 9). However, there were some exceptions; notably the additives proline and NDSB gave recoveries of soluble protein as high as 50% but did not give similarly high recoveries of monomeric and folded protein. To elucidate the effect of pH on the recovery of refolded protein, the thermal stability of natively folded p38 $\alpha$  was examined in a series of buffers (Fig. 10). The results showed that at pH 5.8 (in

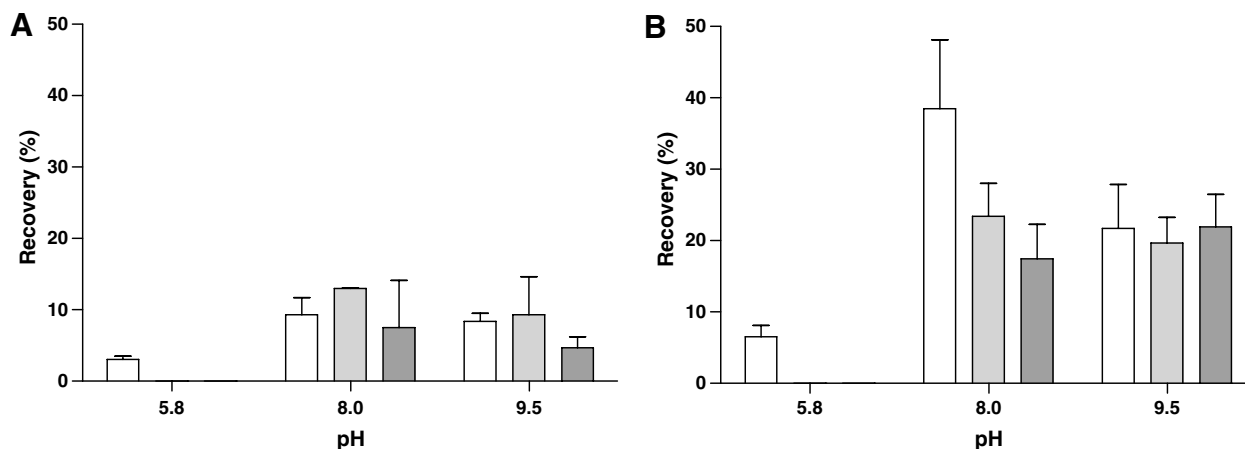


Fig. 9. Recovery of soluble (white), monomeric (light gray), and folded (dark gray) protein when refolding inclusion body protein (A) and denatured soluble protein (B) in buffer alone. Results are the mean of three replicates, and errors bars represent the standard error.

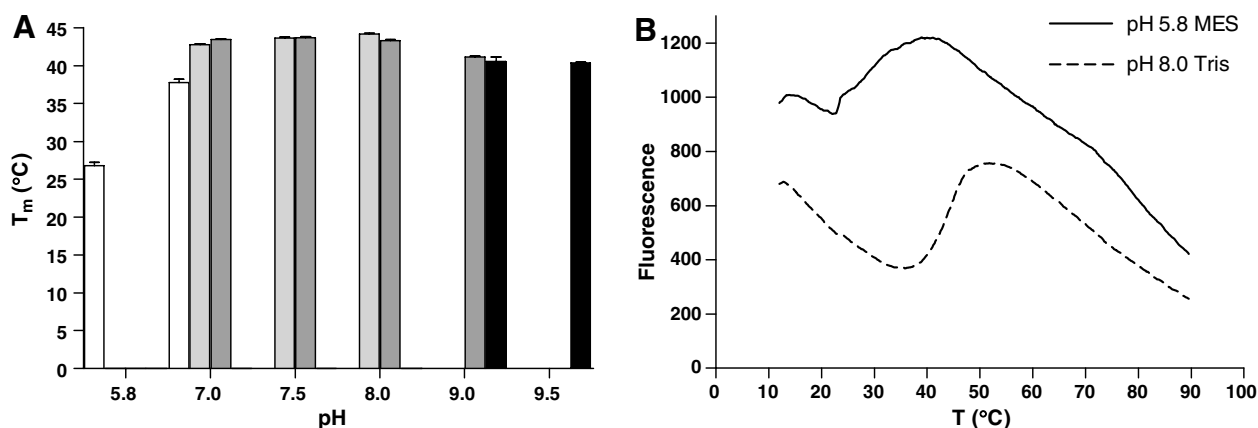


Fig. 10. Thermal melting analysis of native p38 $\alpha$  in a series of buffers and pHs. (A)  $T_m$  of p38 $\alpha$  at a range of pHs in various buffers, white (Mes); light gray (Hepes), darker gray (Tris), and black (Capso). Results are the mean of six replicates, and error bars represent the standard error. Concentrations of buffers were Mes 100 mM, Hepes 10 mM, Tris 50 mM, and Capso 100 mM. (B) Thermal melting profiles of p38 $\alpha$  in 100 mM Mes, pH 5.8, and in 50 mM Tris, pH 8.0. Protein used was 5  $\mu$ g per experiment.

Mes buffer) the protein has a large drop in  $T_m$ , of  $\sim 15$  °C, compared to higher pH conditions. Similar  $T_m$  values were obtained for p38 $\alpha$  in Mes and in other buffers at higher pH, indicating that the reduction in  $T_m$  is due to the pH as opposed to an effect of Mes. This indicates that pH 5.8 causes a destabilization of the structure of the native form of the protein.

High levels of recovery of refolded protein were observed in conditions that included no chemical refolding additive (Fig. 9). Recoveries of approximately 10% were observed with inclusion body protein and 20% with native, denatured protein. These recoveries indicate that p38 $\alpha$  refolds fairly easily, as evidenced by the high number of conditions under which high recoveries of refolded protein were achieved (Figs. 8A and 8B). Although p38 $\alpha$  was found to refold relatively easily at high pH, many of the refolding additives that were used enhanced the refolding of p38 $\alpha$ . Figs. 8C and 8D illustrate the effectiveness of the groups of additives at high pH. A wide spread of results was observed for some additive groups, particularly osmo-

lytes, denaturants, alcohols, and polyols. This wide range indicates that, though some additives gave high recoveries of refolded protein, a few additives in the group gave low recoveries. Lower average recoveries within these groups show that these groups, with the exception of usually a single exceptional additive, are not efficient at enhancing the refolding of p38 $\alpha$ . Examples of these exceptional additives were betaine and PEG, which showed enhancement of the refolding of p38 $\alpha$ , whereas similar additives with the same groups failed to do so. From Figs. 8C and 8D, it is clear that detergents are generally efficient in enhancing the refolding of p38 $\alpha$  compared to conditions lacking a specific refolding additive, as seen by the high average recovery with this particular additive group and narrow range of the results in this group.

A clear difference was noted in the recoveries of soluble, folded, and monomeric protein obtained when refolding inclusion body protein and native denatured p38 $\alpha$ . The recovery of refolded protein obtained with denatured soluble protein was generally approximately two-fold greater

Table 2

Comparison of refolding recoveries of p38 $\alpha$  from screen runs and scaled-up experiments

Refolding experiment	Renaturation buffer	Soluble recovery <sup>a</sup>	Monomeric recovery <sup>b</sup>	Folded recovery <sup>c</sup>
Screen	0.1 M Capso	9 $\pm$ 2	13.0 $\pm$ 0.1	8 $\pm$ 2
Screen	0.1 M Capso + 10% Betaine	15.3 $\pm$ 0.7	11.0 $\pm$ 0.7	15 $\pm$ 1
Scaled-up	0.1 M Capso	13	10	9
Scaled-up	0.1 M Capso + 10% Betaine	15	13	13

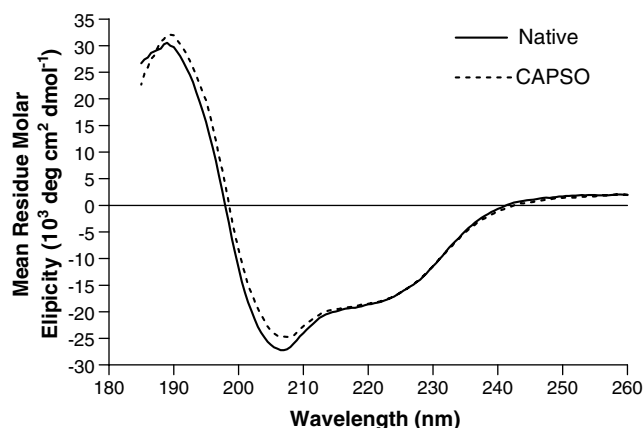
Protein recoveries are expressed as percentage of total initial protein. Refolding screen recoveries are mean of three experiments  $\pm$  SE of mean.<sup>a</sup> Soluble protein recovery was assessed by capillary electrophoresis.<sup>b</sup> Monomeric protein recovery was measured by analytical size exclusion chromatography.<sup>c</sup> Folded protein was evaluated by binding activity via a surface plasmon resonance method (see Materials and methods).

Fig. 11. Circular dichroism spectra of natively folded p38 $\alpha$  and p38 $\alpha$  derived from inclusion bodies refolded in 0.1 M Capso, 10% betaine, pH 9.5. Protein concentration was 0.1 mg/ml. Spectra were collected at 20 °C and are the average of eight scans.

than the recovery of refolded protein from inclusion body protein at high pH (Figs. 8A and 8B). This was consistent across the full range of additives. There was no difference in the recoveries obtained at low pH. Despite lower yields (Fig. 8), the recoveries of refolded protein obtained with inclusion-body-derived p38 $\alpha$  were still consistent with the production of significant amounts of protein for structural studies.

The refolding of p38 $\alpha$  was performed at a larger scale, and the refolded protein was purified to allow the secondary structure content of the refolded material to be examined and to determine the percentage of protein correctly folded. The examination of the secondary structure content of the refolded protein provides evidence that the screen is identifying conditions producing correctly folded p38 $\alpha$ . This refolding was carried out at a 20-mg scale by dropwise addition of p38 $\alpha$  to refolding buffer under the same conditions as in the refolding screen. Six conditions were selected from the screen that had shown the highest recoveries of soluble, monomeric, and folded protein. The conditions from the screen that were tested were 0.1 M Capso, pH 9.5; 0.1 M Capso, 10% betaine, pH 9.5; 0.1 M Capso, 10% ethylene glycol (v/v), pH 9.5; 0.1 M Capso, 0.04% PEG 3440 (w/v), pH 9.5; 50 mM Tris, pH 8.0; and 50 mM Tris, 1 mM cyclohexanol, pH 8.0. The recoveries of protein from the scaled-up refolding were found to be sim-

ilar to those obtained in the refolding screen (Table 2), showing that scaling up did not affect the yields obtained. The far-UV CD spectra of p38 $\alpha$  natively folded and p38 $\alpha$  refolded from inclusion bodies in the presence of 0.1 M Capso, 10% betaine (w/v), pH 9.5 (after dialysis; see Materials and methods) were identical (Fig. 11), indicating that the secondary structure content of the natively folded and the refolded protein was the same. The other conditions that were scaled up also showed similar far-UV CD spectra (data not shown), indicating that the secondary structure content of these refolded protein samples was also similar to that of the native state.

To identify the percentage of refolded protein that was correctly folded, the inhibition of refolded p38 $\alpha$  to UTDC by the known inhibitor of unphosphorylated p38 $\alpha$ , SB202190 [36] was determined. The binding was measured via SPR and the results fitted to a stoichiometric binding model, as shown in Fig. 12. From these data the concentration of inhibitor binding sites was estimated. Because SB202190 binds to the ATP binding site in p38 $\alpha$ , this inhibition assay can measure the amount of correctly folded protein. The refolded p38 $\alpha$  was therefore found to be 92.7  $\pm$  4% folded. These results show that the screen is able to identify conditions producing correctly folded protein.

The refolding of p38 $\alpha$  from inclusion bodies at a 20 mg scale was used to calculate the yield of soluble, correctly folded protein that could be achieved by this method. An average yield of 13% monomeric, folded protein was obtained from the method used. This corresponded to a recovery of monomeric, folded protein of  $\sim$  50 mg protein per litre of expression culture. This yield would be suitable for the production of protein for structural studies by a refolding procedure.

## Discussion

The refolding of protein kinases is a challenging process with little reported success in the production of protein suitable for structural studies. Overcoming this challenge is likely to require the screening of a large number of conditions, to identify those that allow refolding of the protein. It is also critical that the analytical methods used to quantify refolded protein are robust and sensitive.

Commercial refolding screens usually suggest the use of a functional assay for the target protein as a primary

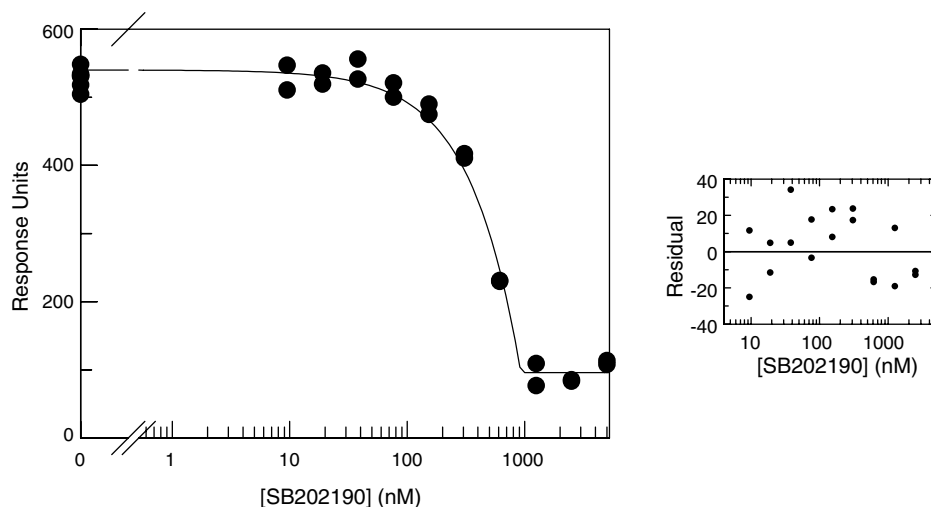


Fig. 12. Inhibition of the binding of refolded p38 $\alpha$  to immobilized UTDC by SB202190 detected with SPR. Data fitted to a stoichiometric binding model, with response unit background at 96 and residuals shown on the right. Protein concentration was 1  $\mu$ M.

method of quantifying the yield of refolding. However, for kinases this is not a viable option for several reasons. First, protein kinases usually require phosphorylation by another protein kinase on key residues for activation. Inclusion-body-derived protein kinases are unlikely to be appropriately phosphorylated and therefore production of activated kinase would require an additional activation step. This is incompatible with high-throughput screening for refolding conditions. In addition, some of the refolding additives may interfere with the phosphorylation of the refolded kinase, leading to false negatives. Second, many structural approaches in the pharmaceutical industry make use of mutants, such as kinase-dead mutants, that show no catalytic activity. Because of these problems, biophysical methods of protein characterization are required for analyzing the refolded protein.

The refolding of proteins generally requires a low protein concentration to reduce aggregation; this means that the analytical techniques used must be highly sensitive. Methods that had low limits of detection, corresponding to minimum useful yields for refolding protein for structural studies, were chosen. For example, the limit of detection for the staining of SDS–PAGE gels corresponds to a refolding yield of 1.25%. We considered this level of recovery a minimum for which refolding would be viable for the generation of protein for structural studies.

The refolding screen makes use of several refolding additives, some of which are present at high concentrations. These additives may interfere with particular analytical methods. For example, additives such as guanidine, which cause large changes in the refractive index of the sample, interfere with SPR measurements. Detergents also cause large baseline deflections for denaturing capillary electrophoresis and analytical size exclusion chromatography. The analytical methods selected were chosen to address these concerns. An initial SDS–PAGE screening method that is relatively insensitive to the additives used was

selected and the refolding additives were removed by dialysis into a neutral buffer.

A hierarchical analysis design was used to eliminate unproductive conditions early in the process (Fig. 3) by examining total soluble protein content following refolding and dialysis. Refolding conditions with a low recovery of refolded protein were first eliminated by SDS–PAGE analysis of the refolded protein. Capillary electrophoresis was also used to eliminate conditions with low recoveries of refolded protein subsequent to concentration and dialysis, with conditions with recoveries of less than 5% being eliminated to allow manageable numbers of samples to be progressed to the final two analytical methods.

We demonstrated a series of analytical methods for analyzing the refolded protein obtained from the screen and showed that these methods have low limits of detection. We showed that SDS–PAGE analysis of the refolded protein was able to identify yields of soluble protein in the presence of refolding additives of at least as low as 1.25% or 1.25  $\mu$ g/ml (Fig. 4). Analytical size exclusion chromatography was shown to be able to quantify the recovery of monomeric protein to a limit of 50  $\mu$ g/ml or a recovery of monomeric protein of 5% (Fig. 6). Binding analysis by SPR of refolded p38 $\alpha$  was shown to be able to identify the yield of folded p38 $\alpha$  to a limit of 1% or 10  $\mu$ g/ml (Fig. 7). These limits of detection are expected to be compatible with screening of kinases with lower yields of refolding than those shown by p38 $\alpha$ .

Strong differences in the yields of refolding of inclusion body protein and denatured soluble protein were identified at higher pH values (Figs. 8A and 8B). The reasons for this difference are unclear. There is no discernable difference in the monomeric state of inclusion body protein and native denatured protein in 8 M urea that can be identified by analytical size exclusion chromatography (Fig. 2). This demonstrates that there is no difference in the aggregation states of the denatured protein from both sources. Some



amount of secondary structure is known to be transiently present in the unfolded state of proteins in high concentrations of chaotropic denaturants, such as guanidine and urea [13,37]. However, this is primarily sequence dependant and would be expected to be the same in both protein preparations [13,37]. It is noteworthy that no additional contaminant proteins of inclusion body protein and soluble protein were identified by SDS–PAGE. (Fig. 1).

The refolding screens carried out identified that there is a strong dependence of the refolding ability of p38 $\alpha$  on pH. At the low pH used in the screen, the refolding of p38 $\alpha$  was inefficient, with low yields of soluble protein identified by capillary electrophoresis (Figs. 8A and 8B). The thermal melting results showed that the inefficiency of refolding at lower pH values is due to a destabilization of the native structure of p38 $\alpha$  in low-pH buffers (Fig. 10). We propose that this destabilization is due to changes in the ionization of residues in the protein, since the *pI* of p38 $\alpha$  is around pH 5.5. Whether this effect is common to serine/threonine protein kinases or is a specific effect for p38 $\alpha$  is not known. However, a low *pI* is not a common feature of all protein kinases, with several related protein kinases having calculated *pI*s of at least 8.5.

Conditions identified by the screen leading to comparatively high recoveries of refolded protein were used to refold protein at a higher scale than was performed in the screen to provide sufficient protein for further analysis and to determine whether there were significant differences in the refolding yields obtained at a larger scale. The refolded protein from this higher-scale refolding experiment was used for the structural analysis of the refolded protein. We showed that the far-UV CD spectra of control natively folded p38 $\alpha$  and refolded p38 $\alpha$  were similar (Fig. 10), with a calculated secondary structure content analogous to the published structure of p38 $\alpha$  [38]. We also showed the correctly folded protein to be  $92.7 \pm 4\%$  of the purified, refolded protein by inhibitor binding studies (Fig. 12). This indicates that the protein obtained from refolding inclusion bodies is correctly folded. The similarity of the refolding yields obtained from the refolding screen and from the larger-scale refolding performed here demonstrates that the refolding screen is capable of identifying refolding conditions that are transferable to high-scale refolding for the production of refolded protein for structural studies. This transferability is key for the screen to be useful for identifying refolding conditions for other, more challenging protein kinases that are produced in an insoluble form in *E. coli*.

The screening system that we have established could be adapted to the refolding of other, more challenging protein kinases in addition to p38 $\alpha$ . These protein kinases may require the identification of individually specific or generic compounds capable of being immobilized on available SPR analysis chips, whilst retaining the ability to specifically bind the kinase under study. Specific ligands for many of these kinases and generic compounds capable of binding a wide range of kinases exist, making our screening system a viable strategy for other kinases. The chemical additives

and buffer conditions of the screen may also require adjustment. Details of the folding pathways and intermediates of protein kinases in general are not currently known. It is, therefore, not known whether protein kinases, and in particular serine-threonine kinases, will refold in similar conditions. It is not known whether protein kinases other than p38 $\alpha$  will show similar poor refolding recoveries in low-pH conditions or will have their refolding enhanced or indeed permitted by similar chemical additives. This will require testing of the refolding screen with example protein kinases and the testing of different conditions to create a generic refolding screen for protein kinases. The results of this testing and the patterns of refolding seen with different protein kinases will reveal important information about the similarities in the folding pathways of the protein kinases tested and for protein kinases in general. Analysis of conserved residues may also reveal important residues in the folding of protein kinases based on structural alignment of sequences and the results of similar refolding screens on additional kinases.

We have established a screening system for the refolding of a model protein kinase. We identified and described a series of specific analytical methods that quantify the yields of refolded protein and identified the oligomeric state of the protein and whether the refolded protein has adopted the correct fold. The yield of refolded protein depends strongly on the pH at which the protein is refolded, and the source of the protein to be refolded also strongly affects the yield of refolded protein. In addition, we have shown that refolding can be performed at larger scale, resulting in correctly folded protein without reducing the recovery yields. This initial refolding screening system can be adapted and applied to the refolding of other pharmaceutically important protein kinases, leading to the identification of methods for the production of protein in quantities required for essential structural analysis that underpin drug discovery projects.

## Acknowledgments

The work was supported by the Biotechnology Biological Science Research Council (BBSRC studentship) and AstraZeneca, through a Warwick-AstraZeneca CASE studentship to R.H.C. We thank Malcolm Anderson for support on analytical size exclusion chromatography and ESI-MS, Paul Hawtin for assistance on capillary electrophoresis, and Gareth Davies for assistance with the SB202190 inhibition assay and corresponding data analysis.

## References

- [1] Y.M. Lee, P. Sininski, *Cell Cycle* (2006) 2110–2114.
- [2] J. Westra, P.G. Limburg, *Mini. Re. Med. Chem.* (2006) 867–874.
- [3] J. Raingeaud, A.J. Whitmarsh, T. Barrett, B. Derijard, R.J. Davis, *Mol. Cell. Biol.* (1996) 1247–1255.
- [4] J. Rouse, P. Cohen, S. Trigon, M. Morange, A. Alonso-Llamazares, D. Zamanillo, T. Hunt, A.R. Nebreda, *Cell* (1994) 1027–1037.
- [5] J. Han, J.D. Lee, L. Bibbs, R.J. Ulevitch, *Science* (1994) 808–811.
- [6] C.B. Breitenlechner, D. Bossemeyer, R.R. Eng, *Biochem. Biophys. Acta* (2005) 38–49.

- [7] F. Baneyx, M. Mujacic, *Nat. Biotechnol.* (2004) 1399–1408.
- [8] D. Ami, A. Natalello, G. Taylor, G. Tonon, S.M. Doglia, *Biochem. Biophys. Acta* (2005) 793–799.
- [9] A. Villaverde, M.M. Carrió, *Biotechnol. Lett.* (2003) 1385–1395.
- [10] S. Ventura, A. Villaverde, *Trends Biotechnol.* (2006) 179–185.
- [11] D.J. Owen, A.C. Papageorgiou, E.F. Garman, M.E.M. Nobel, L.N. Johnson, *J. Mol. Biol.* (1995) 374–381.
- [12] A. Caballero-Herrera, K. Nordstrand, K.D. Berndt, L. Nilsson, *Biophys. J.* (2005) 842–857.
- [13] K.A. Dill, D. Shortle, *Annu. Rev. Biochem.* (1991) 795–825.
- [14] C.M. Dobson, *Se. Cell. Dev. Biol.* (2004) 3–16.
- [15] A.J. Caplan, A.K. Mandal, M.A. Theodoraki, *Trends Cell Biol.* (2007) 87–92.
- [16] J.R. Blackwell, R. Horgan, *FEBS Lett.* (1991) 10–12.
- [17] G. Georgiou, P. Valax, *Methods Enzymol.* (1999) 48–58.
- [18] G. Davies, Ph D Thesis, University of Warwick, 2004.
- [19] M.S. Willis, J.K. Hogan, P. Prabhakar, X. Liu, K. Tsai, Y. Wei, T. Fox, *Protein Sci.* (2005) 1818–1826.
- [20] T. Arakawa, D. Ejima, K. Tsumoto, N. Obeyama, Y. Tanaka, Y. Kita, S.N. Timasheff, *Biophys. Chem.* (2007) 1–8.
- [21] B.M. Baynes, D.I. Wang, D.L. Trout, *Biochemistry* (2005) 4919–4925.
- [22] T. Arakawa, S.N. Timasheff, *Biophys. J.* (1985) 411–414.
- [23] T. Arakawa, S.N. Timasheff, *Biochemistry* (1982) 6536–6544.
- [24] N. El Kadi, N. Taulier, J.Y. Le Huerou, M. Gindre, W. Urbach, I. Nwigwe, P.C. Kahn, M. Waks, *Biophys. J.* (2006) 3397–3404.
- [25] C.C. Mello, D. Barrick, *Protein Sci.* (2003) 1522–1529.
- [26] C. Nishimura, V.N. Uverski, A.L. Fink, *Biochemistry* (2001) 4113–4128.
- [27] M. Yasuda, Y. Murakami, A. Sowa, H. Ogino, H. Ishikawa, *Biotechnol. Prog.* (1998) 601–606.
- [28] D.B. Wetlaufer, Y. Xie, *Protein Sci.* (1995) 1535–1543.
- [29] K. Ruan, C. Xu, T. Li, J. Li, R. Lange, C. Balny, *Eur. J. Biochem.* (2003) 1654–1661.
- [30] S. Machida, S. Ogawa, S. Xiaohua, T. Takaha, K. Fujii, K. Hayashi, *FEBS Lett.* (2000) 131–135.
- [31] M. Silow, M. Oliveberg, *J. Mol. Biol.* (2003) 263–271.
- [32] Y. Chong, H. Cheng, *Biotechniques* (2000) 1166–1167.
- [33] M.E. Goldberg, N. Eeret-Bezancon, L. Vuillard, T. Rabilloud, *Fold. Des.* (1996) 21–27.
- [34] G. Zardeneta, P.M. Horowitz, *J. Biol. Chem.* (1992) 5811–5816.
- [35] J.E. Sullivan, G.A. Holdgate, D. Campbell, D. Timms, S. Gerhardt, J. Breed, A.L. Breeze, A. Bermingham, R.A. Pauptit, R.A. Norman, K.J. Embrey, J. Read, W.S. VanScyoc, W.H. Ward, *Biochemistry* (2005) 16475–16490.
- [36] B. Frantz, T. Klatt, M. Pang, J. Parsons, A. Rolando, H. Williams, M.J. Tocci, S.J. O’Keefe, E.A. O’Neill, *Biochemistry* (1998) 13846–13853.
- [37] D. Shortle, M.S. Ackerman, *Science* (2001) 487–489.
- [38] K.P. Wilson, M.J. Fitzgibbon, P.R. Caron, J.P. Griffith, W. Chen, P.G. McCaffrey, S.P. Chambers, M.S.S. Su, *J. Biol. Chem.* (1996) 27696–27700.