

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

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/131596>

Copyright and reuse:

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

Please scroll down to view the document itself.

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

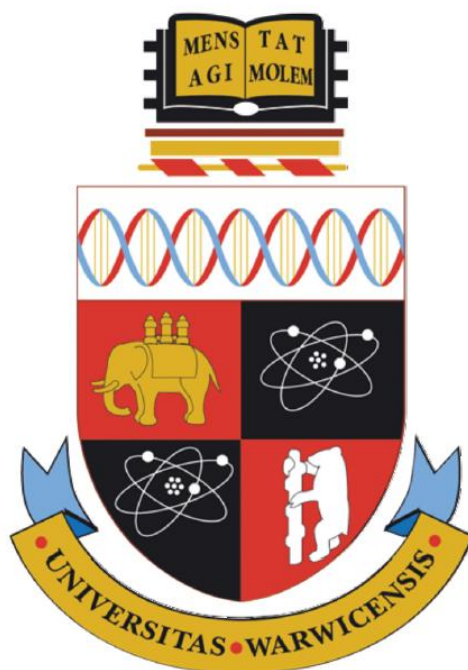
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Engineering self-sufficiency and broadened substrate scope into indole-nitrating cytochrome P450 TxtE

Rakesh Saroay

**Thesis submitted in partial fulfilment of the requirements for the degree
of Doctor of Philosophy in Chemistry**



**University of Warwick
Department of Chemistry
September 2018**

Contents

Contents.....	i
List of figures and tables.....	vi
Acknowledgements.....	xv
Declaration.....	xvi
Abbreviations.....	xvii
Abstract.....	xviii
Chapter 1: Introduction.....	1
1.1 Natural products.....	2
1.2 Natural nitro products.....	3
1.3 Biosynthetic routes to nitro groups.....	4
1.3.1 Enzymatic amine oxidation.....	4
1.3.2 Direct enzymatic nitration.....	6
1.4 Industrial importance of nitro compounds.....	7
1.5 Cytochrome P450 enzymes.....	8
1.5.1 Role of cytochrome P450 enzymes in natural product biosynthesis.....	9
1.5.2 CYP catalytic mechanism.....	11
1.5.3 Redox partners for CYPs.....	13
1.5.4 CYP-Reductase fusion proteins.....	13
1.5.5 CYPs sequence and structure.....	15
1.6 TxtE.....	17
1.6.1 Role in the biosynthesis of thaxtomin A.....	17
1.6.2 TxtE structure and sequence.....	20
1.6.2.1 TxtE flexible loop.....	23
1.6.3 TxtE homologues and mutants.....	24
1.6.4 TxtE substrate scope.....	25
1.7 Introduction to enzyme engineering.....	27
1.7.1 Directed evolution.....	27
1.7.1.1 Random mutagenesis.....	28
1.7.1.2 Targeted mutagenesis.....	28
1.7.1.3 Examples of the directed evolution of CYPs.....	28
1.8 Project aims.....	30

Chapter 2: Investigating wild-type TxtE and creating a self-sufficient fusion protein.....	31
2. Overview.....	32
2.1 Purifying TxtE and reconstructing the in vitro activity.....	32
2.2 UV-Visible spectroscopic characterisation of TxtE.....	33
2.3 Analysis of the TxtE-catalysed nitration reaction.....	34
2.4 Investigating the substrate scope of TxtE.....	35
2.5 Preparation of self-sufficient TxtE-reductase fusion proteins.....	39
2.5.1 TxtE-RhFRED.....	39
2.5.2 TxtE-BM3R.....	42
2.6 Flexibility in the regioselectivity of the TxtE-catalysed reaction.....	45
2.6.1 Scale-up of 4-methyl-DL-Trp nitration reaction.....	47
2.6.2 Comparison to a synthetic standard of 4-methyl-5-nitrotryptophan.....	48
2.7 TxtE-BM3R substrate screen.....	50
2.8 Conclusions.....	52
Chapter 3: Creating libraries of TxtE variants and identification of mutant with increased L-Trp nitration activity.....	54
3. Introduction.....	55
3.1 Analysis of TxtE structure.....	55
3.1.1 Active site.....	55
3.1.2 Flexible loop.....	56
3.1.3 Helix and loop influencing residues.....	57
3.2 Mutation techniques and methods.....	58
3.2.1 Site-saturation library preparation methods.....	58
3.2.2 The 22c-trick method.....	59
3.2.3 Reducing the library screening effort.....	60
3.2.4 Site-saturation mutagenesis techniques.....	61
3.2.5 Quick quality control.....	62
3.2.6 Library screening protocol.....	63
3.3 Initial Validation Screen – H176 library.....	64
3.3.1 Verifying the analytical method.....	64
3.3.2 H176 library preparation.....	65
3.3.3 His176 library screening.....	66
3.4 Analysis of site-saturation library quality.....	68
3.5 High-throughput screening of TxtE-BM3R mutants for improved activity with L-Trp.....	70
3.5.1 Selected positive variants.....	74
3.6 Conclusions.....	75

Chapter 4: Identification of a TxtE-BM3R variant capable of tryptamine nitration.....	77
4. Introduction.....	78
4.1 Screening of R59 mutant library to identify tryptamine nitrating variant...	79
4.2 Characterisation of TxtE-BM3R(R59C).....	81
4.3 Activity of TxtE-BM3R(R59C) with L-Trp.....	83
4.4 Expression and activity of TxtE(R59C) heme domain.....	83
4.5 UV-Vis analysis of substrate binding to TxtE-BM3R and the R59C variant.....	85
4.6 Molecular docking simulations of tryptamine binding in the TxtE-BM3R(R59C) active site.....	87
4.7 Determining the regioselectivity of tryptamine nitration.....	88
4.7.1 Scale-up of the tryptamine nitration reaction.....	88
4.7.2 Synthesis of 4-nitrotryptamine synthetic standard.....	88
4.7.3 Comparison with the authentic standard.....	89
4.8 Conclusions.....	91
Chapter 5: Substrate scope and mechanistic investigation of TxtE-BM3R(R59C).....	92
5.1. Substrate scope of TxtE-BM3R(R59C).....	93
5.1.1 Modification to the 5-position of tryptamine.....	94
5.1.2 Modification of the amino acid of tryptamine.....	95
5.1.3 Tricyclic analogues.....	96
5.1.4 Modification of the indole nitrogen to probe the mechanism of nitration.....	98
5.2 Mechanistic analysis of 2-(benzofuran-3-yl)ethanamine 137 nitration by TxtE-BM3R(R59C).....	98
5.3 Analysis of N-methyl-L-tryptophan binding to TxtE-BM3R(R59C).....	101
5.4 Conclusions.....	102
Chapter 6: Conclusions and future work.....	104
6.1 Completion of aims.....	105
6.1.1 Optimising TxtE for high-throughput or large scale nitration.....	105
6.1.2 Creation of mutant libraries and screening of variants for improved L-Trp nitration.....	105
6.1.3 Identification and substrate tolerance of tryptamine nitrating variant.....	106
6.1.4 Mechanistic investigations into TxtE-catalysed nitration.....	106
6.2 Future work.....	107
6.2.1 Identify the products of TxtE-BM3R and R59C variant-catalysed nitration reactions.....	107
6.2.2 Investigate mutant libraries for activity with other non-natural substrates.....	107
6.2.3 Engineer TxtE-BM3R for further broadening of substrate scope	108

6.2.4 Investigate RufO, a tyrosine nitrating CYP from Rufomycin biosynthesis.....	108
6.2.5 Investigate the biosynthesis of nitro-tetraene-containing natural products.....	109
Chapter 7: Experimental procedures.....	110
7.1 General.....	111
7.1.1 Chemicals and reagents.....	111
7.1.2 Instrumentation.....	111
7.1.3 Plasmids, bacterial strains and enzymes.....	112
7.1.4 Biological reagents and media.....	112
7.2 Biological procedures.....	113
7.2.1 Preparation of pET151_ <i>txtE_rhfred</i>	113
7.2.2 General procedure for protein overproduction.....	115
7.2.3 Protein analysis and characterisation.....	116
7.2.4 Overproduction of TxtE-BM3R in 96-well plates.....	117
7.2.5 Generation of TxtE-BM3R mutant libraries.....	117
7.2.6 Preparation of TxtE-BM3R_R59C variants.....	122
7.3. UV-vis characterisation of heme-containing proteins.....	122
7.3.1 UV-Vis spectroscopic analysis of TxtE.....	122
7.3.2 Qualitative binding assay – determining protein-substrate binding type.....	122
7.3.3 General procedure for the determination of protein-substrate binding constants.....	123
7.3.4 Qualitative binding assay of TxtE-BM3R and TxtE-BM3R(R59C) with 2-(benzofuran-3-yl)ethanamine as the substrate.....	123
7.4 In vitro assays, optimisation and screening methods.....	124
7.4.1 TxtE-catalysed nitration of L-Trp and substrate analogues.....	124
7.4.2 TxtE-reductase fusion protein catalysed nitration of L-tryptophan using purified protein.....	124
7.4.3 TxtE-BM3R reaction buffer and NADPH regeneration screen.....	124
7.4.4 High-throughput screening and analysis of TxtE-BM3R activity assays with L-tryptophan substrate analogues.....	125
7.4.5 Scale-up of the nitration of 4-methyl-DL-tryptophan by TxtE-BM3R using clarified cell lysate.....	125
7.4.6 High-throughput screening and analysis of the TxtE-BM3R mutant libraries.....	126
7.4.7 In vitro activity assays of TxtE-BM3R(R59C) with L-tryptophan and tryptamine.....	127
7.4.8 In vitro activity assays with TxtE(R59C) and redox partner proteins.....	127

7.4.9 In vitro screening and analysis of TxtE-BM3R(R59C) with tryptamine substrate analogues.....	128
7.4.10 Scale-up of the tryptamine nitration reaction with purified TxtE- BM3R(R59C) protein.....	128
7.5 Computational methods.....	129
7.5.1 Molecular docking studies.....	129
7.6 Chemical synthesis.....	130
7.6.1 Synthesis of 4-methyl-5-nitro-L-tryptophan 102.....	130
<i>N</i> -acetyl-4-methyl-5-nitro-D,L-tryptophan108.....	130
4-methyl-5-nitro-L-tryptophan 102.....	131
7.6.2 2-(benzo[b]thiophen-3-yl)ethanamine 138.....	131
7.6.3 4-nitrotryptamine 124.....	132
References.....	134
Appendix.....	142

List of figures and tables

Abstract figure. Nitration of L-Trp by wild-type TxtE and nitration of tryptamine by a TxtE-BM3R variant.....	xviii
Figure 1.1. Natural product structures A) Semi-synthetic drug simvastatin 1 , derived from the polyketide lovastatin 2 isolated from <i>Aspergillus terreus</i> . B) Alkaloid vincristine 3 isolated from <i>Catharanthus roseus</i> . C) Glycopeptide vancomycin 4 isolated from <i>Amycolatopsis orientalis</i>	2
Figure 1.2. Structures of selected natural nitro products.....	3
Figure 1.3. Biosynthetic <i>N</i> -oxygenases A) PrnD-catalysed nitration in the biosynthesis of pyrrolnitrin 8 B) AurF catalysed nitration in the biosynthesis of <i>para</i> -nitrobenzoate 18 (pABA) from <i>para</i> -aminobenzoate 17 (pNBA) in aureothin biosynthesis. C) Proposed pathway to <i>para</i> -nitrobenzylamine 23 via the <i>N</i> -oxidation of <i>para</i> -aminobenzylamine 19 by PrnD.....	5
Figure 1.4. Direct nitration pathways. A) Nitration by <i>Streptomyces fumanus</i> in the biosynthesis of dioxapyrrolomycin 9 B) Biomimetic nitration in the conversion of aknadinine 25 to 1-nitroaknadinine 10 using HNO ₃ C) Enzymatic direct nitration of L-Trp 26 by TxtE.....	6
Figure 1.5. Use of nitro compounds in industry. A) Nitrofurans-containing drugs nifurtimox 28 and nitrofurantoin 29 used to treat Chagas disease B) Nitroimidazole-containing drugs benznidazole 30 and metronidazole 31 C) 2-nitrochlorobenzene as a precursor to the synthesis of Pigment Yellow 12 34	7
Figure 1.6. Examples of CYP-catalysed reactions. A) Hydroxylation of camphor 35 by P450cam B) Hydroxylation of narbomycin 37 by PikC in the biosynthesis of pikromycin 38 C) Selected natural products, P450's and the reactions that they catalyse (as discussed in the text). Functional groups and bonds catalysed by CYPs are shown in red.....	9
Figure 1.7. Structure of heme b 48 (iron protoporphyrin IX).....	11
Figure 1.8. CYP catalytic cycle for substrate hydroxylation. The fate of the dioxygen molecule following binding to CYP heme can be followed from by the black circle.....	12
Figure 1.9. A) P450BM3-catalysed hydroxylation of palmitic acid 57 and 6β-hydroxylation of testosterone 58 by CYP3A4-BM3R incorporating the reductase domain from P450BM3 (BMR). B) Reactions catalysed by naturally self-sufficient P450RhF.....	14
Figure 1.10. A) Domain organisation and redox partner proteins in various CYP systems. Standalone CYP heme domains systems are shown by i) and ii) and self-sufficient systems are shown by iii) and iv). B) Electron transfer pathway for Class I CYP redox systems.....	15
Figure 1.11. Structure and features of CYP enzymes A) Crystal structure of PikC (PDB: 2BVJ) with selected important helices labelled B) Important active site features of PikC showing heme (magenta), ligating Cys (orange) and highly conserved residues T247	

(green) and G350 (orange). C) Open (red, PDB: 1SE6) and closed (cyan, PDB: 2D09) conformations of the F-G region and B-C loop of CYP158A2.....	16
Figure 1.12. Biosynthesis of thaxtomin A. A) Arrangement of the genes in the thaxtomin A 11 biosynthetic gene cluster B) Proposed biosynthesis of thaxtomin A C) Production of NO by the nitric oxide synthase TxtD.....	18
Figure 1.13. A) Proposed catalytic cycle for TxtE-catalysed nitration of L-Trp B) Possible O-O bond cleavage mechanisms in the formation of the active nitrating species NO. The pathway deviating from CYP hydroxylation is shown in blue.....	19
Figure 1.14. A) Crystal structure of TxtE (PDB: 4TPO, substrate not shown) B) Superimposition of TxtE (green) and PikC (grey, PDB: 2BVJ) for comparison of structural features. C) B-C and F-G regions of TxtE and PikC showing a clear substrate access channel to the heme. TxtE has two B' helices in this region compared to PikC. The unresolved TxtE F-G loop is shown by red dots. D) Unusual water network in TxtE extending from HOH256 in the active site to the external HOH88 via a kink in the I-helix near A248. Heme is shown in cyan and waters are shown as red spheres.....	21
Figure 1.15. The structure of the L-Trp bound TxtE active site. Substrate is in green and binding residues are shown in magenta, heme is in cyan and connecting waters are shown as red spheres.....	22
Figure 1.16. A) TxtEH176F/Y/W catalysed nitration of 26 to 71 demonstrating a regioselectivity switch at position 176 B) Release of NO from the nitric oxide donor DEANO.....	23
Figure 1.16. RufO-catalysed nitration of L-Tyr to 3-nitro-L-tyrosine in the biosynthesis of rufomycin.....	24
Figure 1.17. Selected L-Trp analogues previously investigated for nitration by TxtE. Substrates that were readily nitrated are shown by green dots and those that did not nitrate but did bind in a productive manner are shown by black dots. Substrates that did not nitrate and bound to the TxtE heme directly are shown by red dots.....	26
Figure 1.18. Typical directed evolution life cycle and breakdown of the stages involved.....	27
Figure 1.19. Laboratory evolution of P450BM3 A) Reactions catalysed by P450BM3 mutants generated from random mutagenesis experiments B) P450BM3 mutants from a minimal targeted library catalyse hydroxylation of valencene to nootkatone.....	29
Figure 2.1. SDS-Page analysis of TxtE purification. Purified His ₆ -TxtE (48.5 kDa) after IMAC and PD-10 column (right lanes). Cell lysate and unbound protein fractions (middle lane). Molecular weight marker is in the left lane.....	33
Figure 2.2. UV-Vis spectroscopic analysis of TxtE. TxtE in 25 mM Tris-HCl, pH 8 (black), TxtE + sodium dithionite (blue), TxtE + DEANO (red).....	33
Figure 2.3. Analysis of the TxtE-catalysed L-Trp nitration reaction and controls. A) LC-MS EICs at $m/z = 205.1$ and 250.1 corresponding to $[M+H]^+$ for 26 and 27 , respectively. B) ESI-Q-TOF mass spectrum of the product peak 27	34

Figure 2.4. Scheme representing lack of nitro- <i>N</i> -methyltryptophan product when incubated with TxtE under reaction conditions.....	35
Figure 2.5. UV-Vis spectroscopic analysis of L-Trp binding to TxtE. A) Type I difference spectra from titration of TxtE (8 μ M) with L-Trp (0-1 mM). B) Difference in absorbance at λ_{max} (390 nm) and λ_{min} (420 nm) versus L-Trp concentration. Data fitted to a Michaelis-Menten model and points are the mean of three measurements.....	36
Figure 2.6. UV-Vis spectroscopic analysis of Tyr analogue binding to TxtE. A) and C) Difference spectra from titration of TxtE (8 μ M) with <i>o</i> -Tyr and <i>m</i> -Tyr (0-1 mM), respectively. B) and D) Difference in absorbance at λ_{max} (390 nm) and λ_{min} (420 nm) versus <i>o</i> -Tyr and <i>m</i> -Tyr concentration, respectively.....	37
Table 2.1. Summary of substrate binding modes and determined K_d values.....	37
Figure 2.7. Analysis of TxtE-catalysed reaction with Tyr analogues. A) LC-MS EICs at m/z 182.2 and 227.2 corresponding to $[M+H]^+$ <i>o</i> -Tyr and nitro- <i>o</i> -Tyr, respectively. B) LC-MS EICs at m/z 182.2 and 227.2 corresponding to $[M+H]^+$ <i>m</i> -Tyr and nitro- <i>m</i> -Tyr, respectively. i) TxtE-catalysed reaction ii) Boiled TxtE control reaction iii) Substrate + DEANO only.....	38
Figure 2.8. Formation of peroxynitrite species in aqueous solutions and subsequent nitration of activated aromatic species.....	38
Figure 2.9. Agarose gel electrophoretic analysis of the linearised parent plasmid pET151- <i>txtE</i> (6921 bp), the DNA insert <i>rhfred</i> (1083 bp) and the final Gibson assembly product (8004 bp).....	39
Figure 2.10. Analysis of TxtE-RhFRED. A) Domain organisation of TxtE-RhFRED and amino acid composition of the heme-reductase linking region including an additional polypeptide linker (PPL). B) SDS-PAGE analysis of purified TxtE-RhFRED (~84.5 kDa, right lane). Molecular weight marker is in the left lane. C) UV-Vis spectroscopic analysis of TxtE-RhFRED showing absorbance at 420 nm.....	40
Figure 2.11. UV-Vis spectroscopic analysis of L-Trp binding to TxtE-RhFRED. A) Type I difference spectra from titration of TxtE-RhFRED (8 μ M) with L-Trp (0-1 mM). B) Difference in absorbance at λ_{max} (390 nm) and λ_{min} (420 nm) versus L-Trp concentration. Data fitted to a Michaelis-Menten model and are the mean of three measurements.....	41
Figure 2.12. EICs at m/z = 205.1 and 250.1 corresponding to $[M+H]^+$ for tryptophan and 4-nitrotryptophan, respectively, from LCMS analysis of the TxtE-RhFRED-catalysed L-Trp nitration reaction (top) and control (bottom).....	42
Figure 2.13. Analysis of TxtE-BM3R. A) Domain organisation of TxtE-BM3R and amino acid composition of the natural heme-reductase linking region B) SDS-PAGE analysis of purified TxtE-BM3R (~111 kDa, right lane). Molecular weight marker is in the left lane. C) UV-Vis spectroscopic analysis of TxtE-BM3R showing absorbance at 420 nm.....	43
Figure 2.14. UV-Vis spectroscopic analysis of L-Trp binding to TxtE-BM3R. A) Type I difference spectra from titration of TxtE-BM3R (8 μ M) with L-Trp (0-1 mM). B) Difference in absorbance at λ_{max} (390 nm) and λ_{min} (420 nm) versus L-Trp concentration. Data fitted to a Michaelis-Menten model and are the mean of three measurements.....	44

Figure 2.15. Analysis of the TxtE-BM3R-catalysed L-Trp nitration reaction (top) and control (bottom). LC-MS EICs at $m/z = 205.1$ and 250.1 corresponding to $[M+H]^+$ for 26 and 27 , respectively.....	44
Table 2.2. Total turnover numbers (TTN) of wild-type TxtE with redox partners compared to TxtE-reductase fusion proteins calculated from activity assays with L-Trp.....	45
Figure 2.16. The TxtE-catalysed reaction with 4-methyl-DL-Trp indicates either 5- or 7-nitration (top) as indicated by 1H NMR. Reaction with 4-F-Trp has previously been shown to result in 7-nitration (bottom).....	46
Figure 2.17. Screening TxtE-BM3R activity in various reaction buffers and utilising either NADPH or a regenerating system. Percentage conversion determined from LC-MS analysis monitoring peak area for 4-methyl-L-Trp and nitro-4-methyl-L-Trp.....	47
Figure 2.18. Analysis of the scaled-up TxtE-BM3R-catalysed 4-methyl-L-Trp 76 nitration reaction (top) and control (bottom). LC-MS EICs at $m/z = 219.1$ and 264.1 corresponding to $[M+H]^+$ for 76 and nitrated 76 , respectively.....	48
Figure 2.19. Synthesis of 4-methyl-5-nitro-L-Trp 109 from the corresponding indole 106 via an <i>N</i> -acetyl intermediate 108	49
Figure 2.20. 1H NMR spectra of synthetic and enzymatic 4-methyl-5-nitro-L-tryptophan. A) Aromatic region of the 1H NMR spectrum of the isolated product from the enzymatic reaction B) Aromatic region of the 1H NMR spectrum of synthetic 4-methyl-5-nitro-L-tryptophan. C) Aromatic region of the 1H NMR spectrum of a mixture of the enzymatic and synthetic products.....	49
Figure 2.21. Structures of the substrates tested for TxtE-BM3R catalysed nitration. A) Substrates accepted by TxtE-BM3R as determined by LC-MS analysis. B) Substrates for which no detectable levels of nitro-product were observed.....	51
Figure 2.22. Analysis of the TxtE-BM3R-catalysed nitration reaction for positive substrates. A) Left: LC-MS EICs at $m/z = 219.1$ and 264.1 corresponding to $[M+H]^+$ for 109 and nitrated 109 , respectively. Right: Mass spectrum of nitrated 109 . B) Left: LC-MS EICs at $m/z = 223.1$ and 268.1 corresponding to $[M+H]^+$ for 79 and nitrated 79 , respectively. Right: Mass spectrum of nitrated 79 . C) Left: LC-MS EICs at $m/z = 223.1$ and 268.1 corresponding to $[M+H]^+$ for 74 and nitrated 74 , respectively. Right: Mass spectrum of nitrated 74	52
Figure 3.1. Targeted TxtE heme domain active site residues (PDB: 4TPO). A) L-Trp (green) shown with binding residues in magenta. Polar contacts are shown by yellow dashed lines and connecting waters shown by red spheres. B) Residues within a 5 Å shell of the L-Trp indole moiety. Heme is shown in cyan.....	56
Figure 3.2. Overview of the flexible F-G-loop residues K180 and D181 (magenta) and interacting residues, R34 and E391 (blue) from the H176F mutant crystal structure (PDB: 5D3U). Polar contacts are shown by yellow dashed lines and L-Trp and heme are shown in green and cyan, respectively.....	57
Figure 3.3. A) Overview of the R59 and influencing residues K84, V63 and T56 (blue) within the overall structure of TxtE (PDB: 4TPO). Polar contacts are shown by yellow dashed lines. L-Trp and heme are shown in green and cyan, respectively and the R59 coordinating waters are shown as a red spheres.....	58

Table 3.1. Summary of residues targeted in the TxtE heme domain and the categories which they correspond to.....	58
Figure 3.4. Theoretical distribution of nucleotide bases when using the 22c-trick (top), NNK/S degenerate primers (middle) or NNN degenerate primers (bottom). Expected percentage distribution of each nucleotide is represented as pie charts for the three methods.....	60
Figure 3.5. Overview of Q5 Site-directed mutagenesis and QuikChange mutagenesis primer design for substitution at one site within a gene.....	62
Figure 3.6. Schematic showing the overall library preparation and screening protocol for TxtE-BM3R mutant libraries. The screening reactions and LCMS analyses are conducted in 96-well plates following QQC of the site-saturation mutagenesis libraries.....	63
Figure 3.7. EICs at $m/z = 250$, corresponding to $[M+H]^+$ for 4- (27) or 5-nitro-L-tryptophan (71), from LC-MS analyses of A) an authentic standard of 71 and B) 27 from the TxtE-BM3R-catalysed nitration reaction.....	64
Figure 3.8. Sequencing electropherogram of H176 site-saturation library. The sequence of the wild-type codon corresponding to His is shown above the targeted codon.....	65
Figure 3.9. Library screening reaction for L-Trp nitration by TxtE-BM3R(H176) variants. Reaction monitored by LCMS for either 4- or 5-nitro-product using a NADPH recycling system in a 96-well plate and conversions assessed by UV-Vis peak area.....	66
Figure 3.10. H176 mutant library screening plate heat maps. A) Heat map monitoring for product at 1.67 min corresponding to 4-nitro-L-tryptophan. B) Heat map monitoring for product at 1.81 min corresponding to 5-nitro-L-tryptophan. Conversions are represented as a scale of highest to lowest for each plate. Control reactions are as follows: A12-C12 are with TxtE-BM3R protein, D12-F12 are with pET28(a)+ proteins and G12 and H12 are with reaction buffer (100 mM Tris-HCl, pH 8).....	67
Table 3.2. Summary of TxtE-BM3R mutants capable of nitrating L-Trp at the 5-position as determined by high-throughput screening of H176 library variants.....	67
Figure 3.11. Selection of electropherograms corresponding to high quality site-saturation libraries (top row) and low quality libraries (bottom row) from the complete set prepared via the 22c-trick method in this study. The central three peaks correspond to the targeted codon in all cases and demonstrate the variation in library diversity obtained using this method.....	68
Figure 3.12. Obtained distribution of nucleotides for TxtE-BM3R mutant libraries represented as pie charts. The expected theoretical percentage distribution of nucleotide bases using the 22c-trick is shown at the top. Comparison with the distribution obtained for the high quality libraries (left) and low quality libraries (right) show the variation in library quality achieved using this method. Percentage base distribution of prepared libraries calculated from sequencing electropherograms. The wild type codon for the targeted site is shown above each pie chart.....	69
Figure 3.13. A) Bar chart of the average fold-difference in conversion for all 84 enzyme variants per library screening plate compared to the average of the three wild-type reactions for the corresponding plate which is represented as 1.0 (orange dashed line). B) Bar chart of the fold-difference in conversion for the best enzyme variant per	

- library screening plate compared to the average of the three wild-type reactions for the corresponding plate which is represented as 1.0 (purple) dashed line). Often the best mutant per plate exceeded that of the control reaction.....73
- Figure 3.14. Comparison of the seven best mutants selected from the thirty that were picked from across all enzyme variant libraries. Comparisons are made against the average conversion of the three control reactions for the corresponding plate shown as being 1.0 (red dashed line). The specific mutation has been identified and enzymes which were established as being the wild-type protein were excluded.....74
- Figure 4.1. Structures of natural products that incorporate tryptamine in their biosyntheses. A) serotonin **119** B) psilocybin **120** C) 19,20-dihydroakuummicine **121** D) monoterpene indole alkaloid precursor, strictosidine **122**.....78
- Figure 4.2. Hydrogen-bonding between R59 (magenta) and the carboxylic acid of L-Trp (green), the heme propionate (cyan) and active site waters (red spheres) in TxtE...79
- Figure 4.3. A) Library screening reaction for tryptamine nitration by TxtE-BM3R(R59) variants. B) R59 mutant library screening plate heat map monitoring tryptamine nitration. Reaction monitored by LCMS for nitrotryptamine production and conversions assessed by UV-vis peak area.....80
- Figure 4.4. SDS-PAGE analysis of TxtE-BM3R(R59C) (112 kDa) (right-hand lane). Molecular weight marker is in the left-hand lane. B) UV-vis analysis of the TxtE-BM3R(R59C) domain with the peak at 420 nm showing the incorporation of heme cofactor in the enzyme active site.....81
- Figure 4.5. A) EIC at $m/z = 161.1$ and 206.1 corresponding to $[M+H]^+$ for **90** and nitrated **90**, respectively, from LC-MS analysis of the TxtE-BM3R(R59C)-catalysed nitration of **90** reaction revealing one major product. B) EIC at $m/z = 206.1$ corresponding to nitrated **90** in UHPLC-ESI-Q-TOF MS analysis revealing major and minor products. C) i) Mass spectra for the $m/z = 206.1$ ion for the minor product ii) Mass spectra for the $m/z = 206.1$ ion for the major product.....82
- Figure 4.6. EICs at $m/z = 205.1$ and 250.1 corresponding to $[M+H]^+$ of **26** and nitrated **26**, respectively from LC-MS analysis of the TxtE-BM3R(R59C)-catalysed nitration reaction (top). A control reaction using boiled enzyme reagent is also shown (bottom). Inset: Mass spectrum of the $m/z = 250.1$ ion corresponding to $[M+H]^+$ for nitrated **26**.....83
- Figure 4.7. A) SDS-PAGE analysis of TxtE(R59C) (48.5 kDa) (right-hand side). Molecular weight marker is in the left-hand lane. B) UV-vis analysis of the TxtE(R59C) domain with the peak at 420 nm showing the incorporation of heme cofactor in the enzyme active site.....84
- Figure 4.8. EICs at $m/z = 161$ and 206 corresponding to $[M+H]^+$ of **90** (left-hand peak) and nitrated **90** (highlighted, right-hand peak), respectively, from LCMS analyses of assays using TxtE-BM3R(R59C) or TxtE(R59C) coupled with ferredoxin (Fd) and ferredoxin reductase (Fr) or a combination of these (control reactions). NADPH (2.4 mM final) was consistent in all of the above assays.....85
- Figure 4.9. A) UV-Vis difference spectra comparing the binding of **26** to TxtE-BM3R (blue) and TxtE-BM3R(R59C) (red). B) UV-Vis difference spectra comparing the binding of **90** to

- TxtE-BM3R (blue) and TxtE-BM3R(R59C) (red). Final conc. of substrate added to each protein (8 μ M) was 0.1 mM.....86
- Figure 4.10. Molecular docking simulations of **90** in the modelled TxtE-BM3R(R59C) active site. A) The lowest energy binding mode showing the tryptamine (red sticks) indole moiety in a 'flipped' state with its N-H pointing towards the roof of the active site. B) The second lowest energy binding mode showing the tryptamine indole in an 'unflipped' state with its N-H pointing downwards. Substrate was docked as the ammonium form and in both cases hydrogen bonds to residues T296 and N293.....87
- Figure 4.11. Route to **124** used in this study using an engineered TrpB(*Pf2A6*) variant with good activity for **123**, followed by a decarboxylation.....89
- Figure 4.12. LCMS analysis of synthetic and enzymatic **124**. EIC at $m/z = 206$ corresponding to $[M+H]^+$ of **124** (highlighted) for the enzymatic product isolated from the reaction, the synthetic **124** product and a mixture of the two solutions.....90
- Figure 4.13. 1H NMR spectroscopic analysis of synthetic and enzymatic **124** highlighting the aromatic region of A) Isolated product from reaction of TxtE-BM3R(R59C) with tryptamine B) synthetic **124**. C) A mixture of enzymatic **124** isolated from the reaction and synthetic **124**.....90
- Figure 5.1. A) Structure of bufotenin **125** B) Structure of physostigmine **126** C) Structure of pinoline **127** D) Structure of griseofulvin **128**.....93
- Figure 5.2. EICs at A) $m/z = 179.1$ and 224.1 corresponding to $[M+H]^+$ for **129a** and **129b**, respectively, and B) $m/z = 175.1$ and 220.1 corresponding to $[M+H]^+$ for **130a** and **130b**, respectively, from LCMS analysis of nitration assays containing **129a** or **130a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). C) Mass spectrum of the $m/z = 224$ ion corresponding to $[M+H]^+$ of **129b**. D) Mass spectrum of $m/z = 220$ ion corresponding to $[M+H]^+$ of **130b**.....94
- Figure 5.3. A) EIC at $m/z = 191.1$ and 236.1 corresponding to $[M+H]^+$ for **131a** and **131b**, respectively, from LCMS analysis of nitration assays containing **131a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). B) Mass spectra of peaks with $m/z = 236$ at i) $t = 16.1$ min and ii) $t = 16.9$ min, corresponding to $[M+H]^+$ of regioisomers **131bi**) and **131bii**), respectively.....95
- Figure 5.4. EICs at A) $m/z = 191.1$ and 236.1 corresponding to $[M+H]^+$ for **91a** and **91b**, respectively, and B) $m/z = 176.1$ and 221.1 corresponding to $[M+H]^+$ for **86a** and **86b**, respectively, from LCMS analysis of nitration assays containing **91a** or **86a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). C) Mass spectrum of the $m/z = 224$ ion corresponding to $[M+H]^+$ of **91b**. D) Mass spectrum of $m/z = 220$ ion corresponding to $[M+H]^+$ of **86b**.....96
- Figure 5.5. A) EIC at $m/z = 173.1$ and 218.1 corresponding to $[M+H]^+$ for **135a** and **135b**, respectively, from LCMS analysis of nitration assays containing **135a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). B) Mass spectra of peaks with $m/z = 218$ at i) $t = 17.0$ min and ii) $t = 17.2$ min, corresponding to $[M+H]^+$ of regioisomers **135bi**) and **135bii**).....97
- Figure 5.6. A) EIC at $m/z = 162.1$ and 207.1 corresponding to $[M+H]^+$ for **137a** and **137b**, respectively, from LCMS analysis of nitration assays containing **137a** and either TxtE-

BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). B) Mass spectrum of peak with $m/z = 207$, corresponding to $[M+H]^+$ of 137b	98
Figure 5.7. A) The two proposed pathways for indole aromatic nitration catalysed by TxtE-BM3R.....	99
Figure 5.8. A) Tryptophan analogues that are modified at the indole N-H that do not undergo nitration B) Reaction of 137 with TxtE-BM3R(R59C) produces a single nitro product.....	100
Figure 5.9. Reaction of 138 with TxtE-BM3R(R59C) also did produce nitrated product.....	100
Figure 5.10. Incubation of the tryptamine analogue 139 with TxtE- BM3R(R59C) did not result in the formation of nitro product.....	101
Figure 5.11. Molecular docking simulations of 139 binding in the modelled TxtE(R59C) active site. Binding mode showing the substrate in an 'unflipped' state. Substrate was docked as the ammonium form and in both cases this group hydrogen bonds to residues T296 and N293. Heme is shown in cyan.....	102
Figure 5.12. Substrate tolerance of TxtE-BM3R(R59C). Structures of all substrates tested for nitration with TxtE-BM3R.....	103
Figure 6.1. The 3-nitration of tyrosine catalysed by the CYP enzyme RufO using the nitric oxide donor, diethylamine NONOate (DEANO), putidaredoxin (CamA) and putidaredoxin reductase (CamB) redox system and NADH.....	108
Figure 6.2. The structure of lajollamycin with the nitro group on the polyene unit shown in red.....	109
Figure 7.1. General protocol for NEBuilder HiFi DNA Assembly kit.....	113
Table 7.1. Components and quantities of the PCR reactions according to the Q5-Site-Directed Mutagenesis protocol (NEB).....	114
Table 7.2. PCR cycling conditions for the Q5-Site-Directed mutagenesis protocol (NEB). Reactions carried out in 0.2 mL flat top PCR tubes.....	114
Table 7.3. Components and quantities of KLD treatment mixture containing kinase, ligase, DpnI. Reactions incubated at room temperature for 5 mins in 0.2 mL PCR tubes.....	114
Table 7.4. Calculated protein molecular weights and predicted extinction coefficients for TxtE, TxtE-RhFRED and TxtE-BM3R used for the calculation of protein yield after purification.....	115
Table 7.5. Components and quantities of buffers used during the purification of proteins.....	116
Table 7.6. Components and quantities of 10 % acrylamide/bis-acrylamide SDS-PAGE analysis gel.....	116
Table 7.7. List of forward and reverse primers used for TxtE-BM3R site-saturation mutant library generation by the 22c-trick method. The annealing temperatures used for each PCR reaction are also noted.....	120

Table 7.8. Theoretical base distribution when using NDT, VHG and TGG and the mutagenic codons according to the 22c-trick method.....	121
Table 7.9. Percentage distribution of bases in the first, second and third codon of the mutagenic region (NDT, VHG and TGG) of primers when using the 22c-trick method.....	121
Table 7.10. LCMS analysis conditions used during the optimisation and substrate screening experiments for TxtE-BM3R and variants.....	125
Table 7.11. Mass-Directed Auto Preparation conditions for the purification of 4-nitro-L-tryptophan from the scaled-up reaction in cell lysate.....	126
Table 7.12. LCMS analysis conditions for the screening of tryptamine substrate analogues with TxtE-BM3R_R59C variant.....	128
Table 7.13. HPLC elution conditions used for the purification of nitrotryptamine from the scaled-up reaction using purified TxtE-BM3R_R59C protein.....	129
Table 7.14. HPLC elution conditions used for the purification of <i>N</i> -acetyl-4-methyl-5-nitro-D,L-tryptophan.....	130
Table 7.15. HPLC elution conditions used for the purification of synthetic nitro standard 124	133

Acknowledgements

Firstly, I would like to thank Prof. Gregory Challis for giving me the opportunity to work on such an exciting and multidisciplinary project. He has provided constant patience, encouragement and guidance throughout my PhD has made this time a truly rewarding experience.

I'd would also like to extend my gratitude to everyone who took the time to teach me the necessary skills and techniques required to carry out my studies; for biology, Dr. Daniel Griffiths and Dr. Matthew Jenner taught me the joys of protein purification, and Dr. Daniel Zabala helped greatly with the molecular biology aspect of my research. Regarding chemistry, I would like to thank Dr. Shanshan Zhou and Dr. Jade Ronan for helping me with this brief part of my research and for being great fume hood partners.

I would like to thank all the members of the Synthetic Biochemistry laboratory at GlaxoSmithKline, Stevenage, UK, for an immensely enjoyable and gratifying experience during my industrial placement. In particular, I would like to thank supervisor Dr. Doru Roiban to whom I am incredibly grateful for sharing his knowledge and enthusiasm, and Dr. Sarah Lovelock who taught me many of the protein engineering techniques discussed in this thesis.

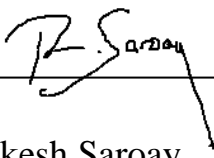
Dr. Lona Alkhalaf is thanked for providing helpful suggestions in the last three years of my PhD and for laboriously reading my first thesis drafts. Dr. Sarah Barry laid the foundation for much of this work and is thanked for helpful suggestions early in my studies. I would also like to thank all other members of the Challis group past and present, but in particular Matt, Chris, Christian, Doug and Richard, who made the laboratory an entertaining place, and also provided many laughs out of it.

Thanks to BBSRC/GSK for funding, my advisory panel, all other members of CBRF and the extended family at GSK who I have had the pleasure of meeting during this journey.

Finally, I would like to thank my Mum, Dad and sister, Kajal, for supporting me not only throughout my PhD studies but also throughout all the other endeavours I have decided to take on.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any other degree. Results from other authors are referenced in the usual manner throughout the text.



Rakesh Saroay

Date: 17/03/19

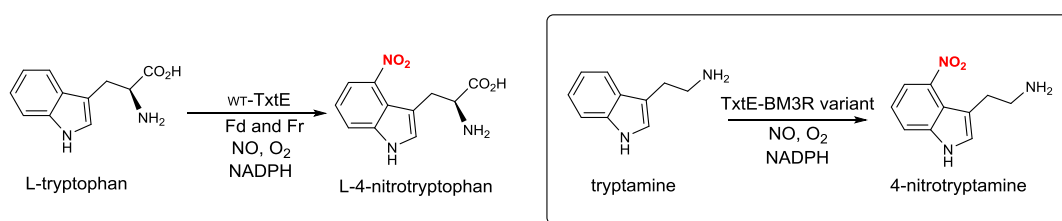
Abbreviations

BGC	Biosynthetic gene cluster
CYP	Cytochrome P450
EIC	Extracted ion chromatogram
EDTA	Ethylenediamine
ESI	Electrospray ionization
FAD	Flavin adenine dinucleotide
FMN	Flavin mononucleotide
GC	Guanosine cytosine
HPLC	High performance liquid chromatography
HR	High resolution
IMAC	Immobilized metal ion affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
LB	Luria-Bertani
LC	Liquid chromatography
MAOI	Monoamine oxidase inhibitor
MS	Mass spectrometry
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
NMR	Nuclear magnetic resonance
NRPS	Nonribosomal peptide synthetase
PCP	Peptidyl carrier protein
PCR	Polymerase chain reaction
PKS	Polyketide synthase
PMSF	Phenylmethylsulfonyl fluoride
PPL	Polypeptide linker
SAM	S-adenosyl methionine
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TEMED	Tetramethylethylenediamine
TOF	Time-of-flight
UHPLC	Ultra high performance liquid chromatography
UV-Vis	Ultraviolet-Visible

Abstract

TxtE is a unique cytochrome P450 that catalyses the direct regioselective nitration of L-tryptophan in the first dedicated step in thaxtomin A biosynthesis.

Highly corrosive concentrated nitric and sulphuric acids are traditionally used under harsh conditions to achieve aromatic nitration. Such reagents are difficult to handle on a large scale. Furthermore, these reactions lack regiospecificity, often resulting in a mixture of regioisomers. In contrast, TxtE catalyses regiospecific nitration of L-tryptophan at the 4-position using the artificial NO donor 2-(*N,N*-diethylamino)-diazenolate 2-oxide (DEANO), molecular oxygen, Fd and Fr redox partners and NADPH. The reaction proceeds at ambient temperature and the reagents are straightforward to handle, making them attractive for use on an industrial scale.



Abstract figure. Nitration of L-Trp by wild-type TxtE and nitration of tryptamine by a TxtE-BM3R variant.

To develop TxtE into an industrially useful biocatalyst, the enzyme was firstly fused to the reductase domain of P450_{BM3}, a naturally occurring self-sufficient bacterial cytochrome P450 from *Bacillus megaterium*, eliminating the requirement for exogenous redox partner proteins. The ability of TxtE-BM3R to nitrate a range of L-Trp analogues and other aromatic compounds was then investigated, revealing TxtE is tolerant to minor substrate modifications, such as indole ring substitutions. A recently-published X-ray crystal structure of the enzyme was exploited to direct high-throughput enzyme engineering efforts focussed on improving turnover and broadening its substrate scope. Twenty-two site-saturation libraries were screened against the natural substrate and initial results revealed several mutants that improve conversion by up to 3.5 times that of the wild-type enzyme. Lastly, an R59X library was screened for tryptamine nitration, identifying two mutants capable of nitrating this biosynthetically-important molecule. Investigations into the substrate scope of this variant uncovered activity towards several molecules not accepted by the wild-type enzyme including a tricyclic compound.

Chapter 1: Introduction

1.1 Natural products

Bacterial, fungal and plant natural products exhibit an extremely broad range of bioactivities and are used as anti-microbial, anticancer and antitumor agents.¹ These compounds are either directly used or serve as leads for the development of therapeutic or agricultural agents and thus are very important for society.² The semi-synthetic cholesterol-lowering drug simvastatin **1** for example is derived from lovastatin **2** (Fig.1.1A), a fungal natural product first isolated from *Aspergillus terreus*.³ A vast array of bioactive compounds are isolated from plants, including the anticancer drug vincristine **3** (Fig.1.1B), a vinca alkaloid produced by the Madagascar periwinkle *Catharanthus roseus*.⁴ Many natural products also act as antibiotics including the glycopeptide vancomycin **4** (Fig.1.1C) isolated from *Amycolatopsis orientalis*, which demonstrates its activity by blocking cell wall biosynthesis.⁵ These are just a few examples of the diverse structural scaffolds created by nature which fall into several chemical classes such as polyketides,⁶ non-ribosomal peptides,⁷ alkaloids⁸ and terpenoids.⁹

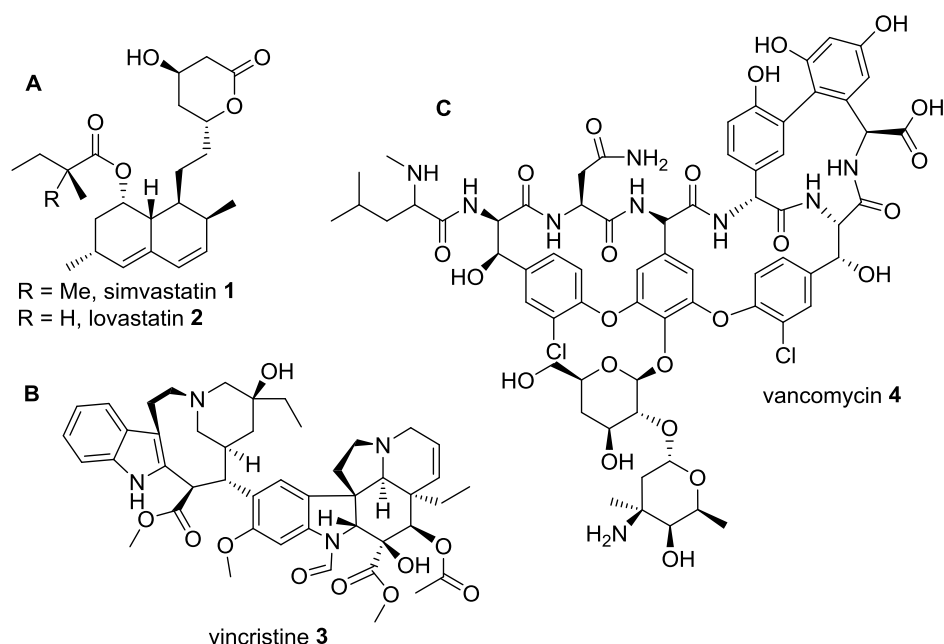


Figure 1.1. Natural product structures A) Semi-synthetic drug simvastatin **1**, derived from the polyketide lovastatin **2** isolated from *Aspergillus terreus*. B) Alkaloid vincristine **3** isolated from *Catharanthus roseus*. C) Glycopeptide vancomycin **4** isolated from *Amycolatopsis orientalis*.

1.2 Natural nitro products

Nitro compounds are often associated with synthetic compounds such as pharmaceuticals, dyes and high-energy explosive materials. Although relatively rare, there are also a significant number of natural nitro products which display a variety of impressive bioactivities.¹¹ One of the most prolific is the antibiotic chloramphenicol **5** first isolated from *Streptomyces venezuelae* in 1948 and still currently in use.¹² The antitumor compound aureothin **6** from *Streptomyces thioluteus* contains a nitrobenzoate unit¹³ whilst the highly decorated peptide hormaomycin, a bacterial hormone from *Streptomyces griseoflavus*, contains the structurally unusual nitrocyclopropylalanine **7** component.¹⁴ Pyrrolnitrin **8** is an antifungal antibiotic isolated from *Pseudomonas fluorescens*¹⁵ and dioxapyrrolomycin **9** was first isolated from *Streptomyces fumanus*.¹⁶ The Chinese plant *Stephania sutchuenensis* is the source of the structurally unusual alkaloid 1-nitroaknadinine **10**,¹⁷ whilst thaxtomin A **11** was first isolated from plant pathogenic *Streptomyces scabies*.¹⁸ Heronapyrrole C **12** is a member of the rare pyrroloterpene class of natural products isolated from the marine organism *Streptomyces* sp. (CMB-M0423) and was one of the first examples of a natural 2-nitropyrrole.¹⁹

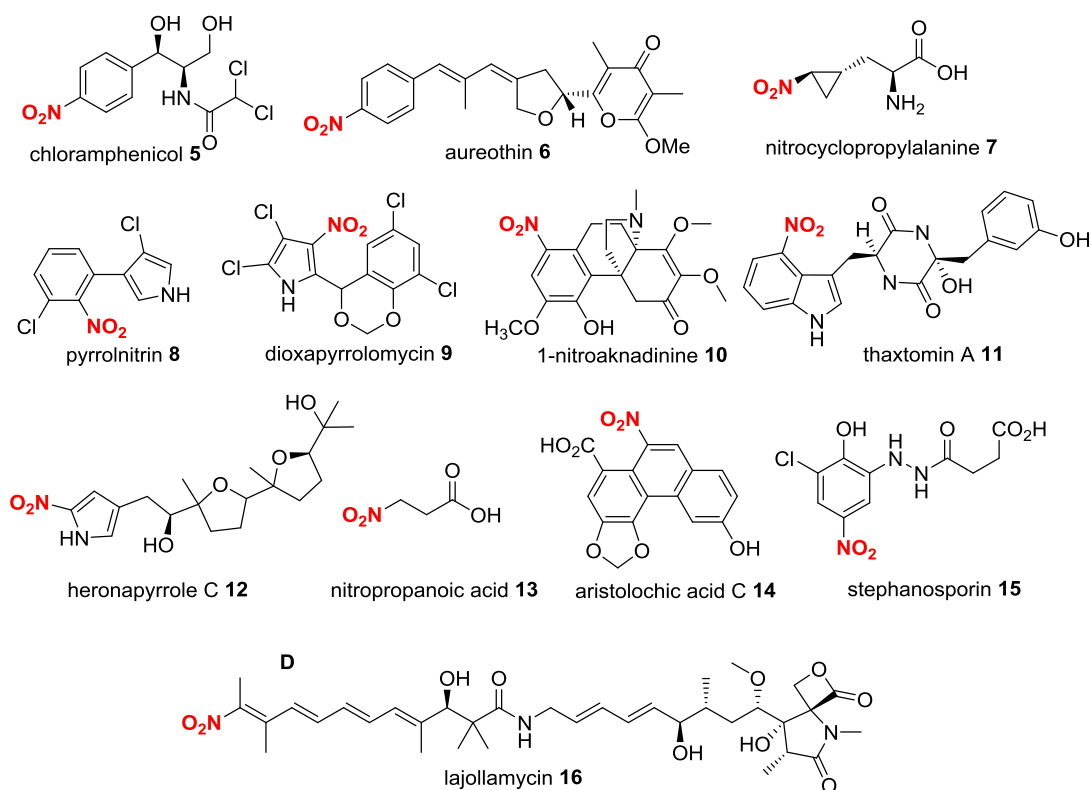


Figure 1.2. Structures of selected natural nitro products.

The simple nitropropanoic acid **13** unit has been found in plants where it has been found to be incorporated into hiptagin from *Hiptage mandoblota* Gaertn²⁰ and also in fungi (*Aspergillus flavus*).²¹ Despite the high toxicity of aristolochic acid C **14** from the Aristolochiaceae family of plants, it has a long history of being used in Chinese herbal medicine.²² 2-chloro-4-nitrophenol is formed from stephanosporin **15** in *Stephanospora caroticolor* following ‘traceless’ loss of the hydrazide side chain.²³ Lajollamycin **16**, which is a nitropolyene first isolated from the marine bacterium *Streptomyces nodosus*,²⁴ is part of the structurally novel tetraene β -lactone- γ -lactam class of antibiotics which is exemplified by the closely related analogue oxazolamycin.²⁵

1.3 Biosynthetic routes to nitro groups

Enzymatic nitration can be achieved via oxidation of an amine for direct nitration.

1.3.1 Enzymatic amine oxidation

Although *N*-oxidation as a mechanism for introduction of a nitro group has been proposed for several natural products including chloramphenicol and 3-nitropropanoic acid, only two confirmed examples are known. One is PrnD, a Rieske non-heme iron-dependent amine oxygenase²⁶ that facilitates the formation of the aryl nitro group of pyrrolonitrin **8**. Mechanistic studies on the formation of the nitro group were conducted by Lee *et al.* using isotopically-labelled O₂ and *p*-aminobenzylamine **19** (pABA) as a probe.²⁷ Previous studies had suggested the successive *N*-oxidation of an amine to give the monohydroxylamine and dihydroxylamine derivatives, such as **20** and **21**.²⁸ Dehydration to a nitroso species was then suggested to be followed by a final oxidation step to give the nitro compound.

Lee *et al.* firstly investigated the origin of the oxygen atoms by incubating pABA **19** with PrnD in an ¹⁸O₂ atmosphere. The pHABA **20** intermediate was found to contain one ¹⁸O atom whilst pNBA **23** contained two confirming both oxygens are derived from the atmosphere. In a second experiment to investigate the intermediates, unlabelled pHABA **20** was incubated with PrnD in an ¹⁸O₂ atmosphere. The pNOBA **22** formed did not include an ¹⁸O atom indicating that a second *N*-oxidation does not occur at this stage and that desaturation takes place instead. When unlabelled pNOBA **22** is incubated under the same conditions, one ¹⁸O atom is incorporated into the final pNBA **23** product showing that the final oxygen is installed after the formation of the nitroso species.

The second enzyme characterised to be a nitro-group forming *N*-oxygenase is AurF involved in aureothin biosynthesis.²⁹ The unusual *p*-nitrobenzoate **18** starter unit required by the polyketide synthase (PKS) to complete aureothin **6** biosynthesis was confirmed to follow the same mechanism as that of PrnD. These studies suggested a common amine oxidation route to nitro groups which is surprising considering the enzymes belong to

different classes. PrnD is a Rieske oxygenase that contains both a [2Fe-2S] cluster and a non-heme iron centre within its catalytic subunit. AurF on the other hand belongs to the di-iron non-heme oxygenase family of enzymes that contain a dinuclear iron centre bridged by an oxygen atom.³⁰

To date, no enzymes have been discovered that can incorporate both atoms of a single dioxygen molecule simultaneously into an amine group.

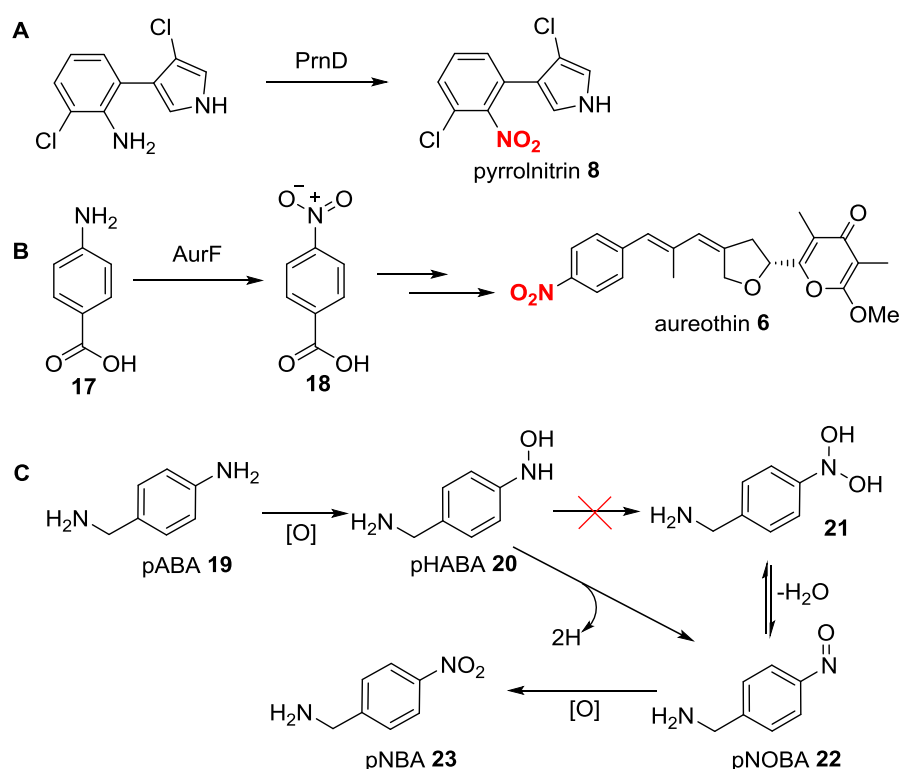


Figure 1.3. Biosynthetic *N*-oxygenases A) PrnD-catalysed nitration in the biosynthesis of pyrrolnitrin **8** B) AurF catalysed nitration in the biosynthesis of *para*-nitrobenzoate **18** (pABA) from *para*-aminobenzoate **17** (pNBA) in aureothin biosynthesis. C) Proposed pathway to *para*-nitrobenzylamine **23** via the *N*-oxidation of *para*-aminobenzylamine **19** by PrnD.

1.3.2 Direct enzymatic nitration

Direct biochemical nitration has been suggested for natural products without an amino derivative. This was first implicated for dioxapyrrolomycin **9**, an antibiotic from *Streptomyces fumamuis*.³¹ Carter *et al.* used $K^{15}N^{18}O_3$ as the sole source of nitrogen during culturing of the bacteria and found that the proportion of ^{15}N and ^{18}O incorporated into the isolated nitro product was similar to that of the nitrate source. This indicates that the organism is able to directly install the nitro group analogous to electrophilic aromatic nitration with NO_2^+ . Similarly, the proposed formation of 1-nitroaknadinine **10** in *Stephania sutchuenensis* was replicated biomimetically by treating the non-nitrated analogue **25** with HNO_3 .¹⁷ This product had similar IR and 1H NMR spectra to 1-nitroaknadinine **10** suggesting direct biochemical nitration is plausible. Despite efforts to study the biosynthesis of these compounds via gene cluster analysis and cloning, the enzymes responsible for these reactions have not yet been identified.

TxtE, a cytochrome P450 enzyme, was the first direct nitrating biosynthetic enzyme to be characterised.³² It is involved in the nitration of L-Trp **26** to give 4-nitrotryptophan **27** required for thaxtomin A **11** biosynthesis and is the focus of this work.

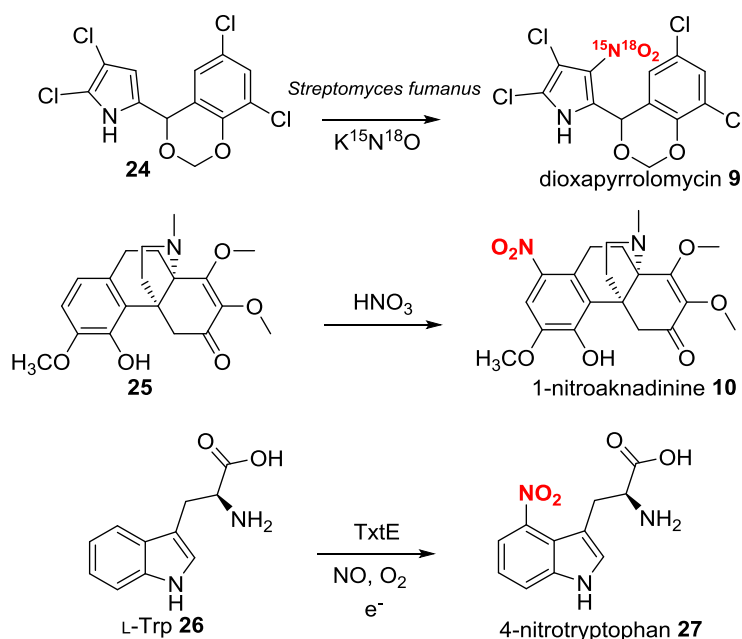


Figure 1.4. Direct nitration pathways. A) Nitration by *Streptomyces fumamuis* in the biosynthesis of dioxapyrrolomycin **9 B) Biomimetic nitration in the conversion of aknadinine **25** to 1-nitroaknadinine **10** using HNO_3 C) Enzymatic direct nitration of L-Trp **26** by TxtE.**

1.4 Industrial importance of nitro compounds

Nitro compounds are important for a wide range of industrial applications. The benzofuran-containing drugs nifurtimox **28** and nitrofurantoin **29** are used to treat Chagas disease³³ and bacterial infections,³⁴ respectively. Benznidazole **30** and metronidazole **31** are two molecules on the World Health Organisation's List of Essential Medicines and demonstrate antiparasitic³⁵ and antiprotozoan³⁶ activity, respectively, and belong to the nitroimidazole class of drugs. In these cases the nitro group is essential for activity and the agents are considered pro-drugs as they must first become activated. The nitro group can undergo bioactivation by enzymatic reductases *in vivo* that form reactive species such as nitro radical ions and it is these species that are understood to account for the biological effects of these drugs.^{37, 38}

Nitro compounds are also extremely important for the synthetic dyes and pigments industry. 2-nitrochlorobenzene **32** is used in the synthesis of 3,3'-dichlorobenzidine **33**, a precursor to several important pigments including Pigment Yellow 12 **34**.³⁹

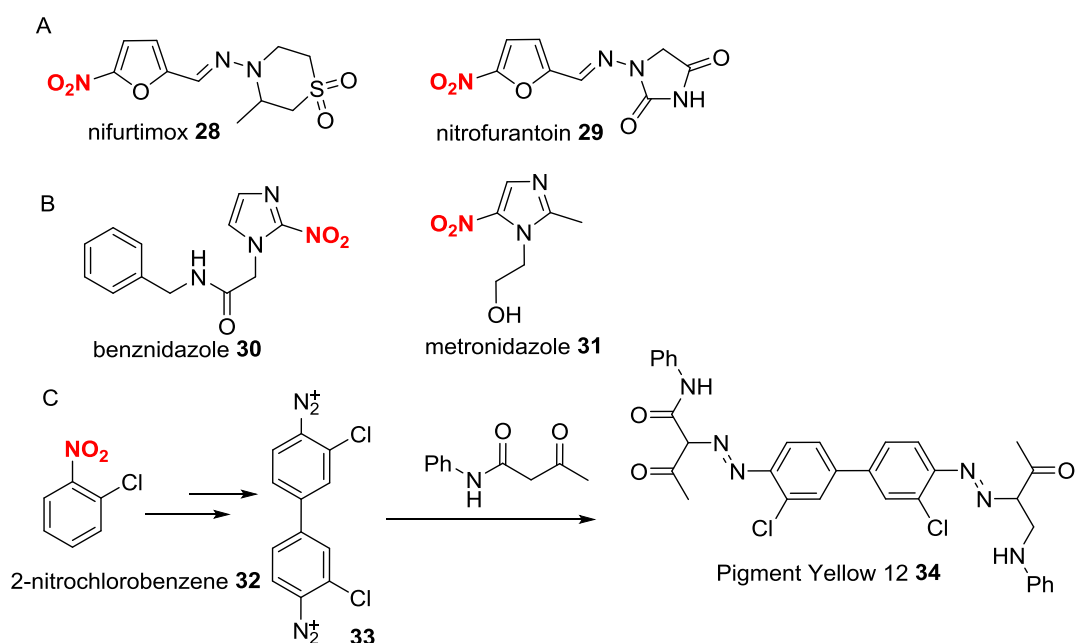


Figure 1.5. Use of nitro compounds in industry. A) Nitrofuran-containing drugs nifurtimox **28** and nitrofurantoin **29** used to treat Chagas disease B) Nitroimidazole-containing drugs benznidazole **30** and metronidazole **31** C) 2-nitrochlorobenzene as a precursor to the synthesis of Pigment Yellow 12 **34**.

The synthesis of these nitro compounds often relies on the use of concentrated nitric and sulphuric acids under harsh conditions.^{40, 41} Such reagents are difficult to handle on a large scale and furthermore the S_NAr type nitration of aromatic compounds often lacks regioselectivity, resulting in a mixture of regioisomers.^{42, 43} The study of biosynthetic nitrating enzymes will assist in the future development of environmentally-friendly and regioselective alternatives that are convenient for use in industry.

1.5 Cytochrome P450 enzymes

1.5.1 Role of cytochrome P450 enzymes in natural product biosynthesis

Biosynthetic tailoring enzymes are often responsible for introducing the diversity and functional complexity of natural products.⁴⁴ These enzymes can be involved in the modification the starter/extender units required by PKS or NRPS multimodular synthases or the tuning of the natural product scaffolds post-assembly.^{45, 46} Some have even demonstrated activity on carrier-protein tethered substrates resulting in the introduction of functionality during assembly of the molecular backbone.^{47, 48, 49}

Cytochrome P450 enzymes (CYPs) are one of the most important classes of tailoring enzymes and are involved in a wide range of oxidation reactions in natural product biosynthesis,⁵⁰ the most common of which are hydroxylation⁵¹ and epoxidation.⁵² They are versatile catalysts acting on a diverse range of substrate structures and often facilitate reactions that would challenge even the most accomplished synthetic chemists. They are heme-dependent catalysts that obtain their name from a characteristic ferrous-CO absorption maximum (Soret band) at 450 nm resulting from a $\pi - \pi^*$ transition which is common in porphyrins.⁵³ Although several other enzyme classes such as flavin-dependent monooxygenases and non-heme iron dependent oxygenases are capable of catalysing oxidative transformations, neither possess the reaction scope demonstrated by CYPs.⁵⁴ As a result these enzymes have been utilised as biocatalysts across multiple biological kingdoms (archaea, bacteria, eukarya), including notably by humans, where a super family of 57 CYPs are encoded.⁵⁵

In biosynthetic pathways, one of the most widely studied CYPs is P450cam involved in the hydroxylation of camphor **35** to 5-*exo*-hydroxycamphor **36** (figure 1.6A).⁵¹ It was the first CYP structure to be determined by crystallography and remains a model example.⁵⁶ Another is PikC (CYP107L1) involved in pikromycin **38** biosynthesis in *Streptomyces venezuelae*. This enzyme catalyses the hydroxylation of the 14-membered ring macrolide narbomycin **37** to give pikromycin **38** (figure 1.6B), but can also hydroxylate the 12-membered ring macrolide YC-17 at two different positions to give methymycin or neomethymycin (not shown).⁵⁷

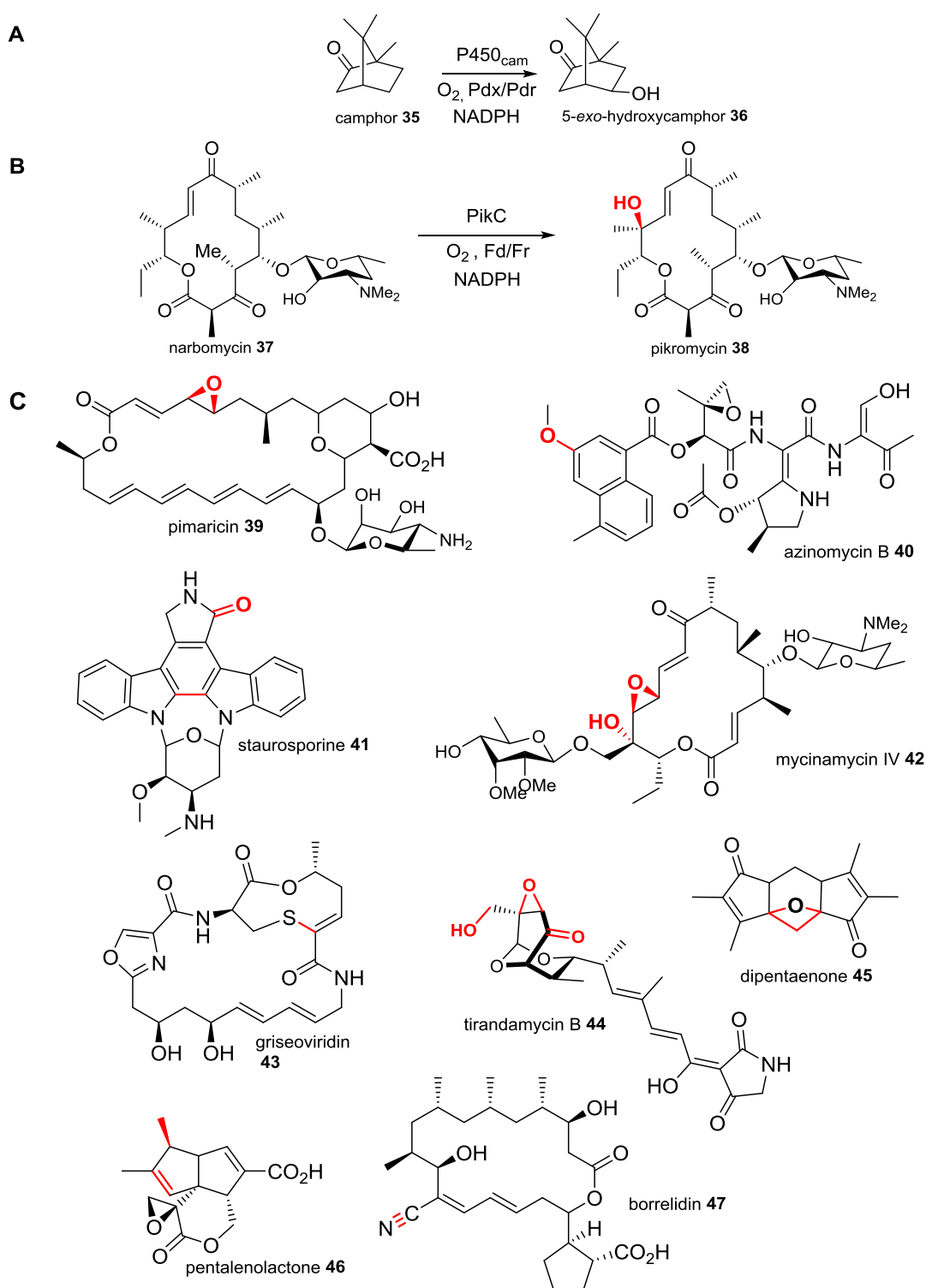


Figure 1.6. Examples of CYP-catalysed reactions. A) Hydroxylation of camphor 35 by P450cam B) Hydroxylation of narbomycin 37 by PikC in the biosynthesis of pikromycin 38 C) Selected natural products, P450's and the reactions that they catalyse (as discussed in the text). Functional groups and bonds catalysed by CYPs are shown in red. Examples from ref. 50.

A range of CYP epoxidations are frequently observed on polyketide scaffolds as a result of the common occurrence of olefins within these structures. PimD (CYP161A2), a CYP from *Streptomyces natalensis* is involved in the formation of the C-4-C5 epoxide of polyene macrolide pimaricin **39**.⁵⁸ CYPs can also catalyse aromatic hydroxylation. AziB1 hydroxylates naphthoic acid in the biosynthesis of azinomycin B **40**, a potent antitumor agent.⁵⁹ These reactions are believed to proceed via initial epoxidation of the aromatic ring followed by opening of the epoxide and restoration of the aromaticity.

MycG is a multifunctional CYP involved in the biosynthesis of the mycinamycins, 16-membered macrolides isolated from *Micromonospora griseorubida*.⁶⁰ MycG catalyses the sequential C14 hydroxylation and C12=C13 epoxidation of mycinamycin IV **42** forming mycinamycin V and II, respectively. The order of these steps cannot be exchanged as epoxidation of mycinamycin IV **42** does not result in further reaction. Additionally, the methyl esters of the mycinose moiety were determined to be crucial for substrate recognition in the MycG active site and potentially serve as a method for discriminating between closely related analogues.

Another remarkable example of a multifunctional CYP is TamI from the tirandamycin biosynthetic pathway in *Streptomyces* sp. 307-9.⁶¹ TamI repeatedly exchanges substrates with an FAD-dependent oxidase TamL. TamI catalyses hydroxylation at C10 before TamL oxidises this group to a ketone. The product then undergoes successive epoxidation at C11-C12 and hydroxylation at the C18 methyl catalysed by TamI to give the final tirandamycin B **44** product.

Given the large variation in sequence and structure, CYPs can catalyse a number of unusual reactions beyond the typical hydroxylations and epoxidations.⁴⁴ Borrelidin **47** is an 18-membered macrolide that demonstrates a variety of bioactivities including antiviral and antimalarial. It contains a rare nitrile group which has been proposed to be catalysed by an enzymatic cascade that includes the CYP BorI.⁶² BorI and a dehydrogenase BorK catalyse aldehyde formation on a macrolide ring methyl group. The aminotransferase BorJ then catalyses aminomethyl formation before BorI converts this group to a nitrile in the final biosynthetic step.

PntM is involved in an unusual oxidative rearrangement in pentalenolactone **46** biosynthesis.⁶³ The substrate substituents generate steric hindrance in the PntM active site that suppresses the formation of reactive radical species that are required for hydroxylation. Instead, electron transfer is favoured allowing the formation of the 1,2-dimethylcyclopent-2-ene moiety to proceed via carbocation formation and subsequent methyl migration.

StaP (CYP245A1) is a CYP involved in the biosynthesis of staurosporine **41**, an alkaloid isolated from *Streptomyces* sp. TP-A0274, and is responsible for formation of the

indolocarbazole scaffold via intramolecular C-C bond formation.⁶⁴ Synthetically, such reactions remain challenging as a result of difficulties in simultaneously controlling both the regioselectivity and configuration of the bond formed.

The redox-independent CYP154A1 from *Streptomyces coelicolor* is an exceptional example of the structural, chemical and mechanistic diversity of CYPs.⁶⁵ In addition to not requiring redox partner proteins, this CYP has its heme cofactor orientated 180° relative to other proteins within this class of enzymes. This unusual feature has been proposed as the reason for the rare intramolecular cyclisation of dipentaenone leading to a Paternó-Büchi-type product **45** resulting from the reaction between the C5 carbonyl group and the C11/C12 double bond.

Finally, SgvP is a CYP involved in griseoviridin **43** biosynthesis in *Streptomyces griseoviridis* NRRL 242, and has been revealed to be responsible for the C-S bridge formation in this compound.⁶⁶ Despite multiple cases of C-S bond forming reactions in microbial biosynthetic pathways, these transformations are often accomplished by non-heme iron dependent oxygenases or S-transferases. SgvP is the first example of a microbial CYP capable of such a transformation.

1.5.2 CYP catalytic mechanism

CYPs catalyse oxidation by using a heme stabilised iron atom to produce a highly reactive oxidising species from molecular oxygen.⁵³

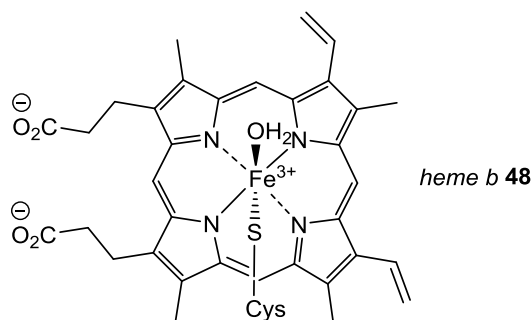


Figure 1.7. Structure of heme b 48 (iron protoporphyrin IX).

The typical catalytic cycle for CYPs is shown in figure 1.8.⁶⁷ In the resting state **49**, the CYP heme iron weakly binds a solvent water molecule in **50** which is displaced upon substrate binding. This causes a reorganisation of the ferric *d*-orbitals and in turn the electronic spin state equilibrium switches from a low-spin state ($S = 1/2$) to a high-spin state ($S = 5/2$), resulting in an increase in the ferric iron reduction potential. This shift encourages the transfer of a single electron from NADPH via redox partners, forming the ferrous species **51** and at this point the catalytic cycle can proceed via the binding of

dioxygen. This mechanism ensures that the generation of reactive species only takes place once a substrate molecule is present in the active site. The dioxygen species **52** can adopt a ferric superoxy form **53** before transfer of a second electron to give the ferric peroxy form **54**. An initial protonation of this species forms the ferric hydroperoxy state **55** which, following further protonation and loss of water, generates the catalytically active ferryl-oxo species (compound I **56**). Compound I **56** is stabilised by the dissociation of a radical into the porphyrin π -system and is considered to be the crucial reactive form required for monooxygenation of substrates. Dissociation of the product from the active site allows binding of a water molecule which reinstates the low-spin ferric species ready for the next cycle.

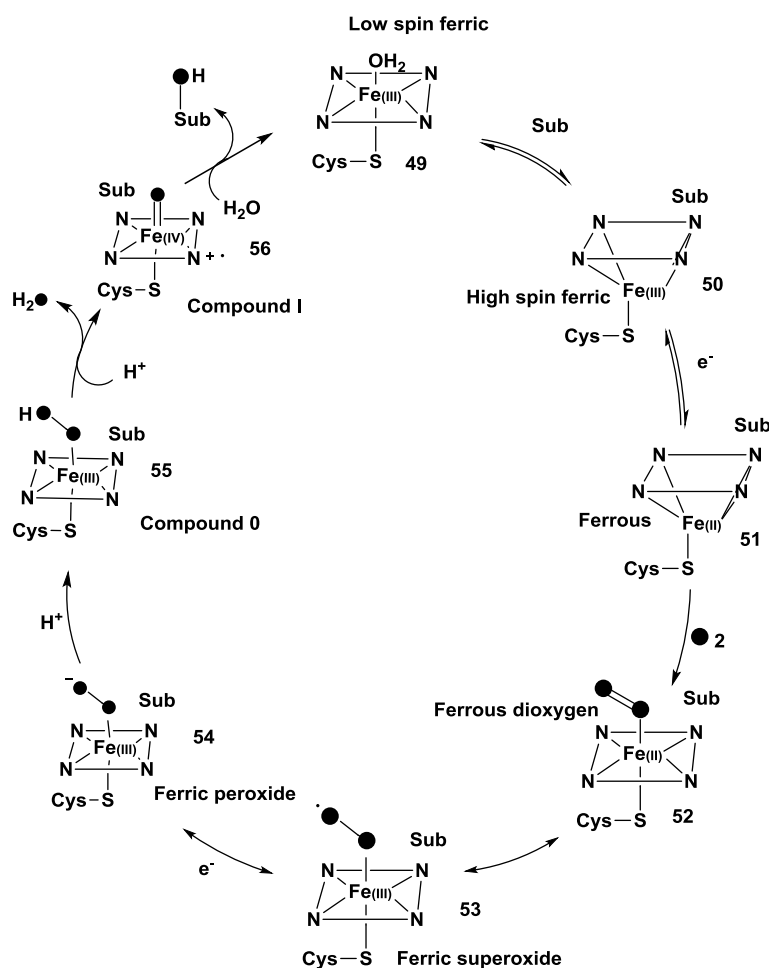


Figure 1.8. CYP catalytic cycle for substrate hydroxylation. The fate of the dioxygen molecule following binding to CYP heme can be followed from by the black circle.

1.5.3 Redox partners for CYPs

The reduction of the ferric-heme species is generally achieved by the transfer of electrons from a suitable source such as NAD(P)H by redox partner proteins.⁶⁸ There are two common types of CYP-redox systems.⁶⁹ The bacterial class I systems consist of soluble cytoplasmic proteins in a three-component system in which electrons from NAD(P)H are transferred via a FAD-containing reductase to an iron-sulphur cluster-containing ferredoxin. The eukaryotic class II systems consist of microsomal membrane-bound proteins in a two component arrangement consisting of an FAD- and FMN-containing reductase for electron transfer from NAD(P)H. Despite a majority of CYP redox systems falling into either class I or II, there are a growing number of exceptions to this rule. Recent discoveries include a self-sufficient CYP fusion protein (CYP102D1) from *Streptomyces avermitilis*,⁷⁰ and a redox-independent CYP (P450_{154A1}) from *Streptomyces coelicolor* that remarkably catalyses an intramolecular cyclisation without the need for NAD(P)H or redox partner proteins.⁶⁵

1.5.4 CYP-Reductase fusion proteins

The requirement of additional redox partner proteins has limited the application of CYPs in industry. Several CYPs are known to bypass this limitation by containing a reductase domain directly fused to the heme domain. The most well characterised example is P450_{BM3} (CYP102A1),⁷¹ a one-component system from *Bacillus megaterium*, capable of hydroxylating saturated fatty acids of various chain lengths with the highest rate of activity for a CYP discovered thus far ($> 15,000 \text{ min}^{-1}$ for arachidonic acid).⁷² This multiprotein fuses a reductase domain containing FAD- and FMN-binding motifs to the heme-bound CYP domain. Single electrons are firstly transferred from NADPH to the FAD domain and then on to the FMN-domain prior to being shuttled to the heme iron. Kinetic studies have shown that it is indeed this electron transfer arrangement that results in the exceptional catalytic properties of the enzyme.⁷³ Another instance is P450_{RhF} from *Rhodococcus* sp. NCIMB 9784, containing a novel FMN- and [2Fe-2S]-containing reductase domain.⁷⁴ Despite its physiological function remaining unclear, P450_{RhF} has been shown to catalyse a myriad of reactions including dealkylation to **60**, epoxidation to **62** and asymmetric sulfoxidation to **63**, in addition to the typical hydroxylation activity to **60**.⁷⁵

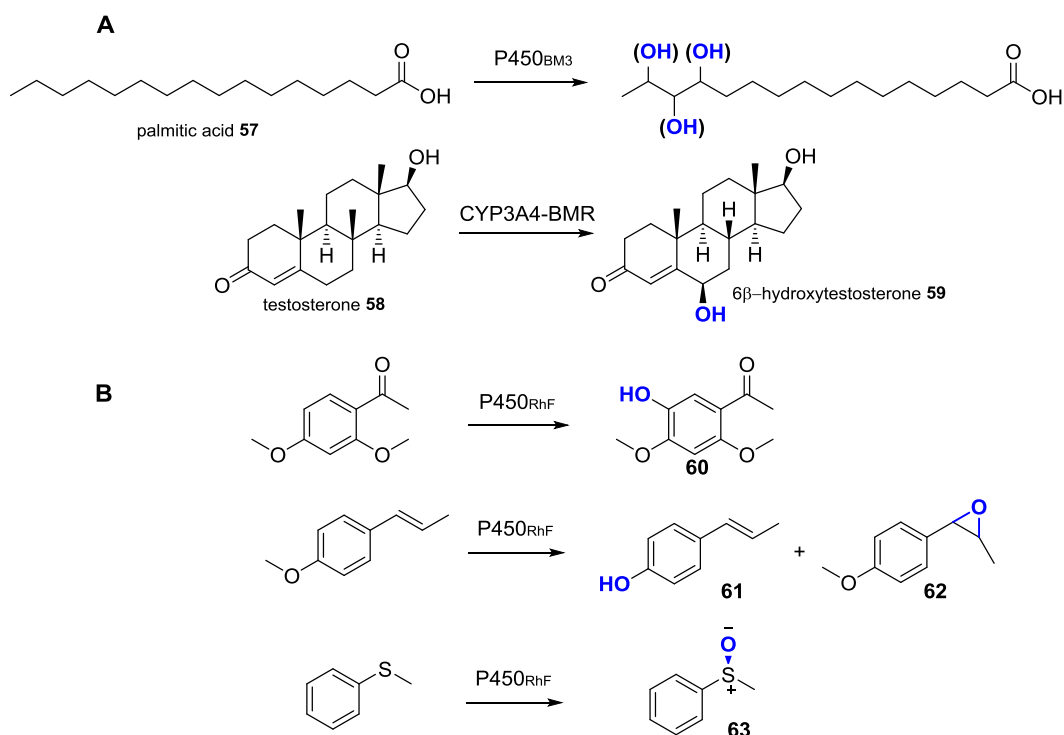


Figure 1.9. A) P450_{BM3}-catalysed hydroxylation of palmitic acid **57 and 6β-hydroxylation of testosterone **58** by CYP3A4-BM3R incorporating the reductase domain from P450_{BM3} (BMR). B) Reactions catalysed by naturally self-sufficient P450_{RhF}.**

The discovery of naturally occurring one-component CYP-reductase proteins has directed others to reconstitute the activity of CYP heme domains by fusing with the corresponding reductase domains from P450_{BM3} (BM3R) and P450_{RhF} (RhFRED). Notable examples include the fusion of the BM3R domain to the human liver CYP 3A4 which is responsible for the metabolism of a vast majority of drugs and xenobiotics.⁷⁶ The solubility of CYP3A4 was increased upon fusion to BM3R and engineering of the heme-reductase linker region provided further improved catalytic activity towards testosterone **58** hydroxylation, a reaction catalysed by the enzyme in humans.⁷⁷

The first example of RhFRED fusion to a bacterial CYP was in the generation of PikC-RhFRED.⁷⁸ This one-component system provided a 4-fold increase in the catalytic activity over wt-PikC coupled with Fd and Fr as determined by the respective K_{cat}/K_m values. Additionally, RhFRED has been fused to MycG,⁷⁹ P450_{cam}, P450_{bzo} and P450_{balk}⁸⁰ with similar success. Often the preparation of the CYP-reductase fusion has provided a platform for further engineering efforts focussed on optimising the activity or broadening substrate tolerance.

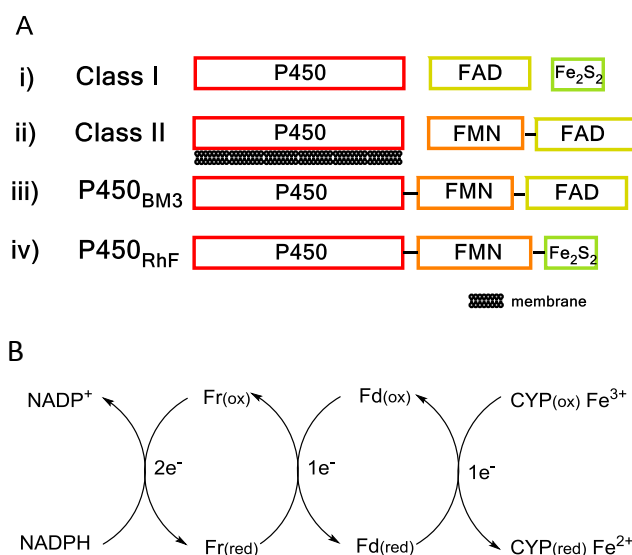


Figure 1.10. A) Domain organisation and redox partner proteins in various CYP systems. Standalone CYP heme domains systems are shown by i) and ii) and self-sufficient systems are shown by iii) and iv). B) Electron transfer pathway for Class I CYP redox systems.

1.5.5 CYPs sequence and structure

Across the P450 family of enzymes there is a general lack of sequence homogeneity or conserved motifs.⁵⁰ One of the few highly conserved residues is the catalytically important Cys that coordinates the heme iron (Fig.1.9.B). This axial cysteinate largely accounts for the unusual spectroscopic properties of CYPs and in particular that of the distinctive ferrous-CO species.^{53,81} Furthermore, this ligand has been proposed to be crucial for the catalytic activity of CYPs. The proximal thiolate of the cysteine has been reported to possess strong-electron releasing character that is able to ‘push’ electron density towards the ferric heme. This has been shown to be critical for heterolytic cleavage of the heme-bound O-O bond and stabilisation of the resulting high-valent Fe^{IV}-species (compound I).⁸² It should be noted that this Cys is absent from several CYPs from various organisms whose function remains unknown,⁸³ and the mutation of Cys to His or Ser in several engineered CYPs has been shown to support the unnatural carbene- and nitrene-transfer reactions⁸⁴ suggesting that natural CYPs lacking the Cys may yet possess unusual and interesting activities.

In general, CYPs are mainly α -helical and share common folds that are structured around the heme cofactor (figure 1.1).⁸⁴ The heme prosthetic group is sandwiched between the large I-helix and a smaller β -sheet rich domain (β 5) that includes the heme-ligating Cys residue. The core structure is important for oxygen activation and is similar for most CYPs. Outside of this region CYPs, particularly those of *Streptomyces* origin, comprise 12

α -helices ($\alpha 1$ - αL) and 10 β -strands ($\beta 1$ - $\beta 10$) that adopt a prism-like structure.⁵⁰ The I-helix usually extends throughout the entire length of the protein and is a key scaffold encompassing all CYPs.

Two highly conserved residues in CYPs include a Thr (T247 in PkC) located in the I-helix and part of the catalytically-important ((A/G)Gx(E/D)T) sequence motif (figure 1.11B). This Thr has been proposed to be important for heme-bound dioxygen activation as well as the prevention of autoinactivation in the presence of substrate.⁸⁵ Mutation of Thr252 to Ala in P450_{cam} has been shown to hinder camphor hydroxylation by up to 95 %. Within the Cys pocket sequence of FxxGx(H/R)xCxG the Gly lying 4 residues upstream of the heme-ligating Cys (G350 in PkC) is also highly conserved and this flexible residue has specifically been implicated in forming a conserved β -hairpin turn (figure 1.11B).⁸⁴

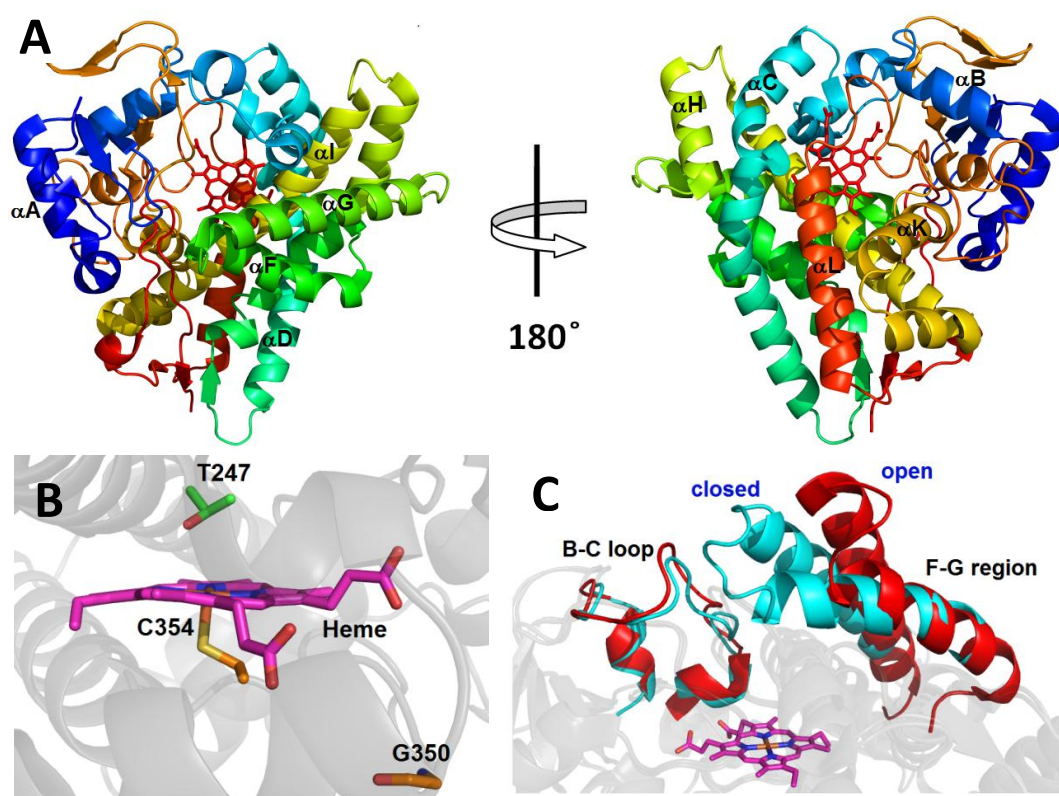


Figure 1.11. Structure and features of CYP enzymes A) Crystal structure of PkC (PDB: 2BVJ) with selected important helices labelled B) Important active site features of PkC showing heme (magenta), ligating Cys (orange) and highly conserved residues T247 (green) and G350 (orange). C) Open (red, PDB: 1SE6) and closed (cyan, PDB: 2D09) conformations of the F-G region and B-C loop of CYP158A2.

In addition, the HxxxR motif in the C-helix is responsible for heme propionate binding is located within the active site and a highly conserved ExxR motif is part of the K helix that forms a salt-bridge.⁸⁶ As a consequence of the variation and low sequence similarity across CYP classes it is currently not possible to deduce CYP function from sequence alone.⁵⁰ Two important regions for substrate recognition and subsequent binding are the B-C and F-G loops that shape the active site entrance (Fig.1.9.B)). Both of these loops have been well studied for the conformational change (open to closed lid state)^{87, 88} upon substrate docking in the active site and it is the motion of these flexible regions that allows CYPs to accommodate substrates of different size and shape.⁸⁹ It also solved the predicament of how substrates entered the active site in the first instance as often CYP binding pockets seems inaccessible from analysis of crystal structures. CYP158A2 is an example that has crystal structures available with this region in both the open and closed lid state (figure 1.11C).⁹⁰ However, it should be noted that the conformational changes of all CYPs are not controlled by substrate binding. The substrate-free crystal structure of PikC has shown it to be in both an open and closed state conformation in this region.⁹¹

1.6 TxtE

1.6.1 Role in the biosynthesis of thaxtomin A

Thaxtomin A **11** was first isolated in 1989 by Lawrence *et al.*¹⁸ and was soon after found to demonstrate its potent phytotoxicity by inhibiting the production of cellulose in plants.⁹² The nitro group at the 4-position of the indole in thaxtomin and an L,L-configuration of the diketopiperazine ring are the minimal requirements for phytotoxicity⁹³ and biosynthetic gene cluster (figure 1.12A) is present in plant pathogenic *Streptomyces* species including *S. scabies*, *S. turgidiscabies* and *S. acidiscabies*.⁹⁴

The *txtD* gene encodes a nitric oxide synthase (NOS).⁹⁵ NOS enzymes are heme thiolate proteins containing an oxygenase domain at the N-terminal and a reductase domain at the C-terminal.⁹⁶ These enzymes catalyse the formation of NO from L-arginine **66** using O₂ and cofactors including FAD and FMN to transfer electrons, provided by NADPH, from the reductase domain to the heme iron. Nitric oxide is released from the *N*-hydroxy-L-arginine intermediate **67** resulting in the formation of L-citrulline **68**. Deletion of this gene in *S. turgidiscabies* or the inclusion of NOS inhibitors in bacterial cultures resulted in diminished levels of thaxtomin A production. Furthermore, a ¹⁵N atom labelling the guanidine moiety of L-arginine was found to be transferred to the nitro group of thaxtomin A, yet the stage at which this occurred in thaxtomin A biosynthesis remained unclear.

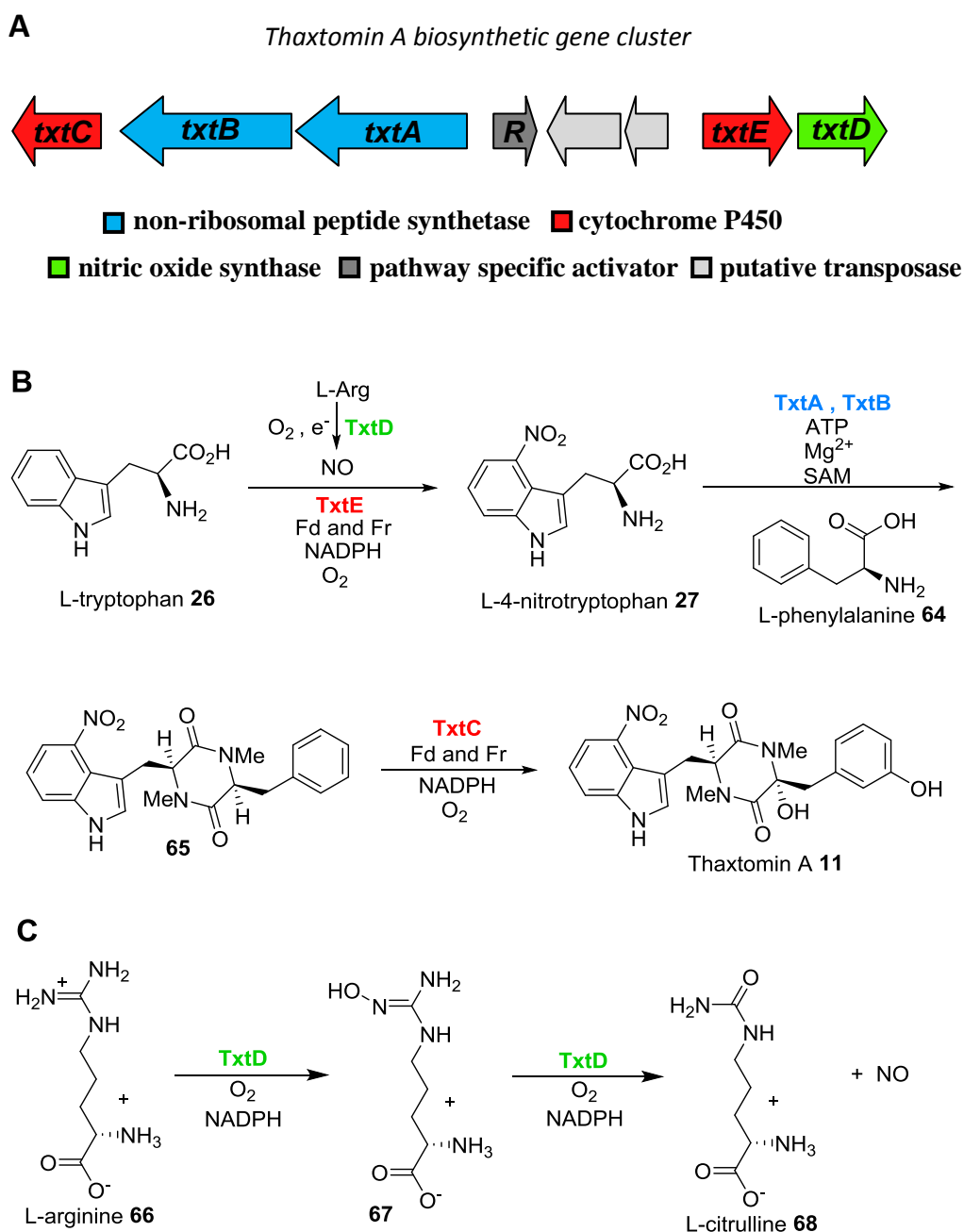


Figure 1.12. Biosynthesis of thaxtomin A. A) Arrangement of the genes in the thaxtomin A 11 biosynthetic gene cluster B) Proposed biosynthesis of thaxtomin A C) Production of NO by the nitric oxide synthase TxtD.

The genes *txtA* and *txtB* were found to encode non-ribosomal peptide synthetases⁹⁷ - large multimodular enzymes that recruit and activate or modify covalently bound substrates before catalysing peptide bond formation with a substrate tethered to an adjacent module.⁷ Knockout mutants in which *txtA* and *txtB* had been individually deleted from *Streptomyces acidiscabies* revealed that 4-nitro-L-tryptophan **27** is a substrate of TxtB whilst TxtA recruits L-phenylalanine. Following *N*-methylation of the two amino acids, a cyclisation reaction forms the *N,N'*-dimethyldiketopiperazine core of thaxtomin D **65**. The *txtC* gene

encodes a CYP that is responsible for dihydroxylation of the scaffold to give the final product **11**.⁹⁹

With 4-nitro-L-tryptophan **27** biosynthesis occurring prior to selection by TxtB, attention was turned towards the role of TxtD in generating the non proteinogenic amino acid. In 2012 a unique CYP nitration enzyme was reported by Barry *et al.*¹⁰⁰ TxtE uses NO and O₂ to directly nitrate L-tryptophan **26** at the 4-position in the first dedicated step in thaxtomin A **11** biosynthesis. The NO required for this transformation is generated directly by TxtD. *In vitro* TxtE catalysis requires spinach ferredoxin (Fd) and ferredoxin reductase (Fr) to transport electrons from NADPH to the iron atom at the centre of the protein as the natural redox partner proteins remain unknown.

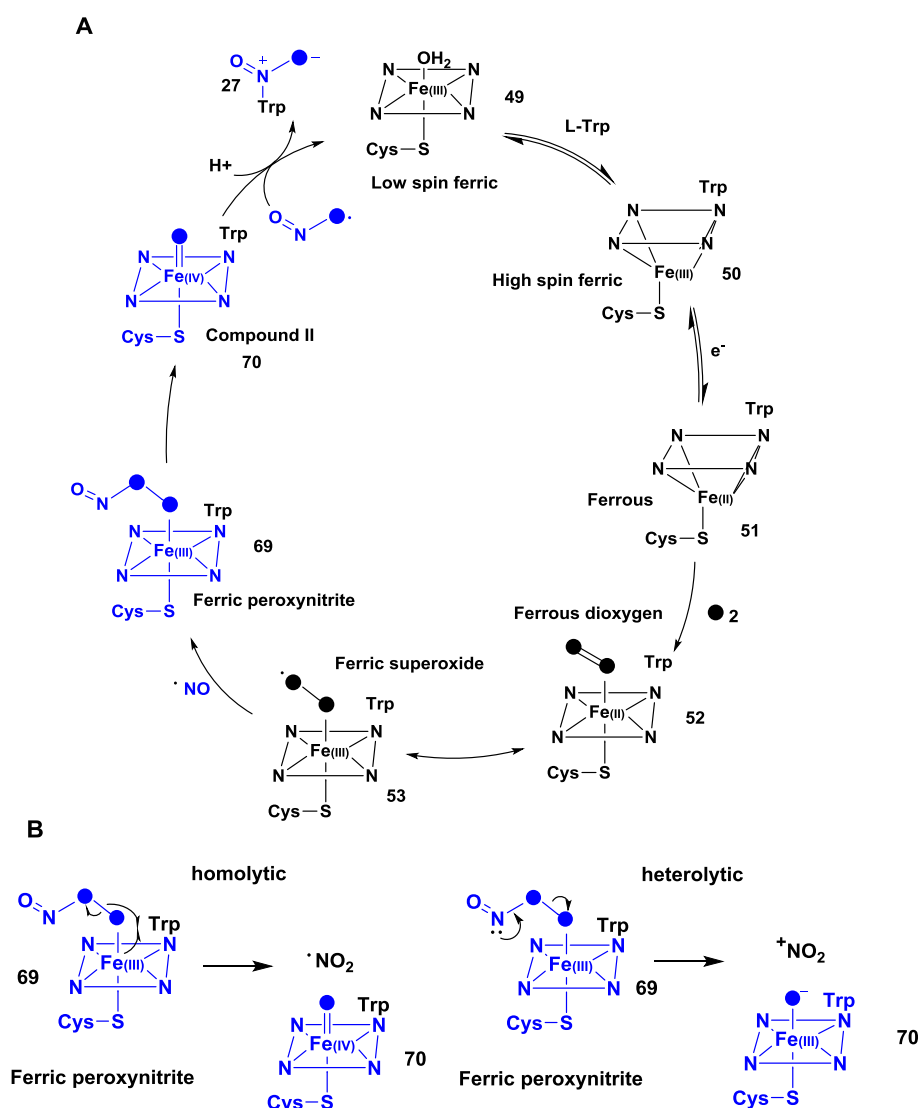


Figure 1.13. A) Proposed catalytic cycle for TxtE-catalysed nitration of L-Trp B) Possible O-O bond cleavage mechanisms in the formation of the active nitrating species NO. The pathway deviating from CYP hydroxylation is shown in blue.

The catalytic cycle proposed for L-Trp nitration is shown in figure 1.13A.¹⁰⁰ In steps analogous to CYP hydroxylation, binding of the substrate to the active site of the enzyme releases a water ligand from the ferric heme accompanied by a change in the Fe spin state from low to high. An electron donated by NADPH prompts reduction of the heme iron to its ferrous form **51** and this in turn triggers the binding of molecular oxygen **52** and generation of a ferric superoxide species **53**. The mechanism is proposed to diverge from CYP hydroxylation at this point by reaction of the ferric superoxide complex with NO to provide the ferric peroxynitrite intermediate **69**. The NO₂ species required for direct nitration is formed via cleavage of the O-O bond of the peroxynitrite complex. This can occur by two mechanisms: homolytic cleavage to provide nitrogen dioxide or heterolytic cleavage to give a nitronium cation (figure 1.13B). Details of the TxtE-catalysed nitration mechanism are discussed in Chapter 5.

1.6.2 TxtE structure and sequence

The industrial potential of TxtE has led several efforts to characterise the crystal structure of this unique enzyme. The first substrate-free (imidazole bound) crystal structure was deduced by Yu *et al.*¹⁰¹ and was later confirmed by the substrate bound structure by Dodani *et al.*¹⁰² (Figure 1.14). Generally the characteristic folds found for CYPs are conserved in TxtE, although the precise positions of these folds differ considerably, especially in the regions which are known to control substrate specificity such as the B-C loop. The typical trigonal-prism like folds parallel to the heme are present and most of the α -helices (A-L) are in positions commonly found for CYPs.

A clearly defined substrate access channel is observed between helices B'1, F and G although the F-G loop which is usually highly disordered is unresolved (figure 1.14C). TxtE has two B' helices (α B'1 and α B'2) near this substrate channel which is a rare feature and together these form a hydrophobic bundle that protects the binding pocket from external solvent and allows easy access by the substrate. Only three other CYPs (AurH,¹⁰³ CYP121¹⁰⁴ and CalO₂¹⁰⁵) are known to form a two helix bundle similar to TxtE.

For catalytic activity, CYPs require protons for the formation of catalytic intermediates and the regeneration of the enzyme resting state.⁵³ These protons are proposed to originate from external solvent waters, however very few CYP structures show a clearly defined water molecule channel. Of the few examples which have been well-characterised (CYP158A2⁹⁰ and P450BM3¹⁰⁶), the hydrogen bond chain is almost vertical to the plane of the heme. Interestingly, TxtE has a unique water molecule network in which the channel is almost parallel to the heme plane and runs from near the B'1 helix towards the I helix and through to external water molecules (figure 1.14D).

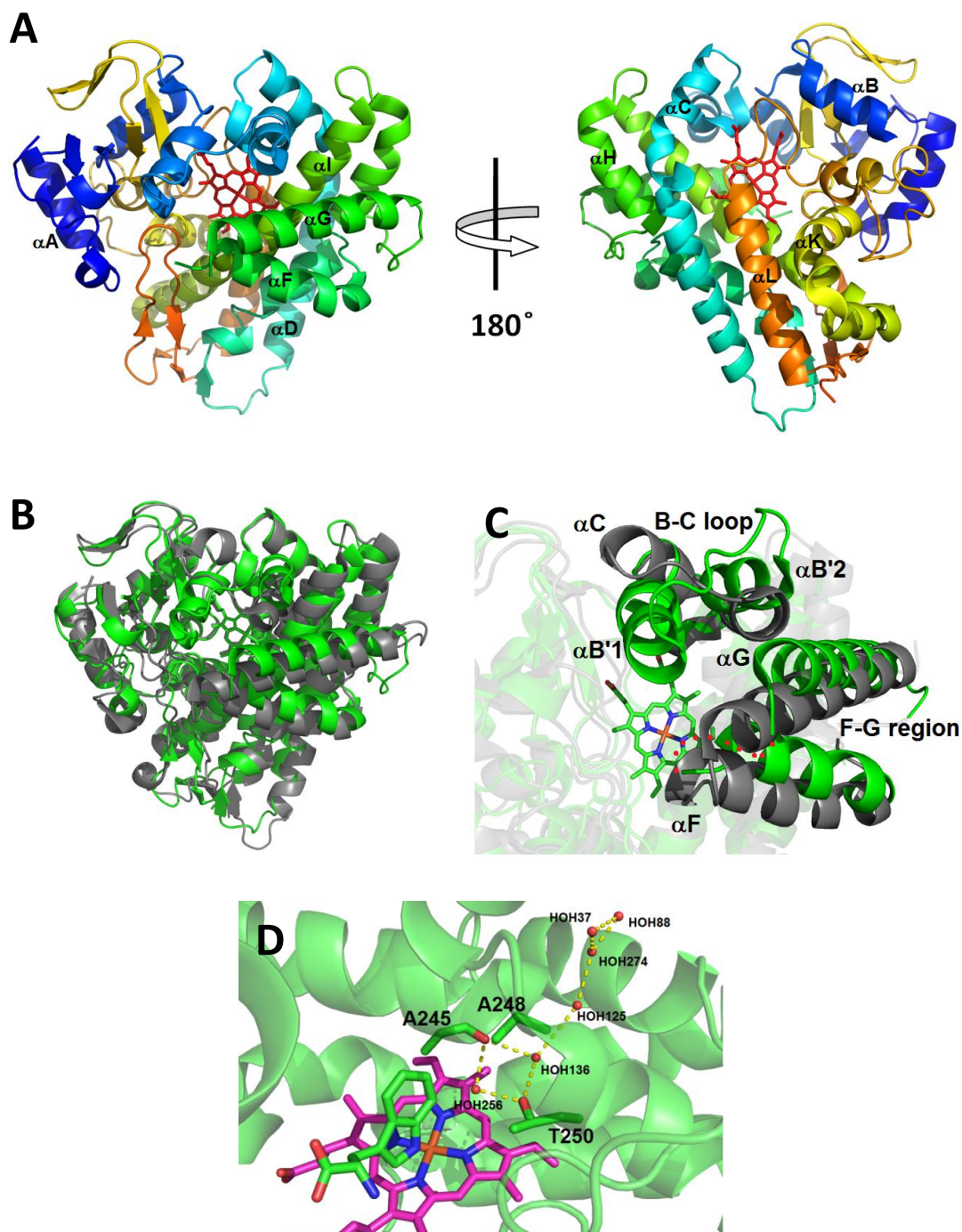


Figure 1.14. A) Crystal structure of TxtE (PDB: 4TPO, substrate not shown) B) Superimposition of TxtE (green) and PikC (grey, PDB: 2BVJ) for comparison of structural features. C) B-C and F-G regions of TxtE and PikC showing a clear substrate access channel to the heme. TxtE has two B' helices in this region compared to PikC. The unresolved TxtE F-G loop is shown by red dots. D) Unusual water network in TxtE extending from HOH256 in the active site to the external HOH88 via a kink in the I-helix near A248. Heme is shown in cyan and waters are shown as red spheres.

In terms of TxtE sequence, there is a significant difference in the region containing the catalytically-important I helix compared to other CYPs. Firstly compared to the PikC sequence (AGHETT), TxtE has an additional residue between A245 and T250 (ALFAPTT). This causes a kink to form at the additional alanine in the TxtE I-helix as a result of hydrogen bonding between the A248 backbone and nearby A245. Other CYPs have shown a kink in this region of the structure (CYP101A1),¹⁰⁷ although it is found to be wider in TxtE and potentially accounts for the water network extending to the external solvent molecules. The T250 residue considered important for dioxygen activation is conserved. Compared to the general Cys pocket consensus (section 1.5.5), G359 in TxtE is mutated to Ala and also a conserved glutamine (E366) residue in the L-helix is mutated to Q.

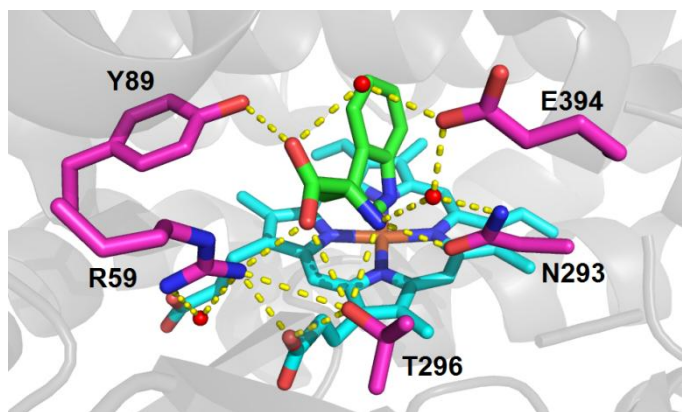


Figure 1.15. The structure of the L-Trp bound TxtE active site. Substrate is in green and binding residues are shown in magenta, heme is in cyan and connecting waters are shown as red spheres.

Focussing on the active site uncovers several important amino acid residues that contribute to L-Trp binding (figure 1.15). Thr296 and Glu394 are important for binding the amine and hydroxyl groups of the amino acid side chain whilst Tyr89 binds the carbonyl group. These interactions keep the substrate anchored in the active site and hold the indole ring above the heme. Rearrangements of the B-C loop containing the two B' helices are known to be crucial for substrate binding as discussed in section 1.5.5. This is also the case for TxtE where Arg59 changes conformation on binding of the substrate in order to provide an interaction with the substrate carboxylate and the heme propionate. Further details of substrate binding interactions are discussed in section 3.1.1.

1.6.2.1 TxtE flexible loop

In March 2016, Dodani *et al.* computationally constructed the F-G loop of TxtE.¹⁰⁷ Previously deemed too disordered to be sufficiently resolved by X-ray crystallography, *in silico* experiments were conducted to reconstruct this loop using homology based methods. Molecular dynamic simulations were carried out on the resulting structure which suggested the F-G loop of TxtE switches between two main modes: one features an open-lid state which reveals the solvent-accessible active site and the other, a closed-lid state, cuts off the active site from solvent waters. Most strikingly, His176, a residue previously unresolved, is observed to be in direct contact with the tryptophan substrate via a close edge-to-face interaction. In none of the open-lid conformations is His176 in contact with the substrate, and instead it points away from the binding pocket.

Furthermore, the indole ring of the **26** occupies the active site in two distinct conformations. One conformation resembles that of the substrate-bound TxtE crystal structure with the indole NH aimed towards the heme iron. The other, described as a ‘flipped’ state, has the indole NH aimed upwards with position 4 of **26** aimed at the heme iron instead, possibly accounting for the observed regioselectivity.

Site-saturation mutagenesis at His176 revealed modification to Phe, Tyr and Trp all resulted in exclusive formation of 5-nitro-L-tryptophan **71**. Computational studies suggested the increased steric bulk provided by these mutations positioned L-Trp such that the indole C-5 position of **26** is now directed towards the heme rather than C-4.

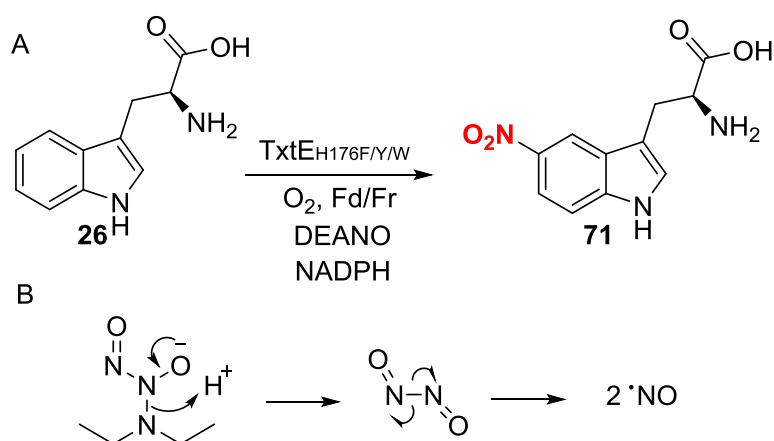


Figure 1.16. A) TxtE_{H176F/Y/W} catalysed nitration of **26 to **71** demonstrating a regioselectivity switch at position 176. B) pH-determined release of NO radical from the nitric oxide donor DEANO**

1.6.3 TxtE homologues and mutants

BLAST analysis uncovered TxtE homologues (>65 % identity) of unknown function that possess a Trp176 residue in their sequence.¹⁰⁷ These include CYPs from *Streptomyces virginiae* (Uniprot ID K9MYX3), *Streptomyces sp.* Mg1 (B4VER3), *Streptomyces lavendulae* (I2FGE0) and *Saccharomonospora marina* XMU15 (H5X7S6). The genes encoding all four enzymes were synthesised and overproduced in *E. coli*. Analysis of activity assays where these proteins were incubated with L-Trp under nitrating conditions determined that these homologues are 5-nitrotryptophan producing CYPs. Mutation of the Trp176 residue in the CYP from *S. virginiae* reverted production to 4-nitro-L-tryptophan, indicating the residue at position 176 acts as a regioselectivity switch for natural nitrating CYPs. Engineering the flexible loop containing this residue for the control of regioselectivity could prove valuable in developing TxtE into a useful industrial biocatalyst.

In 2017, Tomita *et al.* reported the discovery a tyrosine nitration CYP from *Streptomyces atratus*. RufO is involved in the biosynthesis of the rufomycin, a cyclic heptapeptide that has demonstrated anti-mycobacterium activity.¹⁰⁸ The discovery of RufO was aided by analysis of the rufomycin biosynthetic gene cluster which also contains the NOS RufN. RufO catalyses the 3-nitration of tyrosine as demonstrated by *in vitro* assays and yet only shares a 26 % amino acid identity with TxtE. The future discovery of related nitration CYPs relies strongly on the identification of biosynthetic gene clusters containing both a NOS and a CYP.

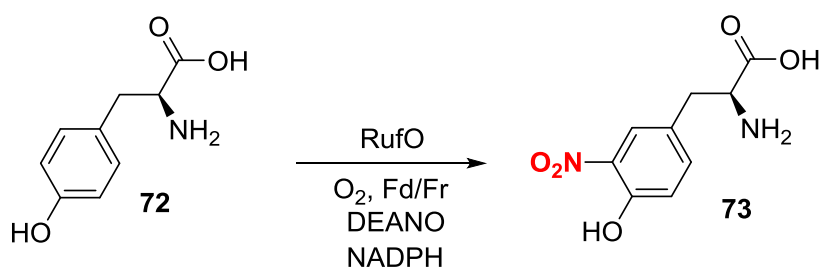


Figure 1.16. RufO-catalysed nitration of L-Tyr to 3-nitro-L-tyrosine in the biosynthesis of rufomycin.

1.6.4 TxtE substrate scope

TxtE has been investigated for its ability to bind and/or nitrate numerous non-native substrates by Dodani *et al.*¹⁰² Binding was determined by UV-visible spectroscopy and is explained in greater detail in chapters 2 and 3. All of the substrates that were found to undergo nitration are depicted in figure 1.17 along with a selection of substrates that either bind to TxtE in a type I fashion without resulting in nitration and those that show type II binding. L-Trp analogues bearing fluoro- (**74**, **79**), methyl- (**76**, **77**, **78**) or methoxy- (**75**) substituents on the indole ring were readily nitrated by TxtE. This included a 4-methyl derivative **76** of L-Trp that implies nitration regioselectivity is flexible. However, substitution by the more electron withdrawing hydroxyl- (**84**) and amino-groups (**85**) abolished reactivity despite these substrates binding in a type I productive fashion similar to L-Trp. This suggests that these polar groups either result in unfavourable interactions with reactive heme-bound species or reduce the electron density of the indole ring beyond the threshold for nitration. Minor alterations to the amino acid chain were tolerated including β -carbon (**81**) and primary amine (**80**) methylation. Addition of a methyl group the indole nitrogen however resulted in loss of productivity even though this substrate demonstrated type I binding. Several substrates (**90**, **91**) adopted type II binding modes that indicate direct coordination to the heme iron and therefore prevent catalysis. In these studies none of the products were isolated and thus the nitration regioselectivity displayed by TxtE upon these substrates was undetermined.

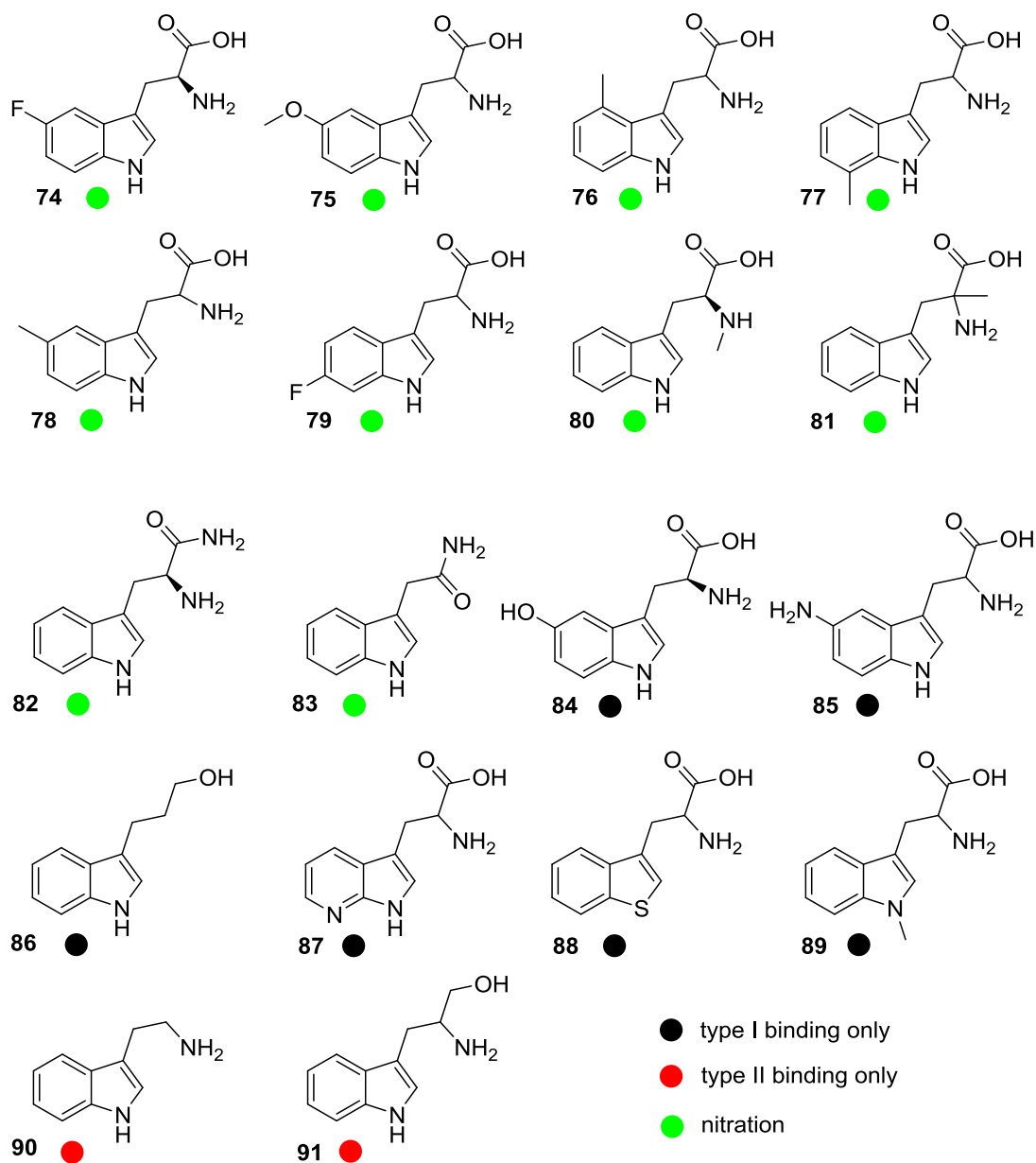


Figure 1.17. Selected L-Trp analogues previously investigated for nitration by TxtE. Substrates that were readily nitrated are shown by green dots and those that did not nitrate but did bind in a productive manner are shown by black dots. Substrates that did not nitrate and bound to the TxtE heme directly are shown by red dots.

1.7 Introduction to enzyme engineering

Given the impeccable chemo-, regio- and stereoselectivity of enzymes, it is often desirable to use biocatalysts for the synthesis of high value products such as pharmaceuticals, agrochemicals and biofuels.¹⁰⁹ Furthermore, enzymes offer an environmentally-friendly alternative to organocatalysts as they are biodegradable and perform under mild conditions in aqueous solutions. However, utilising enzymes for industrial purposes usually requires engineering them to improve their catalytic activity, stability or substrate specificity compared to the wild-type protein.¹¹⁰ As a result of the vast range of biochemical reactions catalysed by CYPs, these enzymes have been the subject of numerous engineering studies.

1.7.1 Directed evolution

Directed evolution is a Darwinian approach to enzyme engineering that mimics natural selection.¹⁰⁹ Proteins are often evolved over several rounds consisting of gene mutagenesis, enzyme variant expression and selection based on the ability of the variant to conduct the desired transformation. The best performing mutant from the round is often isolated and used as a template for the following round. This process is repeated until the desired characteristics have been obtained. In general, there are two established methods for protein engineering: random mutagenesis and rational/targeted design, but neither of these are mutually exclusive and often a combination of methods would be used.

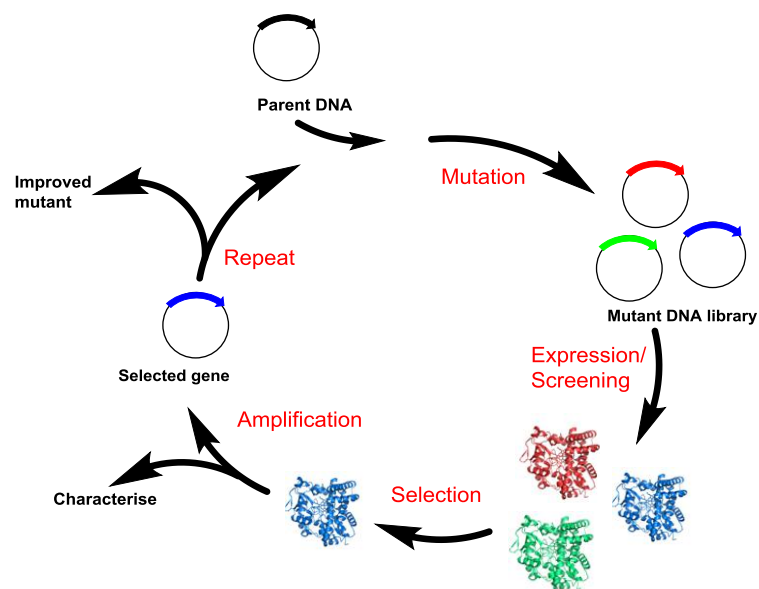


Figure 1.18. Typical directed evolution life cycle and breakdown of the stages involved.

1.7.1.1 Random mutagenesis

One of the simplest directed evolution strategies involves random mutagenesis, via error-prone PCR, across the entire protein domain.¹¹¹ This approach requires no structural information and the number of mutations incorporated can be tuned depending on how tolerating an enzyme is to modification. Given the diversity introduced into the enzymes structure, this method often unveils unexpected beneficial mutations that improve catalytic activity in regions not easily identified by analysing protein structure alone.

1.7.1.2 Targeted mutagenesis

Another approach utilises specific structural or biochemical data to modify an enzymes properties in a targeted manner.¹¹² For example, if it is desirable to broaden the substrate scope of an enzyme, mutation of the active site residues would be expected to have the greatest effect. High-resolution crystal structure data provides the means of targeting these sites which can either be modified singly, or in combination.

1.7.1.3 Examples of the directed evolution of CYPs

P450_{BM3} has been extensively engineered as a result of its high activity and good stability. It has been the subject of both random and targeted mutagenesis.

In a study by Farinas *et al.*, two rounds of evolution produced a variant that possessed five times the activity towards the saturated hydrocarbon octane **92** than the wild-type protein.¹¹³ Sequencing of the mutant revealed Glu236, located on a helix close to the surface of the protein (P450_{BM3} 139-3), was mutated to His. Being approximately 17 Å from the heme cofactor, and in neither the active site nor the substrate access channel, suggested this residue would unlikely have been identified for mutation in targeted studies. In further studies, engineered P450_{BM3} variants have also been found to catalyse the epoxidation of terminal alkenes (**94** to **95**)¹¹⁴ and have even been used for the *in vitro* production of the anti-anxiety drug 6-hydroxybuspirone (**96** to **97**)¹¹⁵ which can be produced in humans by metabolism of buspirone by CYP3A4. P450_{BM3} has also been engineered to catalyse demethylation of protected sugar molecules,¹¹⁶ intramolecular C-H amination^{117, 118} and cyclopropanation.¹¹⁹

P450_{BM3} has also undergone rational design based upon the multitude of structural data available. Seifert *et al.* utilised this data to construct a minimal library of 24 mutants which were simultaneously modified at two positions close to the heme cofactor by five hydrophobic residues (Ala, Val, Phe, Leu and Ile).¹²⁰ The library contained variants of considerably different size and shape binding pockets compared to the wild-type enzyme. Screening for activity with four terpenes revealed mutants capable of oxidising (+)-valencene **98** to the high-value compound (+)-nootkatone **99**, which is not produced by the

wild-type enzyme. The best performing mutant (F87A/A328I) produced the expensive grapefruit aromatic in 26 % conversion.

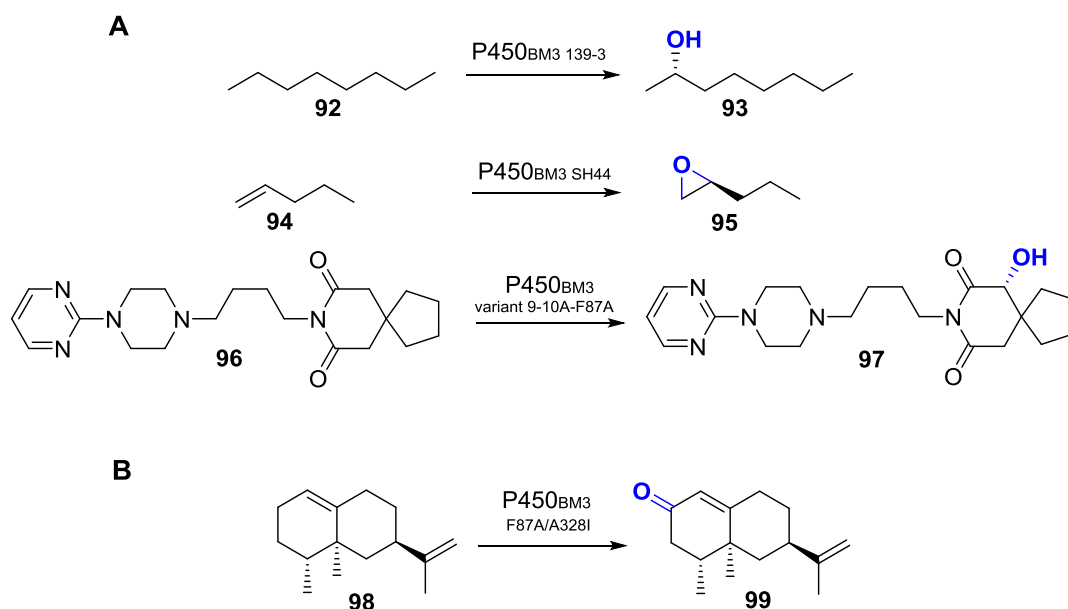


Figure 1.19. Laboratory evolution of P450_{BM3} A) Reactions catalysed by P450_{BM3} mutants generated from random mutagenesis experiments B) P450_{BM3} mutants from a minimal targeted library catalyse hydroxylation of valencene to nootkatone.

1.8 Project aims

For nitration of aromatic compounds by TxtE to be feasible on an industrial scale, the enzyme must be able to demonstrate high activity for a number of diverse substrates. In order to generate TxtE variants easily and efficiently on a high-throughput platform, the aim was to prepare a one-component TxtE-reductase fusion protein.

Following this, TxtE mutant libraries are planned to be prepared and screened against L-Trp to identify variants with increased activity with the natural substrate. In order to achieve this, a reliable assay in 96-well plates and was proposed to be developed using the facilities at GlaxoSmithKline, Medicines research centre, UK.

With an established assay available, the final aim was to identify a TxtE variant capable of catalysing nitration of tryptamine. The substrate tolerance of such a variant could then be investigated.

Chapter 2: Investigating wild-type TxtE and creating a self-sufficient fusion protein

2. Overview

The activity and limited substrate tolerance of wt-TxtE has previously been investigated by Dodani *et al.*¹⁰² This chapter discusses the overproduction and purification of TxtE as well as the use of tyrosine analogues as mechanistic probes.

Efforts were then directed towards creating a self-sufficient cytochrome P450-reductase fusion protein (CYP-Red) to aid scale-up reactions and downstream engineering where the expense and inconvenience of CYP redox cofactors can become an issue. TxtE was fused to the reductase domains of two different naturally occurring CYP-Red enzymes and their activity with the natural substrate L-Trp was analysed. The most active of these fusion proteins was then utilised in both high-throughput substrate screening and large scale nitration reactions.

2.1 Purifying TxtE and reconstructing the in vitro activity

TxtE was overproduced in *E. coli* using plasmid pSB13 prepared by S. Barry.¹⁰⁰ This plasmid contains the *txtE* gene from *S. scabiei* 87-22 cloned into the pET151/D-TOPO vector to allow the production of recombinant N-terminal His₆-tagged TxtE. Expression was induced via the addition of isopropyl β -D-1-thiogalactopyranoside (IPTG) to the bacterial culture, resulting in release of the *lac* repressor from the operon. T7 RNA polymerase encoded by *E. coli* BL21 (DE3) cell lines is now permitted to bind to the T7 promoter region of the vector resulting in efficient transcription of the gene of interest, including a hexahistidine tag to aid purification. A frozen stock of pSB13 was used to transform *E. coli* BL21 (DE3) from which a single colony was selected for overproduction of the protein. Overproduction of TxtE in ampicillin-supplemented Luria-Bertani (LB) medium at 15 °C resulted in the clear presence of a band of the expected size (48.5 kDa) in the SDS-PAGE gel (figure 2.1). Purification by immobilised metal affinity chromatography (IMAC) using a nickel-charged HisTrap column and buffer exchange by PD-10 column to remove imidazole, a known inhibitor of CYPs, yielded highly pure protein suitable for further characterisation.

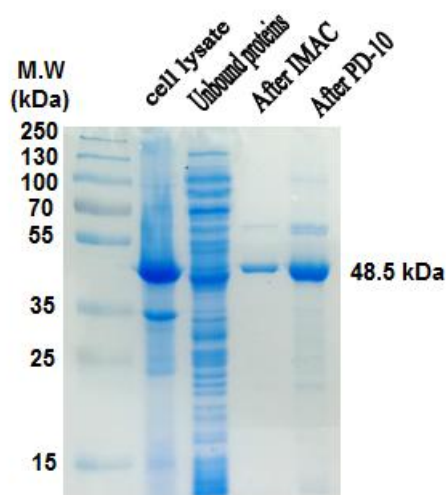


Figure 2.1. SDS-Page analysis of TxtE purification. Purified His₆-TxtE (48.5 kDa) after IMAC and PD-10 column (right lanes). Cell lysate and unbound protein fractions (middle lane). Molecular weight marker is in the left lane.

2.2 UV-Visible spectroscopic characterisation of TxtE

UV-Visible spectroscopy is a valuable tool for characterising CYPs as a result of the distinct spectral properties exhibited by the sulphur-coordinated heme species.⁵³ The resting state TxtE-Fe(III) complex was shown to have a λ_{max} of 417 nm (figure 2.2). The TxtE Fe(III)-NO complex was observed by adding DEANO to a solution of the protein and the resulting species showed a λ_{max} of 431 nm in the UV-Vis spectrum. Adding a solution of sodium dithionite to TxtE produced the Fe(II) complex which exhibited a λ_{max} of 404 nm in UV-Vis spectrum. The spectra are all consistent with the presence of a heme active site.

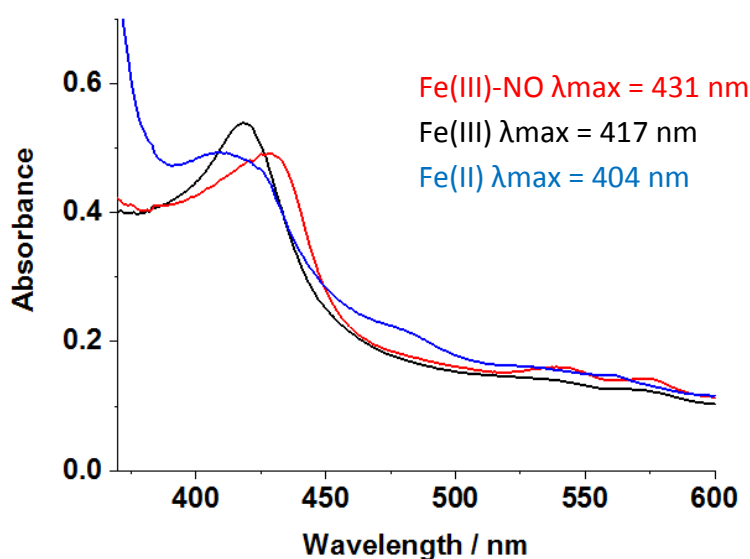


Figure 2.2. UV-Vis spectroscopic analysis of TxtE. TxtE in 25 mM Tris-HCl, pH 8 (black), TxtE + sodium dithionite (blue), TxtE + DEANO (red).

2.3 Analysis of the TxtE-catalysed nitration reaction

In order to assess the catalytic activity of this nitration enzyme, TxtE was incubated with 0.5 mM **26**, the redox cofactors ferredoxin (Fd) and ferredoxin reductase (Fr), both obtained from *Spinachia oleracea*, the NO donor DEANO and NADPH. The reaction was incubated at 25 °C for 2 hours in which time the solution became yellow, consistent with the formation of nitroindoles.¹⁰⁰ The reaction mixture was analysed by LC-MS (figure 2.3) and a product was identified with m/z 250.0824, which was consistent with the molecular formula, and was absent from the control reactions. The existence of a single peak with the corresponding m/z value is consistent with the regioselective formation of **27**.

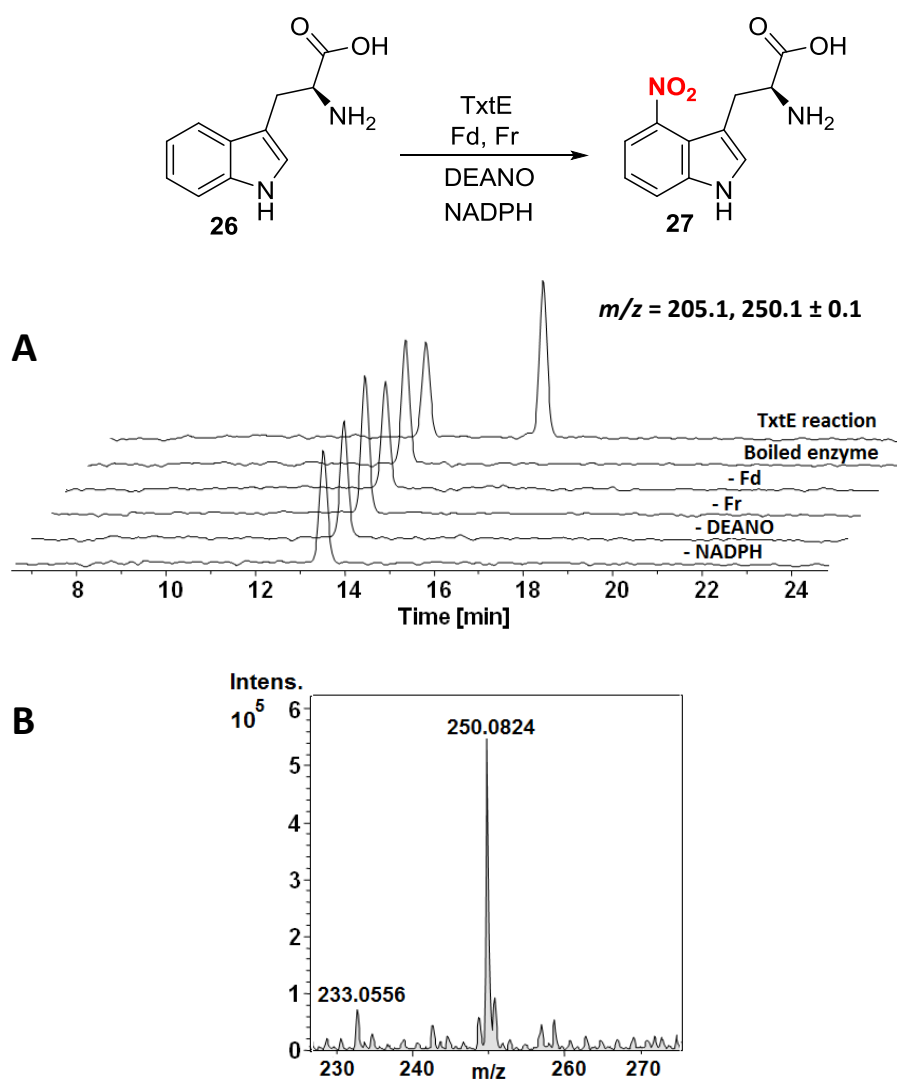


Figure 2.3. Analysis of the TxtE-catalysed L-Trp nitration reaction and controls. A) LC-MS EICs at $m/z = 205.1$ and 250.1 corresponding to $[M+H]^+$ for **26** and **27**, respectively. B) ESI-Q-TOF mass spectrum of the product peak **27**.

2.4 Investigating the substrate scope of TxtE

To assess the substrate tolerance of the wild-type enzyme a range of L-tryptophan analogues were tested by Dodani *et al.* (section 1.6.4).¹⁰² Interestingly methyl substitution at the 1-position of L-Trp (**89**) completely eliminates nitration despite the substrate binding in the active site. This may be because the mechanism requires the hydrogen on the indole nitrogen to be abstracted (section 1.6.1).¹⁰⁰ To further investigate this hypothesis, *ortho*- and *meta*-tyrosine (*o*-Tyr and *m*-Tyr, respectively) were selected as mechanistic probes. Both compounds have the potential to donate a proton via the hydroxyl groups present on the benzyl moiety and these groups would be expected to occupy a similar position in space to the indole N-H of L-Trp within the active site of TxtE.

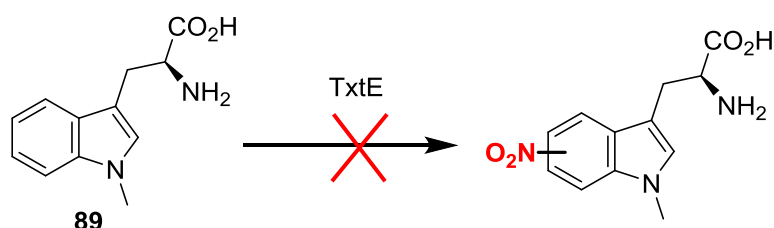


Figure 2.4. Scheme representing lack of nitro-*N*-methyltryptophan product when incubated with TxtE under reaction conditions.

Qualitative binding assays were first carried out with L-Trp and figure 2.5 shows a typical type I difference spectrum produced as the concentration of substrate added to the TxtE solution is gradually increased. Type I spectra are produced when the substrate does not bind directly to the iron atom of the heme.¹²¹ An increase in absorbance at 390 nm and a decrease at 420 nm is the result of the water molecule coordinated to the Fe atom of the heme in the resting state being displaced upon substrate binding (section 1.6.1.). Plotting the difference in absorbance at 390 nm and 420 nm against the concentration of substrate for data sets obtained in triplicate and averaged gives the curve shown in figure 2.5.B. The data was fitted to a Michaelis-Menten model in Origin 9.1 Pro to give the dissociation constant, K_d , of the substrate, using equation 1.1. The dissociation constant of L-tryptophan with TxtE was calculated to be $77.6 \pm 5 \mu\text{M}$, in agreement with a previously calculated value of $60 \pm 6 \mu\text{M}$.¹⁰⁰

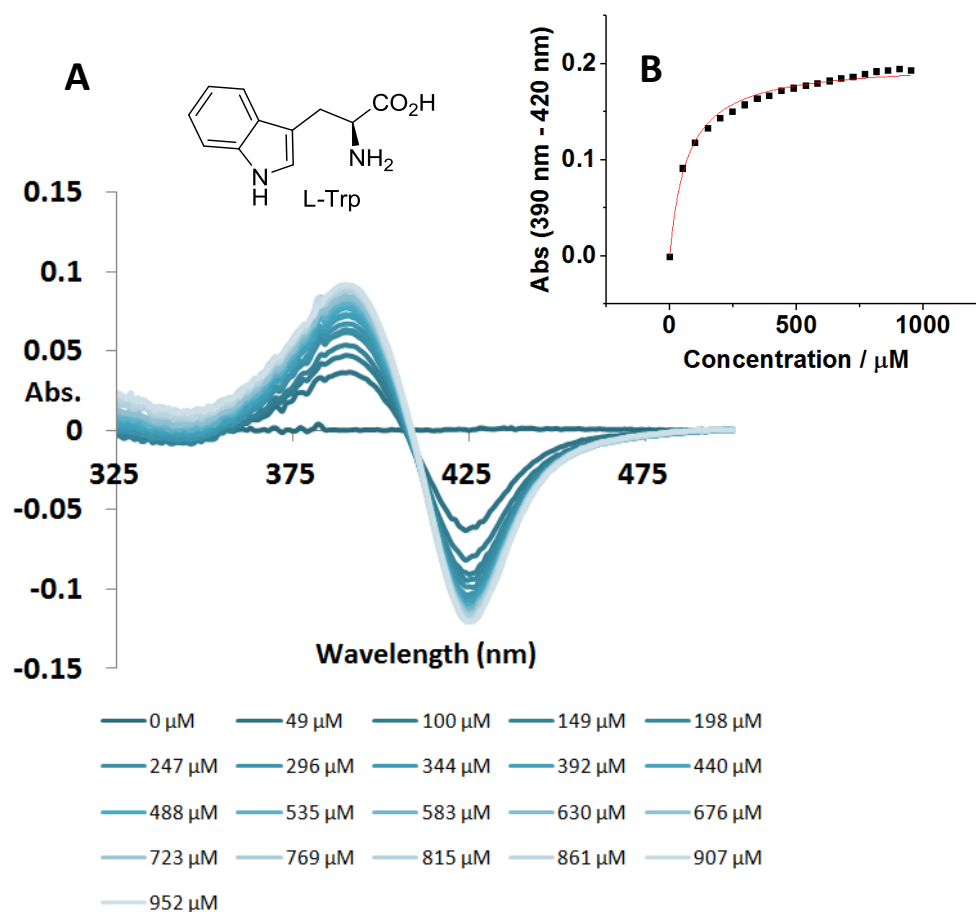


Figure 2.5. UV-Vis spectroscopic analysis of L-Trp binding to TxtE. **A)** Type I difference spectra from titration of TxtE (8 μM) with L-Trp (0-1 mM). **B)** Difference in absorbance at λ_{\max} (390 nm) and λ_{\min} (420 nm) versus L-Trp concentration. Data fitted to a Michaelis-Menten model and points are the mean of three measurements.

$$Y = \frac{B_{\max} * X}{(K_d + X)}$$

Equation 2.1 Determination of K_d from binding curves

Similar binding assays were carried out for the tyrosine analogues. Both molecules exhibited typical type I binding spectra upon titration to TxtE (figure 2.6) and dissociation constants were calculated from the corresponding binding curves. Although considerably higher than that for L-Trp, K_d values of $452 \pm 52 \mu\text{M}$ and $1160 \pm 142 \mu\text{M}$ for *o*-Tyr and *m*-Tyr, respectively, demonstrates that both molecules are able to bind to TxtE.

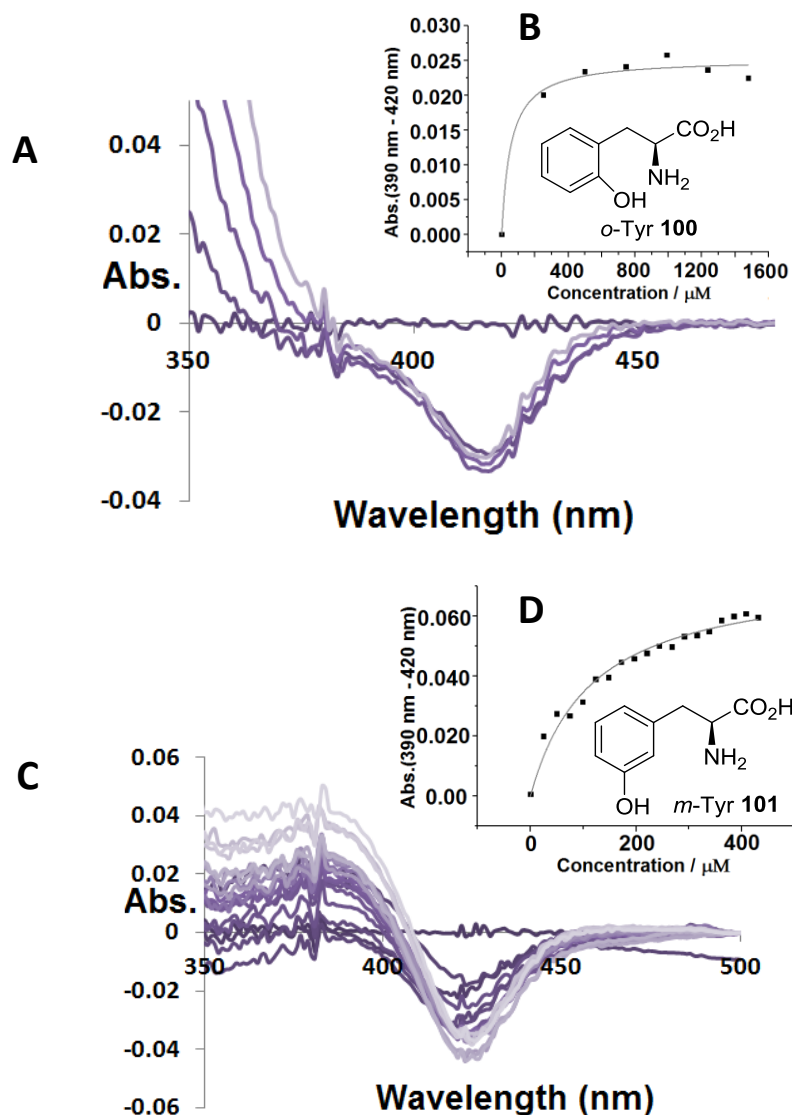


Figure 2.6. UV-Vis spectroscopic analysis of Tyr analogue binding to TxtE. A) and C) Difference spectra from titration of TxtE (8 μ M) with *o*-Tyr and *m*-Tyr (0-1 mM), respectively. B) and D) Difference in absorbance at λ_{\max} (390 nm) and λ_{\min} (420 nm) versus *o*-Tyr and *m*-Tyr concentration, respectively. Data fitted to a Michaelis-Menten model and points are the mean of three measurements.

Table 2.1. Summary of substrate binding modes and determined K_d values

Substrate	Binding mode	K_d (μ M)
L-Trp	Type I	77.6 ± 5
<i>o</i> -Tyr	Type I	452 ± 52
<i>m</i> -Tyr	Type I	1160 ± 142

Encouraged by the preliminary binding results, activity assays were then employed in order to investigate whether *o*-Tyr and *m*-Tyr are able to undergo TxtE-catalysed nitration. The enzyme was incubated with 1.0 mM final concentration of substrate rather than the 0.5 mM substrate used for L-Trp to account for the weaker binding. Redox cofactors, Fd and Fr, DEANO and NADPH were added and the assay was incubated for 2 hours at 25 °C. Unfortunately, LCMS analysis revealed that both *o*-Tyr and *m*-Tyr undergo a large degree of background nitration in the presence of the NO donor, independent of TxtE (figure 2.7). Peaks corresponding to *m/z* 227 can be observed in the reaction assays as well as the boiled enzyme control assays for both substrates. This phenomenon can be attributed to the formation of peroxynitrite *in situ* from superoxide and nitric oxide present on the reaction mixture (figure 2.8) which is both an oxidant and powerful nitrating agent.¹²² This reaction is only observed for *o*-Tyr and *m*-Tyr, not L-Trp, as their aromatic rings are much more activated. No increase in product formation was observed in the presence of TxtE, thus formation of nitrotyrosine products was attributed completely to the non-catalysed reaction.

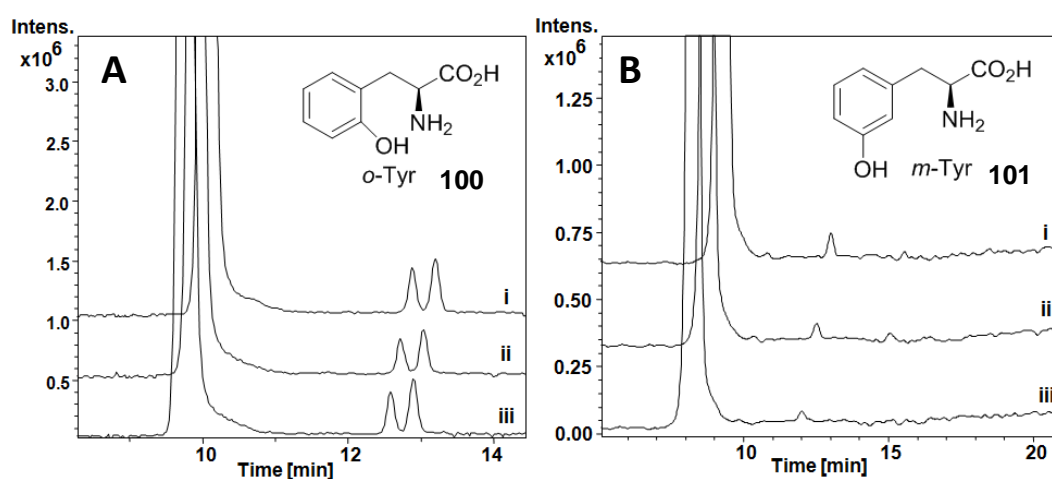


Figure 2.7. Analysis of TxtE-catalysed reaction with Tyr analogues. A) LC-MS EICs at *m/z* 182.2 and 227.2 corresponding to [M+H]⁺ *o*-Tyr and nitro-*o*-Tyr, respectively. B) LC-MS EICs at *m/z* 182.2 and 227.2 corresponding to [M+H]⁺ *m*-Tyr and nitro-*m*-Tyr, respectively. i) TxtE-catalysed reaction ii) Boiled TxtE control reaction iii) Substrate + DEANO only.

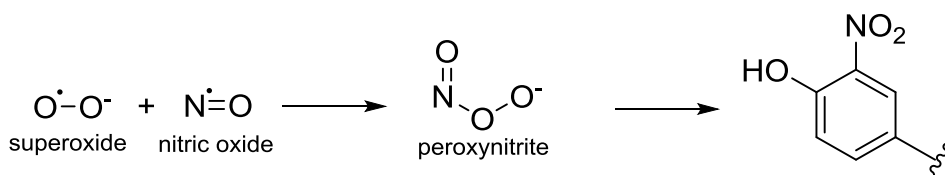


Figure 2.8. Formation of peroxynitrite species in aqueous solutions and subsequent nitration of activated aromatic species.

2.5 Preparation of self-sufficient TxtE-reductase fusion proteins

As well as saving costs, CYP-Reductase fusion enzymes can be more convenient and can potentially increase enzyme efficiency, proving advantageous for use in scale-up reactions where high quantities of enzyme reagent are required.

2.5.1 TxtE-RhFRED

The first reductase domain chosen was from P450RhF, a naturally occurring self-sufficient CYP from *Rhodococcus* sp. NCIMB 9784.⁷⁴ This domain, RhFRED, has previously been used to reconstitute the monooxygenase activity of PikC in the absence of spinach Fd and Fr, with improved kinetic parameters compared to the wild-type enzyme.⁷⁸ Furthermore, a study conducted by the Flitsch group determined that modification of the length of the 21 amino acid natural polypeptide linker between RhFRED and P450cam had a drastic effect on enzyme reactivity and improved total conversion of camphor to its 5-*exo* hydroxy derivative.¹²³ An insertion of 9-11 additional amino acids between the heme domain and RhFRED increased conversion by 10-20 times over the fused enzyme retaining the natural linker. For this reason, TxtE-RhFRED was designed with 11 additional amino acids between the P450 and the FMN domain of RhFRED as a starting point.

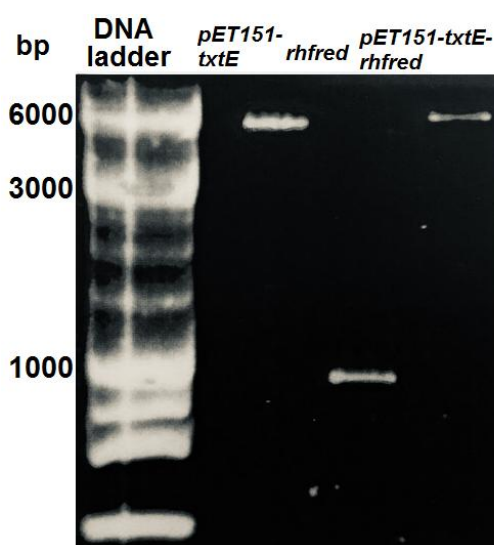


Figure 2.9. Agarose gel electrophoretic analysis of the linearised parent plasmid pET151-*txtE* (6921 bp), the DNA insert *rhfred* (1083 bp) and the final Gibson assembly product (8004 bp).

The complete *txtE-rhfred* gene was generated in pET151 by Gibson assembly.¹²⁴ The vector, pCloneEZ-*rhfred*, containing the reductase domain gene, the 32 amino acid polypeptide linker and 20 bp overlap regions required for Gibson assembly, was doubly digested with *Pvu*II and *Ac*II and purified by agarose gel electrophoresis to yield the pure

reductase domain gene fragment. pSB13 (pET151-*txtE*) was digested with Eco53KI and similarly purified to create the cleaved vector backbone ready for Gibson assembly which was achieved by using the NEBuilder HiFi DNA assembly kit (NEB) (figure 2.9) (section 7.2.1). Following identification of a positive clone, the stop codon and the nine nucleotides between the *txtE* gene and the *Eco53KI* restriction site were deleted using the Q5-Site-Directed Mutagenesis Kit (NEB) to give the final construct pET151-*txtE-rhfred*.

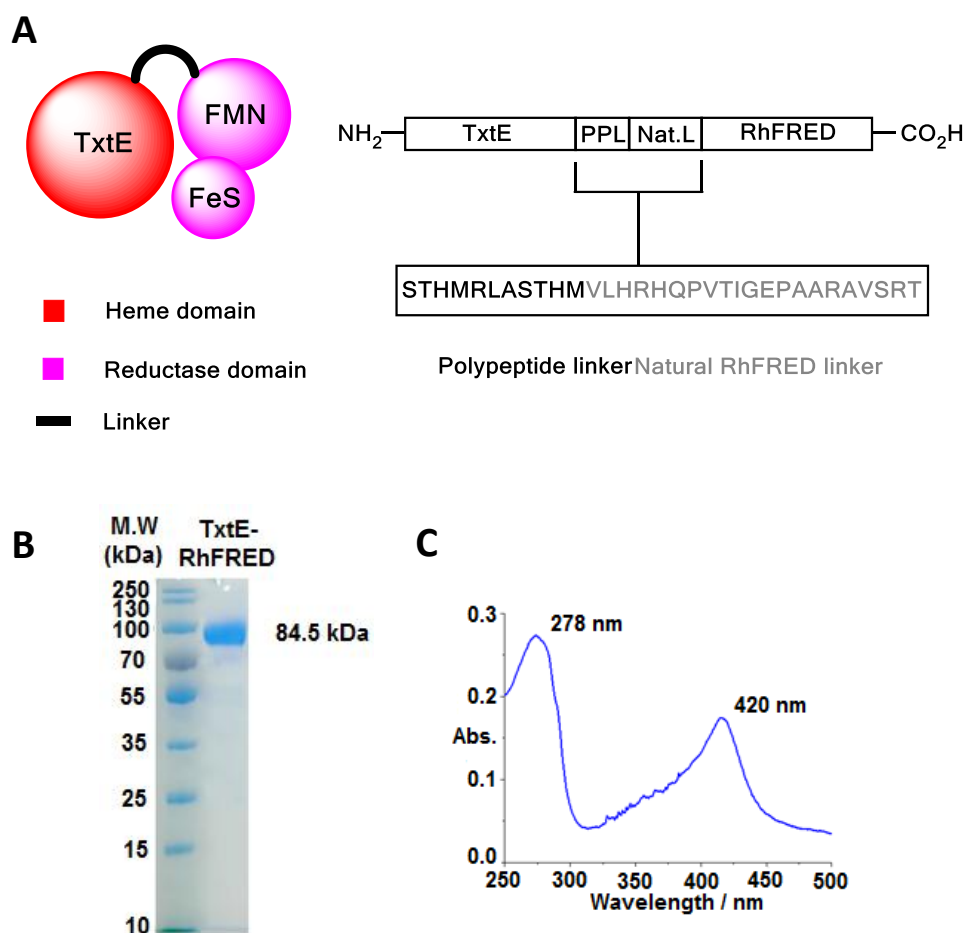


Figure 2.10. Analysis of TxE-RhFRED. A) Domain organisation of TxE-RhFRED and amino acid composition of the heme-reductase linking region including an additional polypeptide linker (PPL). B) SDS-PAGE analysis of purified TxE-RhFRED (~84.5 kDa, right lane). Molecular weight marker is in the left lane. C) UV-Vis spectroscopic analysis of TxE-RhFRED showing absorbance at 420 nm.

The hybrid cytochrome P450-reductase fusion was overproduced and purified as previously described,¹⁰⁰ except 5-aminolevulinic acid (5-ALA), FeCl₃ and thiamine were added to the culture during overproduction.¹²³ SDS-PAGE analysis confirmed a ~85 kDa protein had been purified which was consistent with N-terminal His₆-tagged TxE-

RhFRED (figure 2.10B). UV-Vis analysis showed a distinctive absorption band at 420 nm relating to the heme cofactor (figure 2.10C) indicating the presence of a large secondary domain had little effect on heme incorporation.

The binding constant was then determined for the chimera with the natural substrate **26**. Consistent with the previous study on PikC, it was found that the TxtE fusion protein binds its substrate slightly more tightly than the wild type at $55.5 \pm 1.7 \mu\text{M}$ and $77.6 \pm 5.0 \mu\text{M}$, respectively.

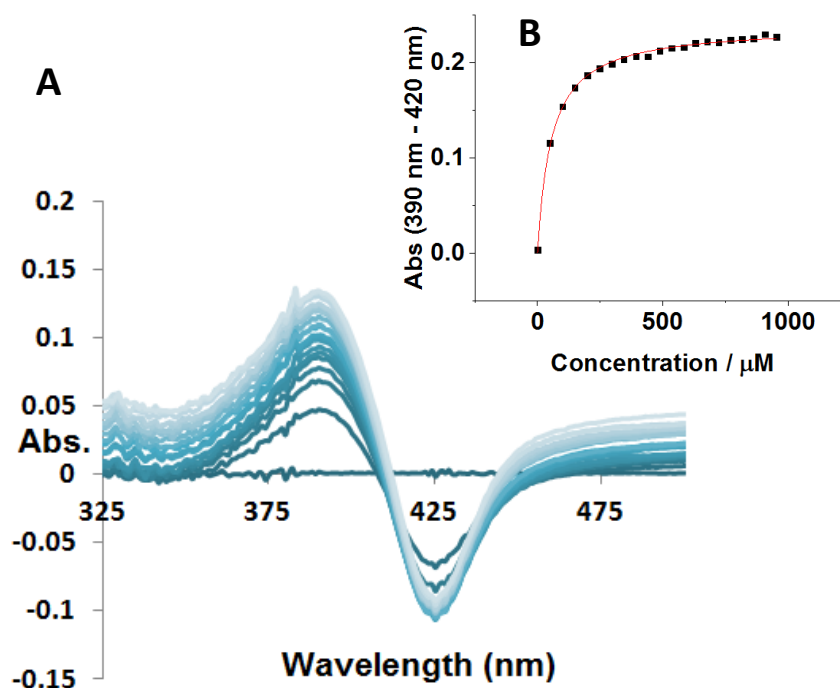


Figure 2.11. UV-Vis spectroscopic analysis of L-Trp binding to TxtE-RhFRED. A) Type I difference spectra from titration of TxtE-RhFRED (8 μM) with L-Trp (0-1 mM). B) Difference in absorbance at λ_{max} (390 nm) and λ_{min} (420 nm) versus L-Trp concentration. Data fitted to a Michaelis-Menten model and are the mean of three measurements.

The ability of TxtE-RhFRED to nitrate L-tryptophan in the absence of Fd and Fr was tested by adding the enzyme (30 μM) to a solution of the substrate and NADPH. DEANO was added to start the reaction and after 2 hours the solution had become yellow. LCMS analysis revealed a peak with mass corresponding to 4-nitrotryptophan confirming that a self-sufficient nitration biocatalyst had been created (figure 2.12). Despite these encouraging results, it was apparent that a high quantity of protein reagent was required to achieve a conversion to product ~60 % at 0.1 mM substrate concentration.

During the course of this study Zuo *et al.* described a similar work on a TxtE-RhFRED fusion protein.¹²⁵ They also observed that although TxtE-RhFRED could catalyse nitration of **26** to **27** in the absence of exogenous reductase partners, turnover was low and high

concentration of protein was required. An alternate TxtE fusion protein using the reductase domain from P450_{BM3}, BM3R, was more successful. P450_{BM3} is a self-sufficient monooxygenase from *Bacillus megaterium*,⁷¹ and its reductase domain was linked to TxtE via a 27 amino acid polypeptide linker, creating TxtE-BM3R. The BM3R domain contains a FMN binding site similar to RhFRED, but differs in the architecture of the C-terminal region of the enzyme where it contains a FAD-binding domain rather than an iron-sulphur cluster binding motif. Conversion to **27** was much improved using this variant over TxtE fused to RhFRED using either the native linker or an extended linker with 8 additional residues appended as described previously.

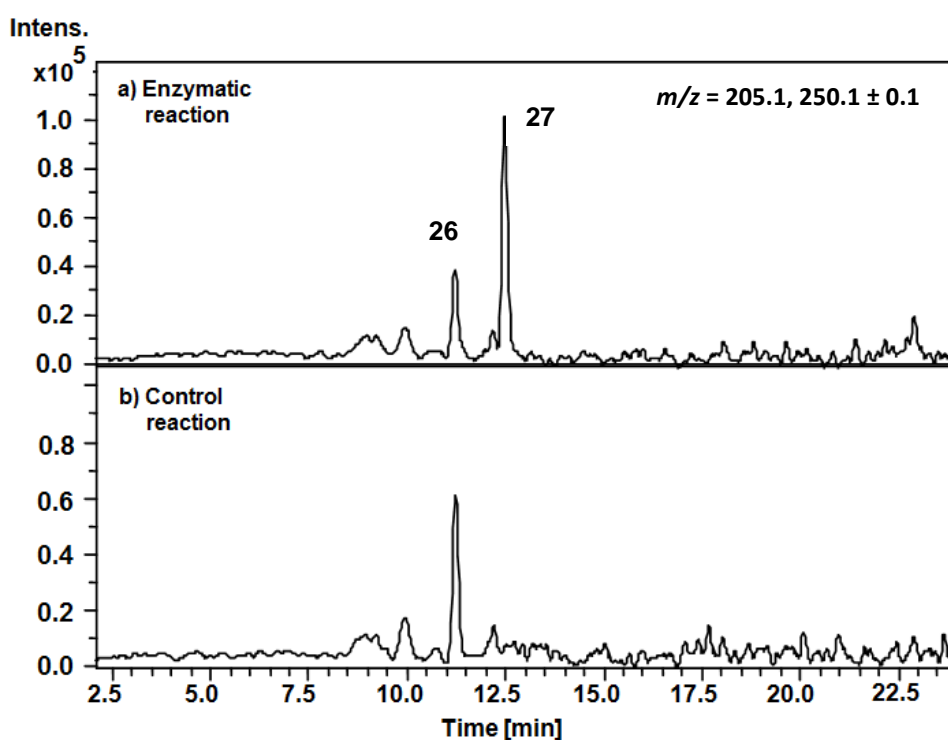


Figure 2.12. EICs at $m/z = 205.1$ and 250.1 corresponding to $[M+H]^+$ for tryptophan and 4-nitrotryptophan, respectively, from LCMS analysis of the TxtE-RhFRED-catalysed L-Trp nitration reaction (top) and control (bottom).

2.5.2 TxtE-BM3R

Our attention therefore turned towards designing and creating a TxtE-BM3R variant. The gene encoding TxtE-BM3R and natural linker was synthesised and cloned into pET28(a)+ with an N-terminal His₆-tag by GenScript, NJ, USA and designated pRS100. The entire 3036 bp gene was subjected to GenScript OptimumGeneTM Codon Optimisation Analysis, an algorithm designed to maximise protein expression in *E. coli* by taking into account codon use bias, GC content and repeated DNA sequences, amongst several other parameters (*US Patent 8,326,547: Document Identifier US 20110081708 A1*).

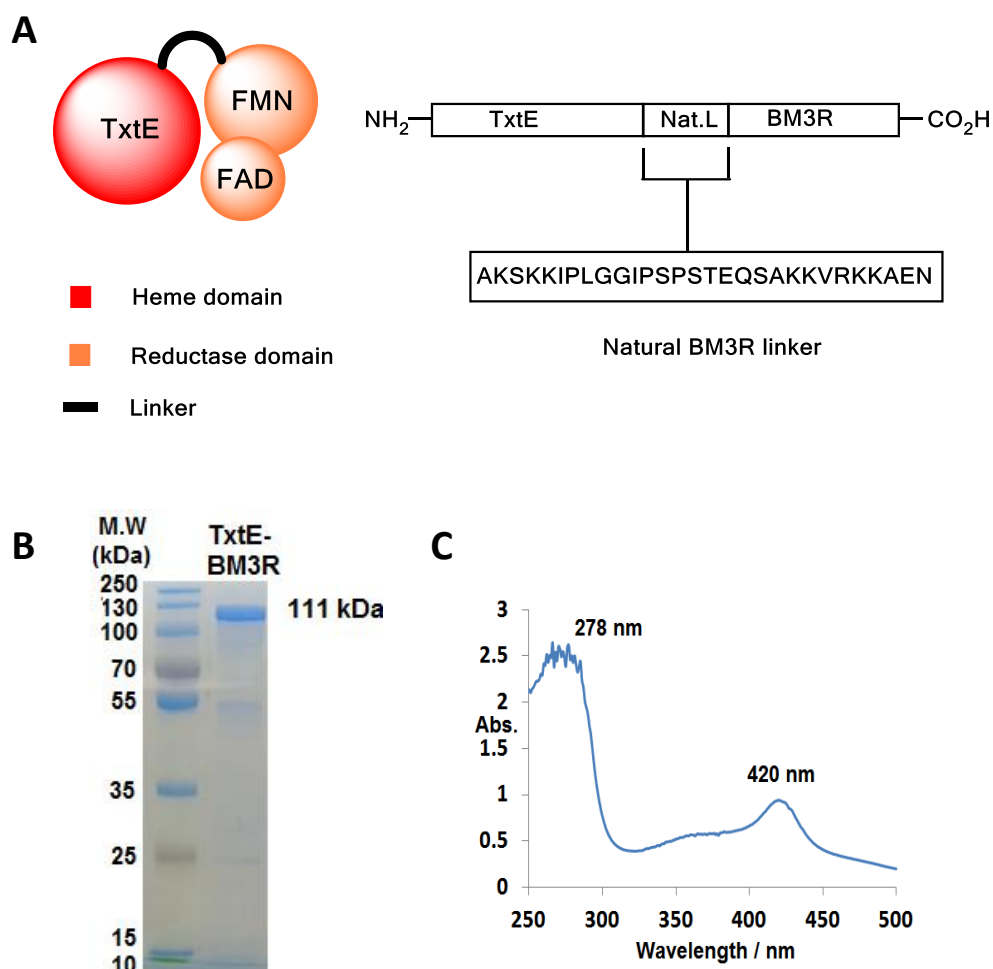


Figure 2.13. Analysis of TxtE-BM3R. A) Domain organisation of TxtE-BM3R and amino acid composition of the natural heme-reductase linking region B) SDS-PAGE analysis of purified TxtE-BM3R (~111 kDa, right lane). Molecular weight marker is in the left lane. C) UV-Vis spectroscopic analysis of TxtE-BM3R showing absorbance at 420 nm.

The larger 111 kDa TxtE-BM3R was purified as described previously, yielding relatively pure protein after Ni^{2+} NTA column chromatography (figure 2.13B). UV-Vis spectroscopy (figure 2.13C) confirmed the incorporation of the heme unit by the presence of the characteristic absorbance band at 420 nm. An L-Trp binding constant of $84.1 \pm 2.3 \mu\text{M}$ was obtained when the substrate was titrated into a solution of the protein (figure 2.14), and is comparable to that of both TxtE and TxtE-RhFRED.

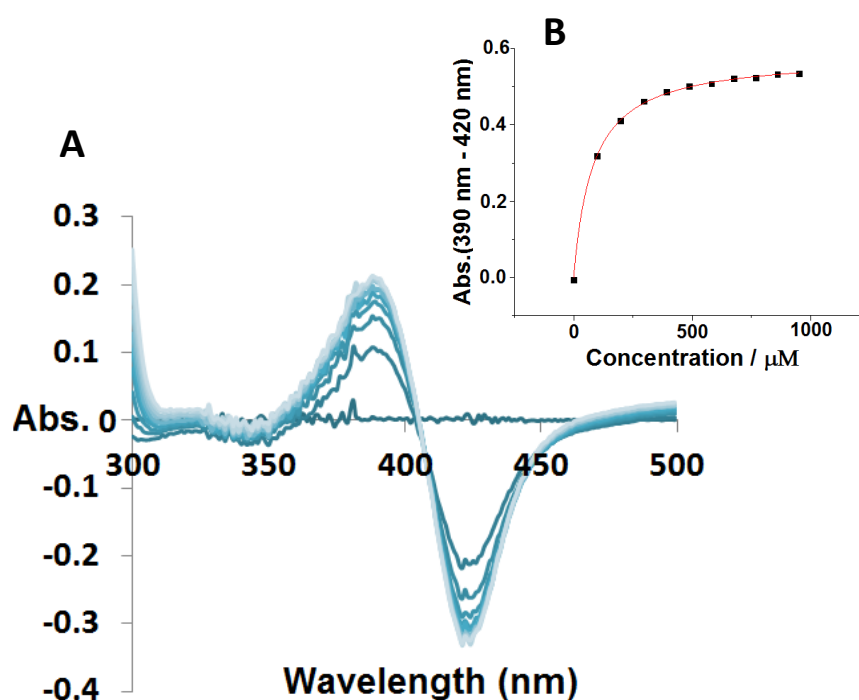


Figure 2.14. UV-Vis spectroscopic analysis of L-Trp binding to TxtE-BM3R. A) Type I difference spectra from titration of TxtE-BM3R (8 μ M) with L-Trp (0-1 mM). B) Difference in absorbance at λ_{\max} (390 nm) and λ_{\min} (420 nm) versus L-Trp concentration. Data fitted to a Michaelis-Menten model and are the mean of three measurements.

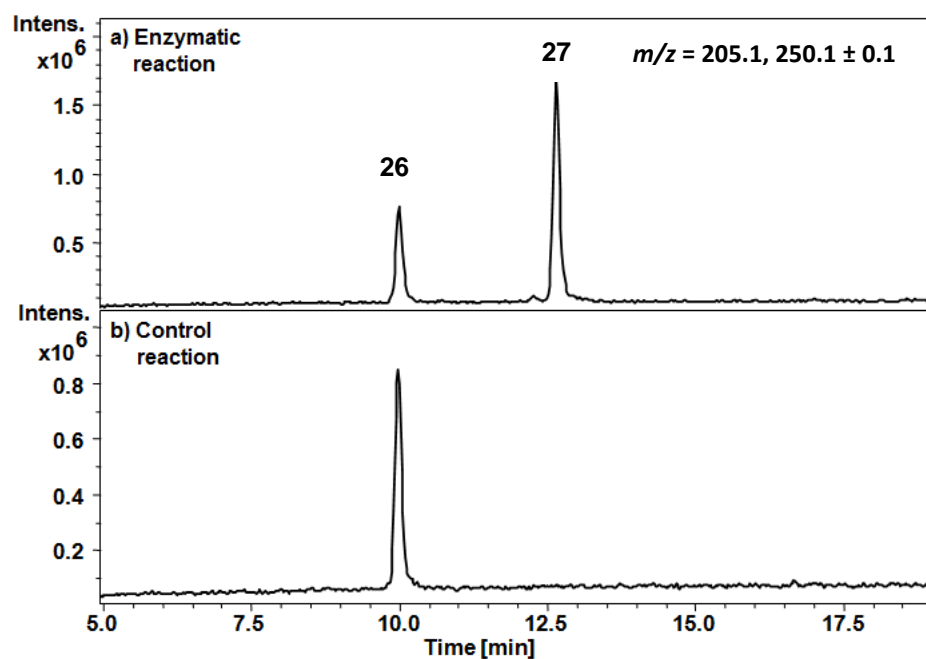


Figure 2.15. Analysis of the TxtE-BM3R-catalysed L-Trp nitration reaction (top) and control (bottom). LC-MS EICs at $m/z = 205.1$ and 250.1 ± 0.1 corresponding to $[M+H]^+$ for 26 and 27, respectively.

Once binding of L-Trp to TxtE-BM3R had been confirmed, an activity assay with the natural substrate was conducted as described previously (figure 2.15). Direct comparison of TTN values for the two chimeric proteins with L-Trp as a substrate revealed a substantial difference in the number of total reactions catalysed per enzyme molecule (table 2.2). Whereas TxtE-RhFRED could only catalyse 2 reactions prior to inactivation, TxtE-BM3R was capable of 450 reactions per enzyme molecule. On the other hand wild-type TxtE coupled with spinach Fd and Fr achieved a TTN value of 580. Importantly, a TxtE variant capable of being employed in large scale and high-throughput experiments had been obtained.

Enzyme	TTN
<i>wt</i> -TxtE	580
TxtE-BM3R	450
TxtE-RhFRED	2

Table 2.2. Total turnover numbers (TTN) of wild-type TxtE with redox partners compared to TxtE-reductase fusion proteins calculated from activity assays with L-Trp.

$$\text{Total turnover number (TTN)} = \frac{\text{no. mol product}}{\text{no. mol enzyme}}$$

Equation 2.2 Determination of total turnover numbers.

2.6 Flexibility in the regioselectivity of the TxtE-catalysed reaction

A substrate of particular interest is 4-methyltryptophan **76** as previous investigations have focused on determining the site of nitration when the 4-position, where nitration occurs on L-Trp, is occupied. Scale-up of the reaction by S. Barry (unpublished) to a total volume of 5 mL, followed by purification by HPLC and NMR analysis concluded that the product isolated was either the 5- (**102**) or 7-nitro (**103**) analogue of **76** as judged by the presence of two doublets and a singlet in the aromatic region of the NMR. To unambiguously identify the product of the reaction, NMR spectra must be compared to that of an authentic standard of the expected products. During the course of the current experiment, work by Zuo *et.al.* revealed that 4-fluoro-L-tryptophan **104** undergoes nitration at the 7-position of the indole ring to **105** (figure 2.16).¹²⁵ TxtE-BM3R provides a convenient method for the scale-up reaction with 4-methyl-DL-tryptophan **76** to directly compare the product NMR

against a synthesised standard whilst avoiding the use of large quantities of costly Fd and Fr redox partner proteins.

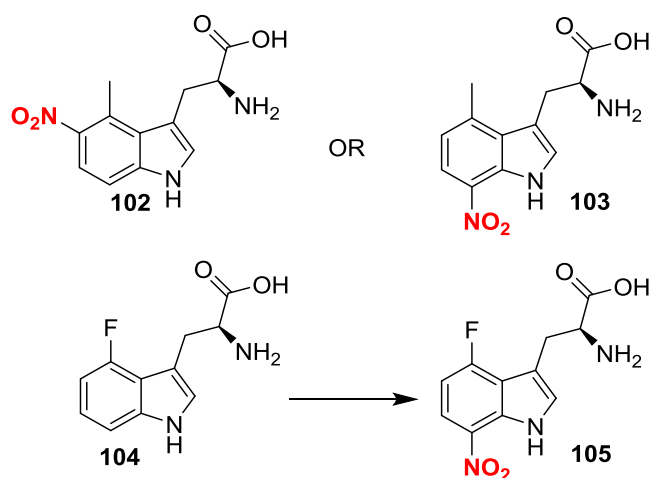


Figure 2.16. The TxtE-catalysed reaction with 4-methyl-DL-Trp indicates either 5- or 7-nitration (top) as indicated by ^1H NMR. Reaction with 4-F-Trp has previously been shown to result in 7-nitration (bottom).

To facilitate the large scale reaction without the need to purify the protein, which can be both lengthy and inconvenient, the decision was taken to scale-up the reaction using cell lysate. The preparation involves cell pellets containing TxtE-BM3R being incubated with lysozyme, polymixin B and benzonase nuclease for 2 hours and centrifuged to yield the clarified lysate. With the aim of maximising conversion to nitrated product, a buffer screen was conducted using this TxtE-BM3R clarified cell lysate and 4-methyl-DL-Trp as a substrate on a 600 μL scale. Tris-HCl and potassium phosphate buffers at different pH values were selected for the screen and this opportunity was also taken to determine the optimum conditions for a NADPH regeneration system, consisting of glucose dehydrogenase (GDH), NADP^+ and glucose.¹²⁶ Glucose and NADP^+ are substrates of GDH which forms gluconic acid whilst reducing NADP^+ . The NADPH generated can then be utilised by TxtE-BM3R for the nitration reaction, forming another molecule of NADP^+ in the process. As long as glucose is added to the assay in excess, the reaction can theoretically repeat until completion of the CYP reaction.

For comparison of total conversion to product, assays using 1 eq. NADPH were also conducted in each buffer system. The experiment was conducted in a 96-well plate to mimic later experiments which would involve the high-throughput screening of enzyme variants.

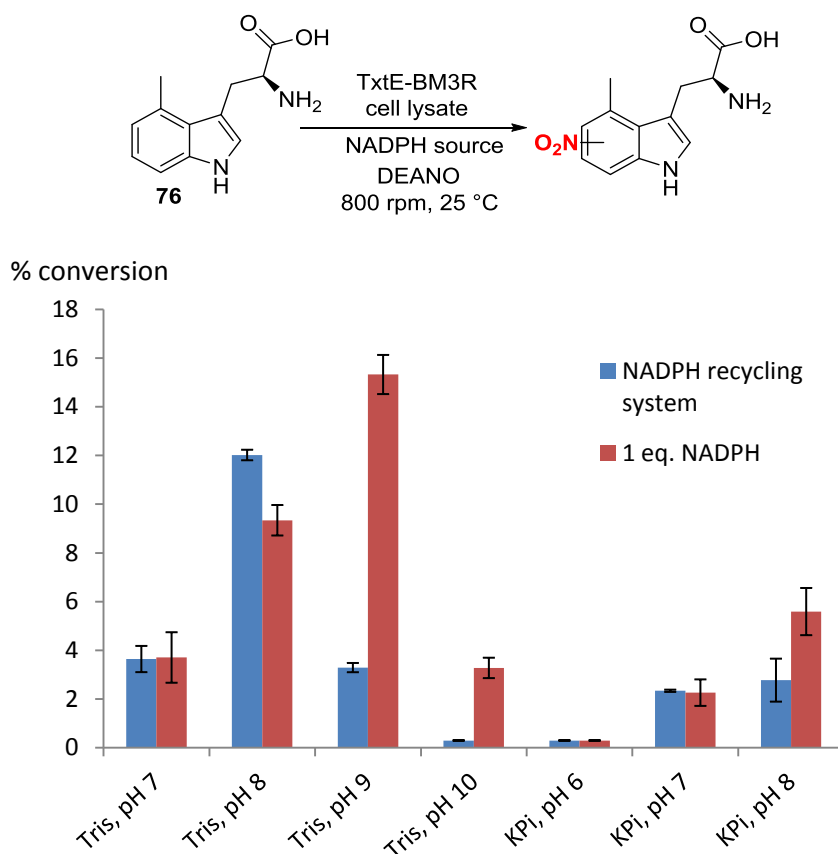


Figure 2.17. Screening TxtE-BM3R activity in various reaction buffers and utilising either NADPH or a regenerating system. Percentage conversion determined from LC-MS analysis monitoring peak area for 4-methyl-L-Trp and nitro-4-methyl-L-Trp.

Figure 2.17 shows the % conversion to nitrated 4-methyl-DL-Trp based on relative peak heights of starting material and product which was identified by a peak corresponding to m/z 264 in LCMS analysis. The graph shows a clear preference for Tris-HCl buffers over potassium phosphate, regardless of pH. Of the three Tris-HCl buffers tested, the assay utilising NADPH worked best at pH 9 (15 %), with activity diminishing rapidly at a higher pH. However, the assays using the cost effective NADPH regeneration system produced a comparable conversion of 12 % at an optimum pH of 8. For this reason, this buffer was chosen for the 4-methyl-DL-Trp scale-up reaction and all future experiments thereafter.

2.6.1 Scale-up of 4-methyl-DL-Trp nitration reaction

With an efficient nitration biocatalyst in hand and optimised conditions verified, efforts were focused on the 4-methyl-DL-Trp reaction. *E. coli* cell paste harbouring the TxtE-BM3R protein was resuspended in 120 mL of 100 mM Tris-HCl (pH 8) buffer. Lysis by microfluidisation and centrifugation generated the clarified TxtE-BM3R-containing lysate. The reaction was performed on a 100 mL scale with a substrate concentration of 5 mM (110 mg starting material) and carried out in a 100 mL EasyMax Chemical Reactor (Mettler Toledo) for efficient temperature control. After 20 hours the reaction was

quenched and worked-up to remove proteins prior to analysis by LCMS (figure 2.18) and purification via an Auto-preparatory system (Waters) using an m/z of 264. Combination of the pure fractions produced 8 mg of nitrated tryptophan analogue which displayed a similar NMR spectrum to that collected by S. Barry. The spectra each contain two doublets and a singlet at similar ppm values which is consistent with either 5- or 7-nitration. The yield for the reaction was 8 %, which although seemingly low, is the result of the first attempt of an enzymatic nitration reaction at this scale. This value could be improved in future scale-up experiments by testing several variables such as rate of NO addition, type of NO donor and reaction temperature, which were not considered during this study as a result of time constraints.

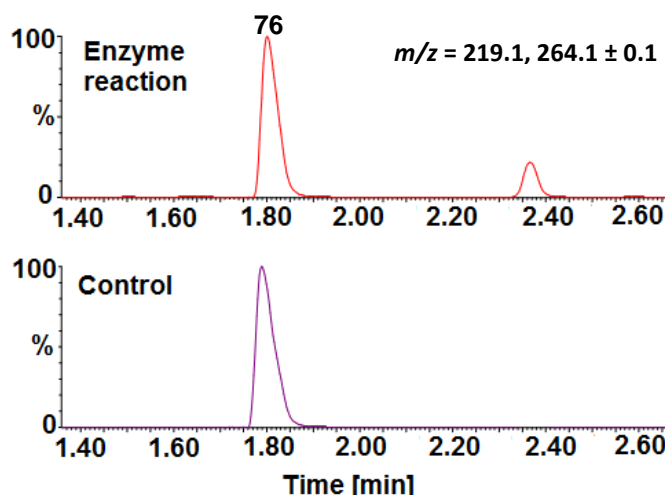


Figure 2.18. Analysis of the scaled-up TxtE-BM3R-catalysed 4-methyl-L-Trp **76** nitration reaction (top) and control (bottom). LC-MS EICs at $m/z = 219.1$ and 264.1 corresponding to $[M+H]^+$ for **76** and nitrated **76**, respectively.

2.6.2 Comparison to a synthetic standard of 4-methyl-5-nitrotryptophan

To confirm which of the nitro-4-methyl-DL-Trp regioisomers (**102** or **103**) is produced by the enzymatic reaction, a standard of **102** was synthesised, enabling direct comparison of NMR spectra. Previous literature precedent for the formation of enantiopure tryptophan analogues has involved formation of an *N*-acetyl amino acid intermediate from the substituted indole, followed by selective deprotection of the acetyl group via Amano acylase,¹²⁷ giving the pure L-Trp analogue (figure 2.19). The 4-methyl-5-nitro-1H-indole **106** starting material is readily available and a similar procedure was followed generating pure **102** in 2 steps in an overall yield of 14 %.

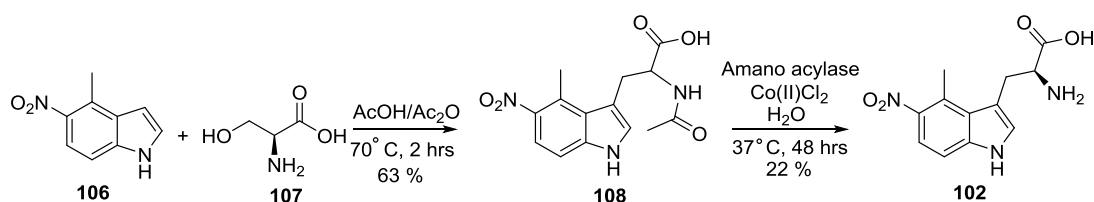


Figure 2.19. Synthesis of 4-methyl-5-nitro-L-Trp 109 from the corresponding indole 106 via an *N*-acetyl intermediate 108

Direct comparison of the aromatic region of the ^1H NMR spectra of the nitro-4-methyl-L-tryptophan enzymatic product and synthetic 4-methyl-5-nitro-L-tryptophan **102**, is shown in figure 2.20. Both compounds exhibit the same splitting pattern of 2 doublets and a singlet, with the same chemical shifts indicating these are identical. However it has previously been reported that a difference in concentration or pH can have a significant effect on the ^1H NMR chemical shifts of nitro-tryptophans, with spectra obtained in base generally exhibiting chemical shifts upfield of those in acidic conditions.¹²⁸ Although the spectra were recorded on the same instrument under the same conditions, to unequivocally confirm that the two products are identical, an equal volume of each NMR sample was mixed together and a new NMR spectrum was recorded. The spectrum of the mixture showed a single set of signals confirming that the two compounds are the same entity. It should be noted that during these studies a paper was published by Zuo *et al.* describing the isolation of **102** and **103** from reaction of 76 with a TxtE-BM3R variant in a ratio of 10:1.¹⁵⁴ However, the isolation of 0.8 mg of **102** was 10-smaller than isolated in these studies.

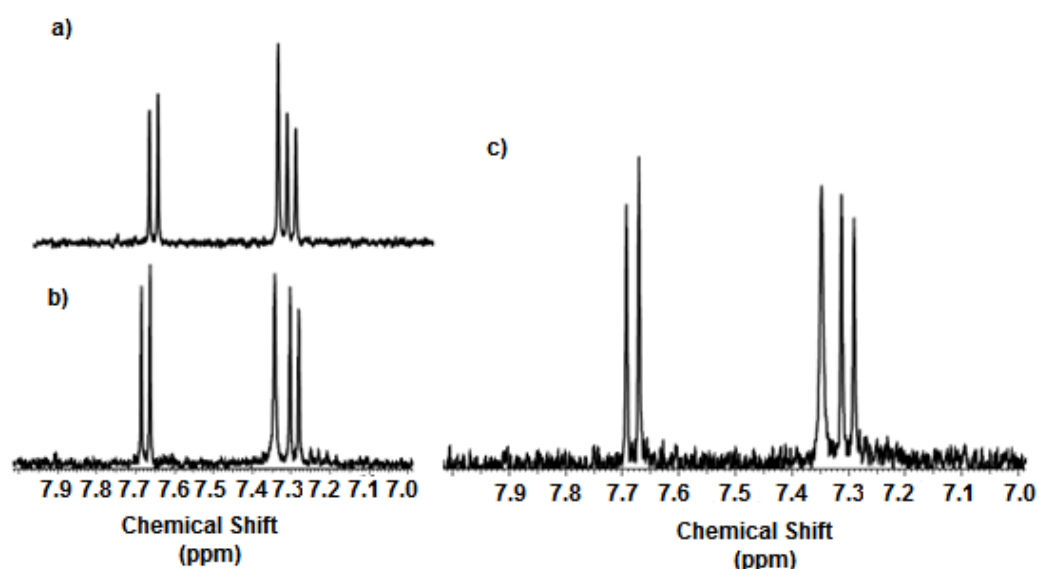


Figure 2.20. ^1H NMR spectra of synthetic and enzymatic 4-methyl-5-nitro-L-tryptophan. A) Aromatic region of the ^1H NMR spectrum of the isolated product from the enzymatic reaction B) Aromatic region of the ^1H NMR spectrum of synthetic 4-methyl-5-nitro-L-tryptophan. C) Aromatic region of the ^1H NMR spectrum of a mixture of the enzymatic and synthetic products.

2.7 TxtE-BM3R substrate screen

The substrate screen of TxtE-BM3R was conducted in 96-well plates using clarified cell lysate using the previously optimised conditions. A total of 17 indole substrates were tested consisting of both commercially available compounds as well as compounds synthesised in-house in GlaxoSmithKline laboratories in the UK and USA. Figure 2.21 illustrates the substrates investigated.

LCMS analysis revealed that only substrates **74**, **79** and **109** underwent nitration by TxtE-BM3R. Nitro-2-methyltryptophan and the nitro-5- and nitro-6-fluoro-substituted tryptophans were readily detected as single regioisomers (figure 2.21). Each of these substrates has the 4-position 'free' for reaction and it is reasonable to assume that the nitro-group is introduced at this position in each product. On the otherhand, 5-substituted L-Trp analogues **71**, **75**, **84** and **110** did not undergo nitration despite also having the 4-position available. The substituent at the 5-position has a considerable effect on the electronic properties of the indole ring, and hence reactivity. The hydroxyl- and methoxy- groups in substrates **84** and **75**, respectively, are both electron donating, increasing the electron density within the indole ring, and thus would be expected to increase reactivity towards electrophiles. However, the presence of these polar functional groups could provide undesirable interaction with the heme, the catalytic intermediates or the protein residues.¹⁰² The nitro group at position 5 in **71** on the other hand is extremely electron withdrawing and most likely prevents productive chemistry by making the indole ring too electron deficient. Substrates **75**, **82** and **83** have previously been determined to undergo nitration,¹⁰² however we were unable to detect nitro product for these under our conditions using cell lysate.

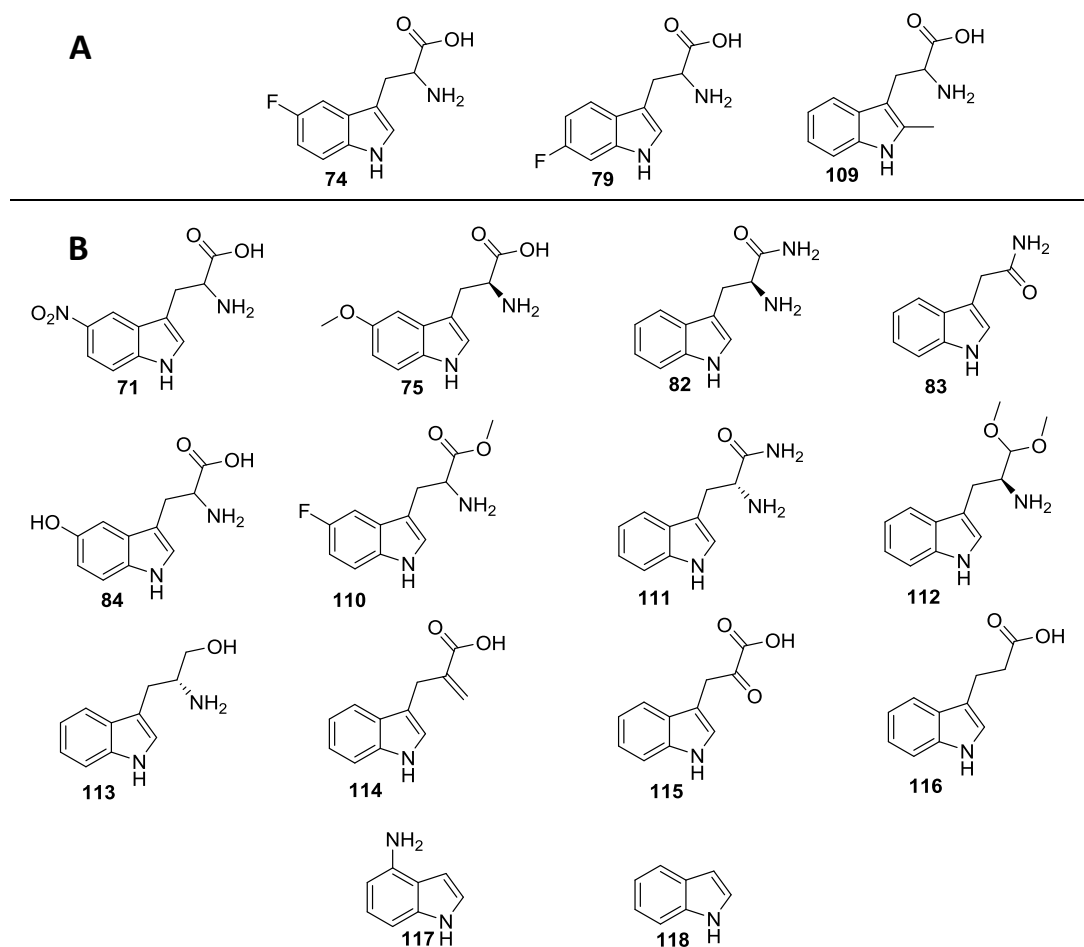


Figure 2.21. Structures of the substrates tested for TxtE-BM3R catalysed nitration. A) Substrates accepted by TxtE-BM3R as determined by LC-MS analysis. B) Substrates for which no detectable levels of nitro-product were observed.

As already suggested by others, any substrate in which the amino acid side chain is altered significantly (**112-118**) are not accepted by TxtE. The side chain has been deemed crucial for substrate recognition and docking. The key interactions provided by the amine and the carboxylic acid are responsible for anchoring the substrate in the active site and positioning the indole ring of L-Trp over the heme ready for reaction. Substrates **112-118** do not fulfil these criteria, hence no product is observed.

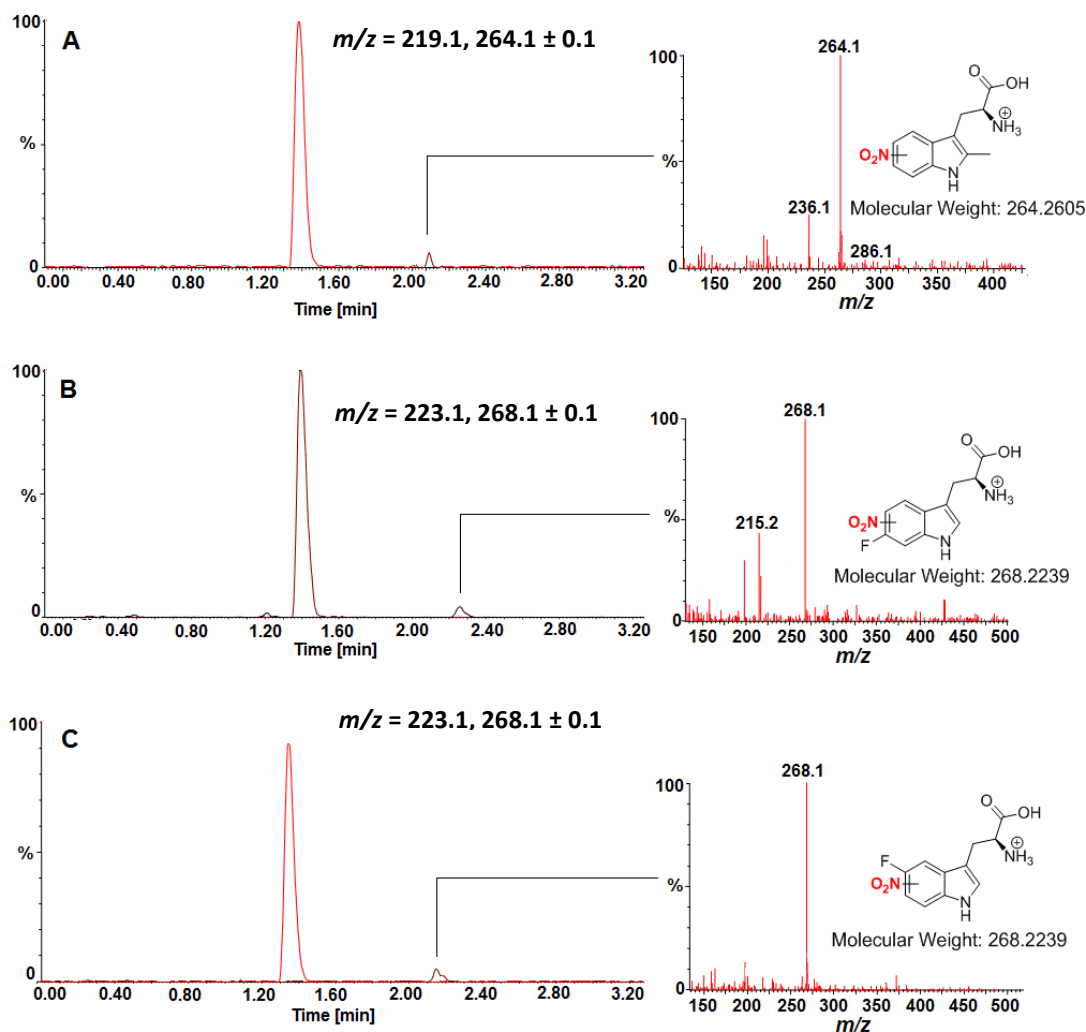


Figure 2.22. Analysis of the TxtE-BM3R-catalysed nitration reaction for positive substrates. A) Left: LC-MS EICs at $m/z = 219.1$ and 264.1 ± 0.1 corresponding to $[M+H]^+$ for 109 and nitrated 109, respectively. Right: Mass spectrum of nitrated 109. B) Left: LC-MS EICs at $m/z = 223.1$ and 268.1 ± 0.1 corresponding to $[M+H]^+$ for 79 and nitrated 79, respectively. Right: Mass spectrum of nitrated 79. C) Left: LC-MS EICs at $m/z = 223.1$ and 268.1 ± 0.1 corresponding to $[M+H]^+$ for 74 and nitrated 74, respectively. Right: Mass spectrum of nitrated 74.

2.8 Conclusions

The reported substrate scope of TxtE by Dodani *et al.*¹⁰² was further investigated via the testing of Tyr analogues. Following incubation with boiled TxtE, both **100** and **101** underwent nitration independently of active enzyme and even in its absence. This is most likely a combination of the formation of the nitrating species peroxynitrite in solution and the increased activation of the aromatic rings in these substrates compared to **26**.

During the investigation of TxtE-reductase fusion proteins, work by Zuo *et al.* described the creation of a self-sufficient TxtE variant fused to the reductase domain of P450_{BM3}.¹²⁵ Encouraged by these reports, we similarly created a TxtE-BM3R one-component protein, which was deemed superior to the TxtE-RhFRED variant, which would aid scale-up

reactions and high-throughput screening efforts. Following optimisation of reaction conditions and NADPH regeneration systems, TxtE-BM3R was used in the scale-up of 4-methyl-L-Trp **76** nitration. Isolation of 8 mg of nitro-product confirmed the product is the 5-nitro regioisomer **102** by comparison to a synthetic standard. A high-throughput substrate screen of the TxtE-BM3R protein revealed indole ring substitution is tolerated as nitro-product was observed for 5-F, 6-F and 2-Me L-Trp analogues. However, substrates in which the amino acid had major modifications did not react, supporting work by others.

Chapter 3: Creating libraries of TxtE variants and identification of mutant with increased L-Trp nitration activity

3. Introduction

Significant effort has been deployed in engineering CYPs from biosynthetic pathways for the oxidation of simple aliphatic molecules as well as large natural products because of the chemo-, regio- and stereoselectivity often displayed by these enzymes.¹⁰⁹ However achieving a comparable level of activity or selectivity with non-natural substrates remains troublesome.¹¹⁰ Nevertheless, since the reagents used for traditional aromatic nitration make this particular reaction notoriously difficult, the possibility of engineering an enzyme such as TxtE is an exciting prospect.

A common CYP engineering technique includes the random substitution of amino acids across the entire heme domain or substrate binding site in a technique known as directed evolution.¹²⁸ Several groups have used this Darwinian approach to successfully engineer CYPs with enhanced reactivity or altered selectivity towards unnatural substrates.¹¹⁴ Whilst these experiments have returned compelling results, the techniques often require a substantial screening effort, particularly when several rounds of evolution are performed.¹²⁹

Others have demonstrated the synergistic effect of simultaneous amino acid substitutions at multiple targeted positions in a CYP.¹³⁰ A prerequisite of this method is that the crystal structure of the CYP must be known in order to identify active site residues. These mutations can often be effective as single mutants, but more recently evidence has shown that some of these are only of use in combination.¹³¹ As the crystal structure of TxtE has previously been solved,¹⁰² and to reduce the overall screening effort required, a targeted approach to engineering TxtE was taken. From the outset the goals were: a) to alter the regioselectivity of the TxtE-catalysed nitration reaction with the natural substrate L-Trp from the 4-position to elsewhere on the indole ring, b) to expand the substrate scope and c) to improve conversion.

3.1 Analysis of TxtE structure

3.1.1 Active site

The TxtE crystal structure has been solved to 1.18 and 1.22 Å resolutions, with and without L-Trp bound, respectively.¹⁰² The substrate-bound active site structure is shown in figure 3.1 and reveals multiple residues within a 5 Å shell of L-Trp that can potentially affect TxtE regioselectivity and/or substrate preference. Figure 3.1A shows the water-mediated interactions and reveals that the amine also coordinates to E394. T296 and N293 also form several interactions via water molecules, leading to a complex hydrogen bonding network which could be easily affected by single mutations within this region. The indole moiety of L-Trp is in particular surrounded by several residues in close proximity (figure

3.1B). Of note are I244, A248 and T250, three hydrophobic amino acids that reside near the 5-, 6- and 7- positions of L-Trp. In contrast, position 4 is relatively unhindered and points directly towards the large open substrate access channel. Of the residues that directly interact with L-Trp via hydrogen bonding interactions, R59, Y89, N293 and T296 are of particular interest for mutation as they either bind the carboxylic acid, α -amine or both simultaneously. R59 has been shown to be a flexible residue, orientating itself to bind the heme propionate groups as well as the substrate carboxylic acid of L-Trp once it enters the active site.

Consequently we elected to mutate 15 residues in the TxtE active site speculating that amino acids within close proximity are likely to have the greatest effect on TxtE-catalysed nitration.

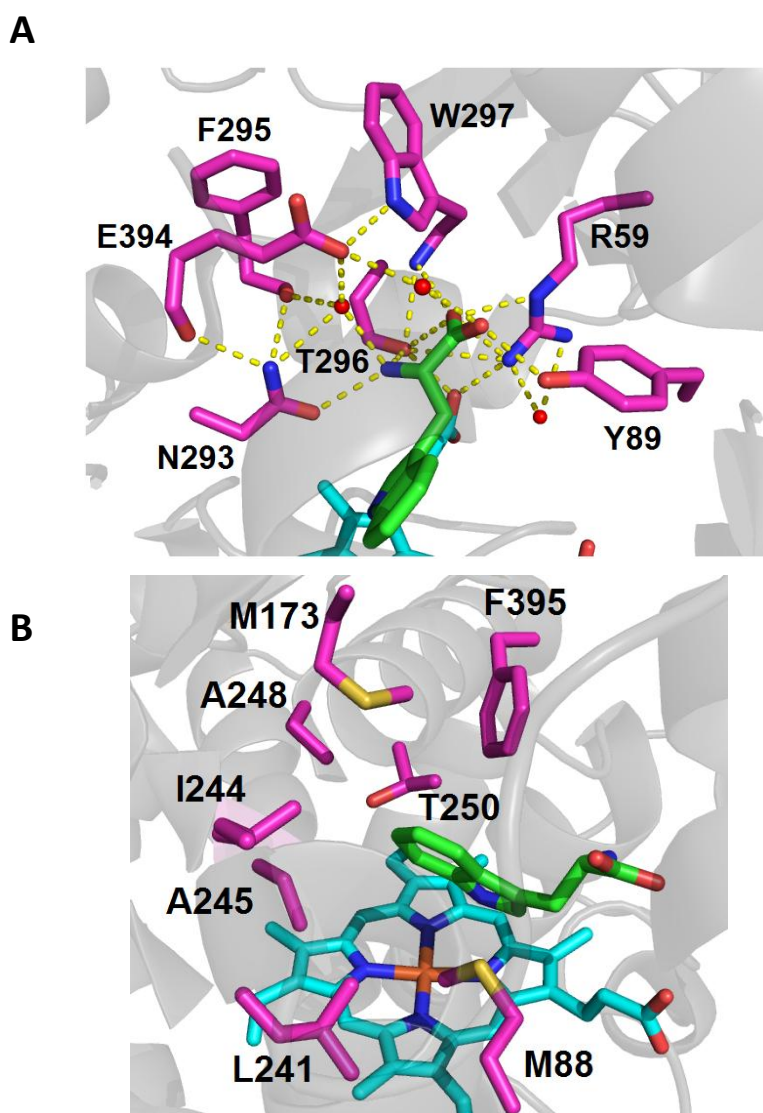


Figure 3.1. Targeted TxtE heme domain active site residues (PDB: 4TPO). A) L-Trp (green) shown with binding residues in magenta. Polar contacts are shown by yellow dashed lines and connecting waters shown by red spheres. B) Residues within a 5 Å shell of the L-Trp indole moiety. Heme is shown in cyan.

3.1.2 Flexible loop

The importance of the flexible F-G loop has previously been discussed for its role on TxtE-catalysed L-Trp nitration (Residues Tyr175 to Asp181).¹⁰⁷ Mutation of H176 to Tyr, Phe or Trp results in the formation of 5-nitro-L-tryptophan **71**. As a result of this minimal point mutation having a profound effect on reactivity, we sought to mutate 7 residues located within the flexible F-G loop (figure 3.2) hypothesizing that residues neighbouring H176 could potentially orientate it differently within TxtE active site, and thus interact with L-Trp in a distinctive manner.

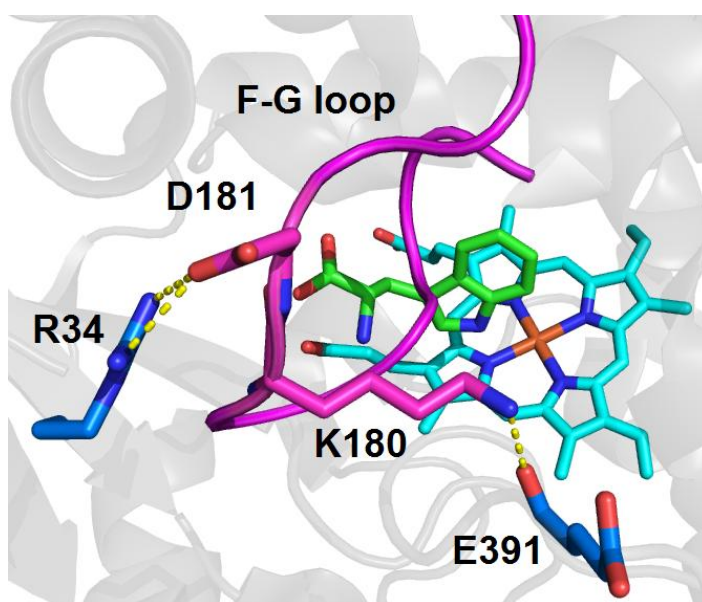


Figure 3.2. Overview of the flexible F-G-loop residues K180 and D181 (magenta) and interacting residues, R34 and E391 (blue) from the H176F mutant crystal structure (PDB: 5D3U). Polar contacts are shown by yellow dashed lines and L-Trp and heme are shown in green and cyan, respectively.

3.1.3 Helix and loop influencing residues

Several residues were identified that were not directly within the active site or F-G loop but are predicted to influence these regions.* R34 and E391 lie within the vicinity of the F-G loop and form hydrogen bonding interactions with K180 and D181 within the loop (figure 3.2). Modification of these residues could therefore alter the way this loop is positioned, thus indirectly influencing substrate binding. Similarly, T56, V63 and K84 are located close to R59, a key substrate binding residue situated at the centre of the protein structure as previously discussed (figure 3.3).

*Pamela J. Thomas (Computational and Modelling Sciences, GlaxoSmithKline Medicines Research Centre, UK) conducted structural analysis of TxtE and is responsible for identifying the helix and loop influencing residues R34, E391, K180 and D181.

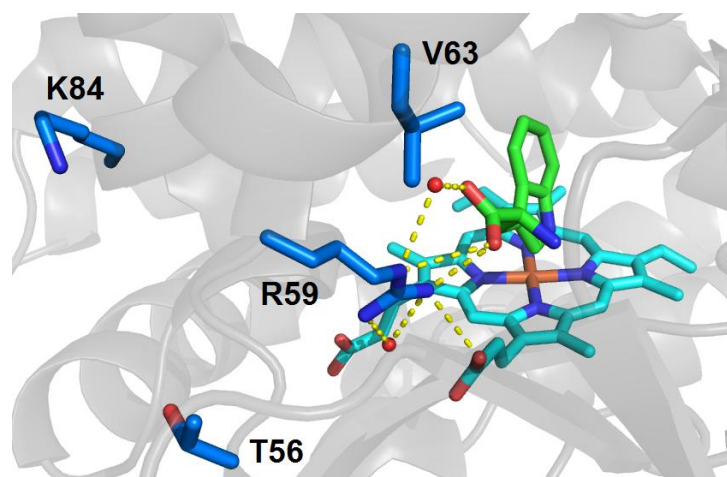


Figure 3.3. A) Overview of the R59 and influencing residues K84, V63 and T56 (blue) within the overall structure of TxtE (PDB: 4TPO). Polar contacts are shown by yellow dashed lines. L-Trp and heme are shown in green and cyan, respectively and the R59 coordinating waters are shown as a red spheres.

Table 3.1. Summary of residues targeted in the TxtE heme domain and the categories which they correspond to.

Category	Residue
Active site (5 Å shell)	M88, M173, I244, A245, A248, F295, W297, F395, L241, T250, E394,
Active site (Trp binding)	Y89, R59, N293, T296
F/G loop member	Y175, H176, S177, G178, P179, K180, D181
F/G loop interacting residues	R34, E391
R59 interacting residues	V63, T56, K84

3.2 Mutation techniques and methods

3.2.1 Site-saturation library preparation methods

Investigating the effect of modifying protein residues through the creation of a set of diverse protein variants is of vast interest for both a fundamental understanding of molecular interactions within the protein and between the protein and the substrate. Saturation mutagenesis is a powerful tool for interaction studies, protein engineering or directed evolution as a particular site can be modified to each and every other possible amino acid in a single step.¹³² However, one of the problems of saturation mutagenesis is

the vast quantity of variants produced. Saturation mutagenesis at a particular site is often achieved through the use of degenerate primers – a mix of oligonucleotide sequences differing only by the codon at the site of modification. Codons specified as NNN account for every possible nucleotide (A, T, C and G) at each of the three positions within it and this method is often favoured over the systematic mutagenesis of the site using primers specifically encoding each amino acid as it only requires one PCR experiment and costs the price of one set of primers. In addition, the mixture of mutated DNA can be sequenced to ensure satisfactory variation of DNA at the mutagenic site and protein variants can be screened in one step.

Systemised forms of protein engineering by site-saturation include Iterative Site Saturation mutagenesis (ISM)¹³⁰ and Combinatorial Active-site Saturation Test (CAST),¹³³ both of which are methods pioneered by Reetz *et al.* However, the large number of variants generated still requires an extensive screening effort which can be a major constraint on time and resources.

3.2.2 The 22c-trick method

The practical limitations associated with the aforementioned methods are largely a result of the redundancy of the genetic code, and are in fact not optimal at the genetic level. For example, methionine and tryptophan are encoded by only one codon each, ATG and TGG, respectively. In stark contrast, serine, leucine and arginine are each encoded by six codons each, thus the representation of some amino acids within a given library will be much higher than others. However, this degree of disproportionation is only true when saturation libraries are created using NNN degenerate primers. Attempts to reduce this bias often involve using NNK/NNS (K = T/G and S = G/C) to encode all 20 amino acids, however a certain level of redundancy still remains within the 32 codons (three codons for Arg, Leu, Ser and two codons for Ala, Gly, Pro, Thr and Val).¹³³

To shift closer to the ideal 20:20 ratio of codons to encoded amino acids, a more rigorous evaluation of the genetic code must be made. Reetz *et al.* had previously investigated protein function using NDT degeneracy (D = A/G/T) encoding 12 different amino acids by 12 codons.¹³⁴ The amino acids encoded within this set are of a mixed nature (Arg, His, Asp = charged, Asn, Ser = polar uncharged and Ile, Tyr, Phe = hydrophobic) but notable omissions include Lys, Cys and Trp. In studies building upon this, a combination of NDT, with VHG (9 amino acids, V = A/C/G, H = A/C/T) and TGG (1 amino acid, Trp), ensures that all 20 amino acids are represented within a total of 22 codons (figure 3.4). The method labelled the ‘22c-trick’ only has two redundant codons in which Leu and Val are encoded twice each when NDT, VHG and TGG are mixed in a ratio of 12:9:1 in the PCR

mixture.¹³⁵ For these reasons the 22c-trick was employed in the current study as a means to obtain a set of TxtE-BM3R variants within our focussed libraries.

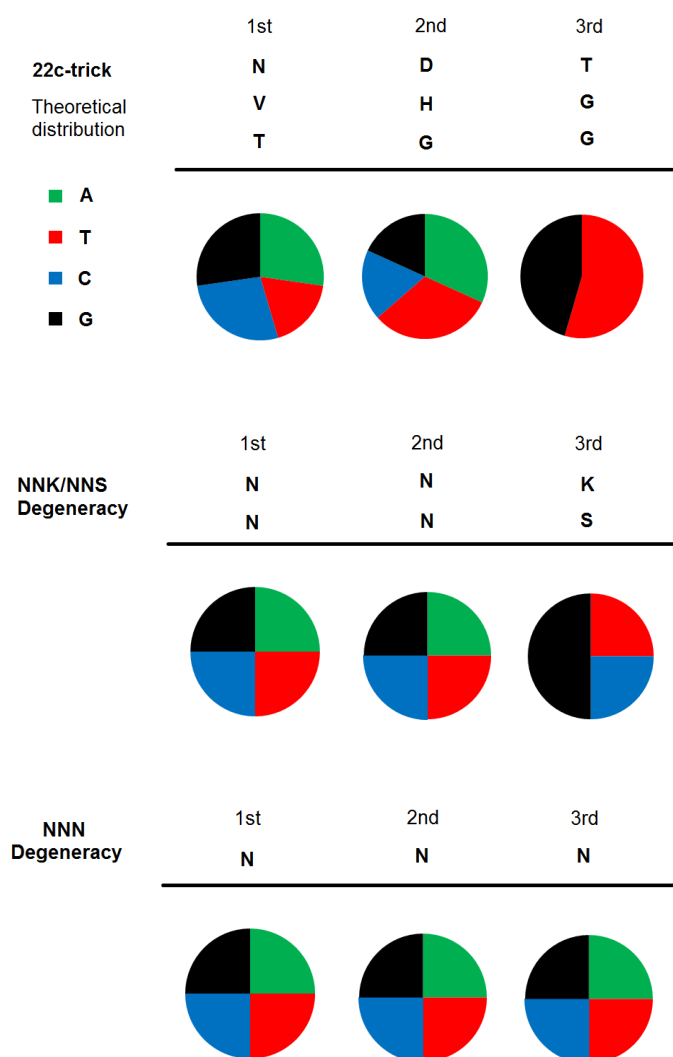


Figure 3.4. Theoretical distribution of nucleotide bases when using the 22c-trick (top), NNK/S degenerate primers (middle) or NNN degenerate primers (bottom). Expected percentage distribution of each nucleotide is represented as pie charts for the three methods. Symbol: N = A, G, C, T; D = A, G, T; V = A, C, G; K = G, T; S = C, G. Adapted from ref 135.

3.2.3 Reducing the library screening effort

The library screening effort refers to the total number of samples required to be analysed by the screening methods to assess the effect of mutagenesis on the targeted protein space.

The screening effort can be defined in one of two ways:¹³⁶

Library coverage refers to the proportion of the variant sequences sampled if a given number of colonies are selected.

Full coverage probability on the other hand, relates to the number of colonies that must be picked in order to sample the complete variant sequences.

It is important to appreciate the differences between these two definitions as this can have an immense effect on the number of colonies that must be picked. To screen the entire protein space using the *full coverage probability* definition, it has been established that it requires between 10-20 times oversampling, that is to cover 20 amino acids represented equally within a variant library, between 200 and 400 colonies must be screened. On the contrary, to achieve 95% *library coverage* of the mutant sequences, only around 3 times oversampling is required. The library size for the latter method can be calculated using the following equation:¹³⁷

Equation.3.1: Determining total library size

$$L = -V \ln (1 - F)$$

where L is the number of library samples required, $V = X^n$, where X is the total number of codons represented by the mutagenic DNA and n is the number of saturated sites, and F is the library completeness i.e. 0.95 for a 95 % library coverage. When using the 22c-trick this equates to 66 colonies [$L = -22^1 \ln(1 - 0.95)$] to cover 95 % of the variant space. Conversely, for NNK/S redundancy which encodes 20 amino acids via 32 codons, this increases to 96 colonies. For comparison, to achieve a similar coverage with NNN redundancy, 192 colonies would need to be picked with a vast misrepresentation of certain amino acids existing within the library.

These statistical definitions do not take into account limiting factors such as incorporation of mutagenic DNA at the genetic level during PCR. The equation assumes all codons are included equally and has been deemed a sufficient measure of library size in previous studies.¹³⁷

3.2.4 Site-saturation mutagenesis techniques

The total number of oligonucleotides required depends on the technique utilized, the total number of saturation sites and the distance between these sites. As our libraries will only be modified one site at a time, the distance between the residues is not a factor in primer design. The much used QuikChange Site-Directed mutagenesis kit (Agilent, CA, USA) was the technique of choice in the 22c-trick study our investigations are based on.¹³⁵ However, this method requires both a mutagenic forward and reverse primer for the PCR experiment, and therefore 3 sense and 3 antisense primers per saturation library when

using the 22c-trick. For the planned 27 libraries, this would mean synthesising 162 oligonucleotides. In contrast, the NEB Q5 Site-Directed mutagenesis protocol (New England Biolabs, MA, USA) requires a mutagenic forward primer and a non-mutagenic reverse primer. In this case, a total of 4 oligonucleotides (3 sense and 1 antisense) are required per 22c-trick library and therefore only 108 for the full set of libraries. Reducing the total cost of primer synthesis by 33% is of even greater significance if multiple sites are saturated simultaneously. The 22c-trick in combination with the NEB Site-Directed mutagenesis kit was therefore chosen to create the TxtE-BM3R variants, striking a balance between an economical PCR procedure and an efficient screening process.

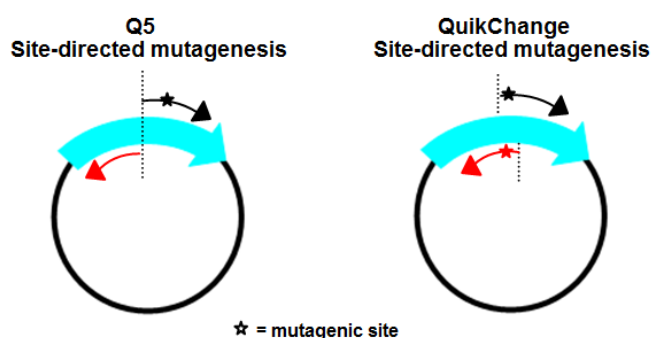


Figure 3.5. Overview of Q5 Site-directed mutagenesis and QuikChange mutagenesis primer design for substitution at one site within a gene.

3.2.5 Quick quality control

Assessing the quality of mutant libraries prior to screening is extremely important for determining whether genetic diversity has been introduced within the targeted region of the parent DNA. For a site saturation library to be deemed a success the distribution of nucleotides within the mutagenic codon should be as close to expected as possible. Not only is screening a library of no or minimal diversity a waste of resources, time and effort, it can also vastly misrepresent the data and create false outcomes. Therefore to overcome such issues, Reetz *et al.* have developed a Quick Quality Control (QQC) method which evaluates library variation prior to screening efforts.¹³⁴ Briefly, the technique requires the mutant DNA libraries, obtained after the PCR experiment, to be transformed and plated on appropriate solid media. The colonies are then scraped from the plate and the plasmid DNA is isolated, generating a mixture of the library DNA. This library is sequenced and assessed for genetic diversity by comparing the distribution of nucleotides at each position in the mutagenic codon. If it is verified that the parent sequence has been substituted and the distribution of nucleotides is similar to the expected sequences, the library is of high quality and it is suitable for screening to commence.

3.2.6 Library screening protocol

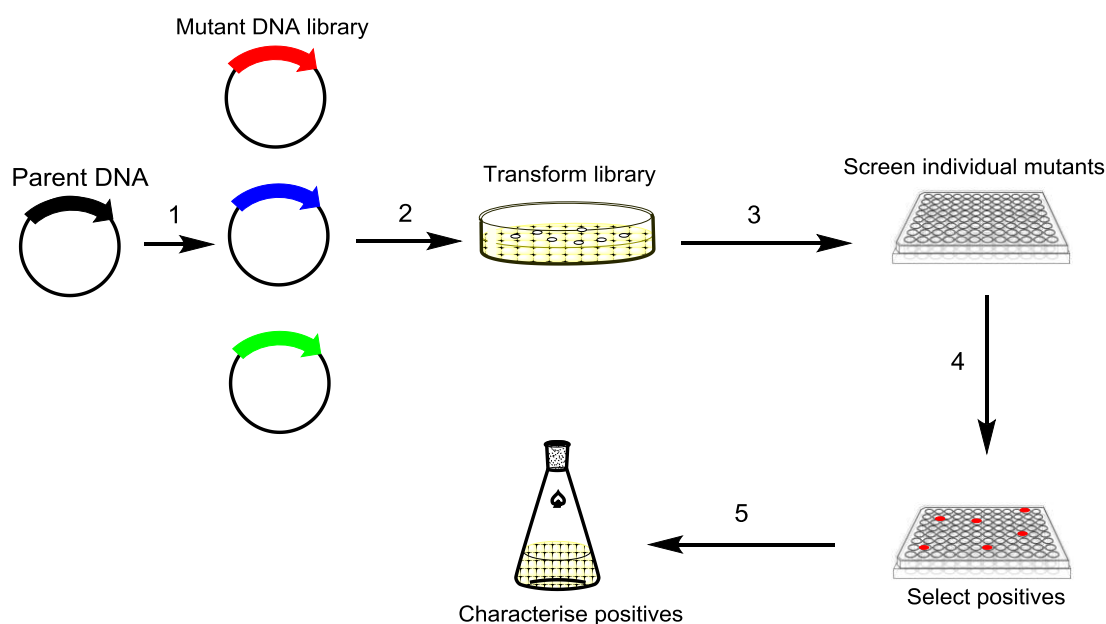


Figure 3.6. Schematic showing the overall library preparation and screening protocol for TxtE-BM3R mutant libraries. The screening reactions and LCMS analyses are conducted in 96-well plates following QQC of the site-saturation mutagenesis libraries.

The general experimental protocol for the preparation of mutant libraries and their subsequent screening is depicted in figure 3.6.

1. Library preparation: The parent DNA is modified at a targeted site by saturation mutagenesis creating a library of enzyme variants.
2. Quality Control: The generated mutant library is assessed for diversity at the targeted site by screening the collective DNA sample.
3. Expression in 96-well plates/reaction: If successful the mutant library is plated on solid agar plates and a single colony is used to inoculate individual wells of a 96-well plate. The enzyme variants are cultured and overproduced in these wells, followed by the screening reaction.
4. Analysis of the screening reaction: A sample of the reaction is analysed and assessed to determine whether any of the new mutants show improved activity over the parent enzyme.
5. Culturing of productive mutants: Mutants showing improved conversion are grown from a parent plate containing the glycerol stocks of the bacterial cell cultures. The plasmid DNA is extracted from these cultures and sequenced to identify the mutation.

3.3 Initial Validation Screen – H176 library

As a validation of our methods and techniques, the H176 library was chosen as the first set of variants to be prepared. Previous studies have shown mutation of this residue to Phe, Tyr or Trp resulted in a switch from 4-nitration to 5-nitration of L-Trp.¹⁰⁷ Creation and screening of this library would therefore show whether: a) the mutagenesis protocol was effective at producing a diverse range of DNA sequences, b) the screening protocol was sufficient for detecting nitro-L-Trp regioisomers and monitoring % turnover c) whether any of the substitutions at this position effect the regioselectivity of this reaction and d) whether the presence of the fused reductase domain impacts on the regioselectivity switch.

3.3.1 Verifying the analytical method

Comparison of the product peak (m/z 250) resulting from the reaction of L-Trp with TxtE-BM3R (4-nitro-L-Trp **27**) and a 5-nitro-L-Trp **71** product standard showed that there was a clear difference in the retention times of the two nitro-L-Trp regioisomers using the LCMS analytical method (figure 3.7). This was an important aspect of the experiment as it would allow the rapid automated detection of the different regioisomers by the analytical software, crucial for high-throughput screening of the mutants.

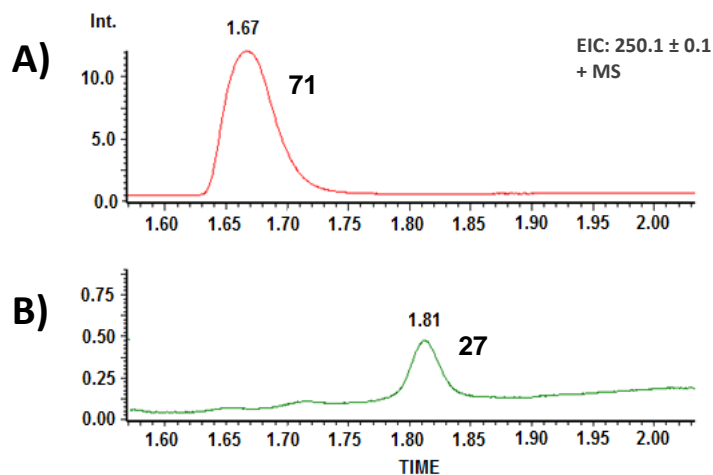


Figure 3.7. EICs at m/z = 250, corresponding to $[M+H]^+$ for 4- (**27**) or 5-nitro-L-tryptophan (**71**), from LC-MS analyses of A) an authentic standard of **71** and B) **27** from the TxtE-BM3R-catalysed nitration reaction

3.3.2 H176 library preparation

As the mutagenic primer contains a mixture of DNA sequences, the melting temperature of the primer, T_m , covers a large range of values. Adjusting the T_a of the mutagenesis PCR experiment is therefore vital to achieving an optimised distribution of bases in the library and to ultimately obtaining a diverse set of enzyme variants. Often several T_a 's must be used to achieve a balanced set which is easily accomplished by splitting the PCR mixture into several smaller portions, running the PCR experiment and then combining the resulting portions afterwards prior to QQC. A detailed explanation regarding the selection of T_a values is described in the methods section (7.2.5).

A QQC analysis of the purified H176 library DNA was performed and the electropherograms of the sequencing analysis is shown in figure 3.8. Evidently the three bases which were targeted by mutagenesis show a great deal of variation and the wild-type codon at this position, CAC, is no longer present. In particular, the first and second positions include all four nucleotides as expected. Position 3 also shows the expected base distribution of G and T, represented in black and red, respectively, but once again the relative distribution of these two bases differs, as judged by visual inspection, from the expected values of 54.5 % for T and 45.5 % for G. Importantly, the parent codon at this position has been eliminated completely as C is no longer present at the 3rd base and also indicates that *DpnI* digestion of the parent plasmid is complete. A His residue could however be reintroduced via the 22c-trick degenerate primers using a CAT codon.

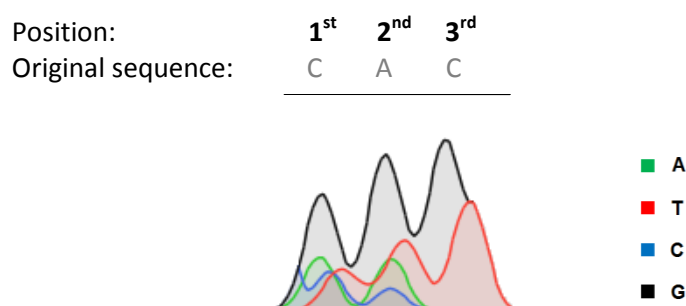


Figure 3.8. Sequencing electropherogram of H176 site-saturation library. The sequence of the wild-type codon corresponding to His is shown above the targeted codon.

3.3.3 His176 library screening

With a diverse H176 DNA library in hand, the screening of enzyme variants in 96-well plate format could proceed as described in section 7.2.4. Once LCMS analysis data had been obtained, the detection of **27** by the screening method was first verified (figure 3.9). Figure 3.10A shows the 96-well plate heat map monitoring the LCMS signal at $t = 1.81$ min, the retention time of **27**. The heat map confirmed the presence of **27** in the three TxtE-BM3R positive control wells (A12, B12 and C12) and the absence of **27** in the negative control wells containing either cells containing empty pET-28(a)+ vector (D12-F12) or buffer (G12-H12). Of the cells producing TxtE-BM3R variants (columns 1-11) only six wells contained variants capable of nitrating L-Trp at the 4-position. The majority of mutations (78/84) resulted in loss of activity, defined as less than 10 % of the conversion shown by the positive control reactions.

A similar heat map was then created to look for variants capable of catalysing nitration at the 5-position by monitoring the LCMS signal at $t = 1.67$. The heat map showed no conversion to **71** by wt-TxtE-BM3R, as expected, but **71** was detected in four of the wells (C3, C5, D2 and D10) (figure 3.10B). As no conversion to **27** was observed for these variants, the regioselectivity was completely switched to 5-nitration.

The corresponding *E. coli* glycerol stocks were cultured and their plasmids were purified. Subsequent plasmid sequencing revealed that two of the variants contained Trp at the 176 position, one contained Tyr and one contained Phe (table 3.2). This is in agreement with previous studies in which mutation of H176 to these three residues resulted in formation of **71** and shows that the presence of the fused reductase domain does not affect the outcome.

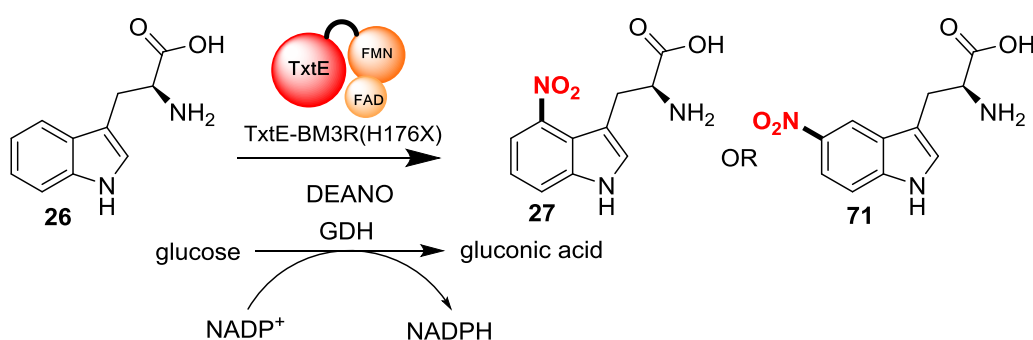


Figure 3.9. Library screening reaction for L-Trp nitration by TxtE-BM3R(H176) variants. Reaction monitored by LCMS for either 4- or 5-nitro-product using a NADPH recycling system in a 96-well plate and conversions assessed by UV-vis peak area.

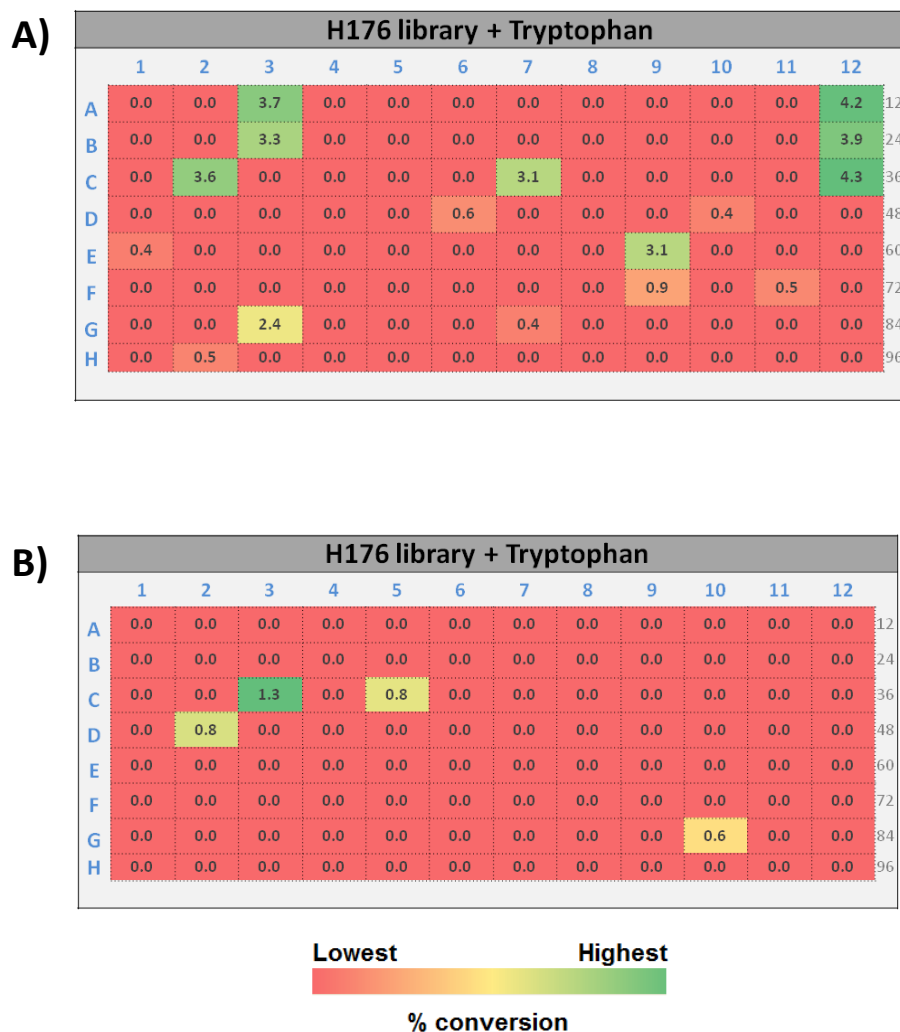


Figure 3.10. H176 mutant library screening plate heat maps. A) Heat map monitoring for product at 1.67 min corresponding to % conversion to 4-nitro-L-tryptophan. B) Heat map monitoring for product at 1.81 min corresponding to % conversion to 5-nitro-L-tryptophan. Conversions are represented as a scale of highest to lowest for each plate. Control reactions are as follows: A12-C12 are with TxtE-BM3R protein, D12-F12 are with pET28(a)+ proteins and G12 and H12 are with reaction buffer (100 mM Tris-HCl, pH 8).

Well	Residue at position 176
C3	Y
C5	W
D2	W
G10	F

Table 3.2. Summary of TxtE-BM3R mutants capable of nitrating L-Trp at the 5-position as determined by high-throughput screening of H176 library variants.

3.4 Analysis of site-saturation library quality

After the initial validation results, the remaining 26 libraries could be prepared with confidence that both the saturation mutagenesis technique is effective and the screening methods are sufficient. It was evident during the preparation of the full set of libraries that not all libraries had the same quality as the H176 set. We observed libraries of higher diversity (figure 3.11A) than that of H176 but also libraries of significantly lower diversity (figure 3.11B) as judged by QQC.

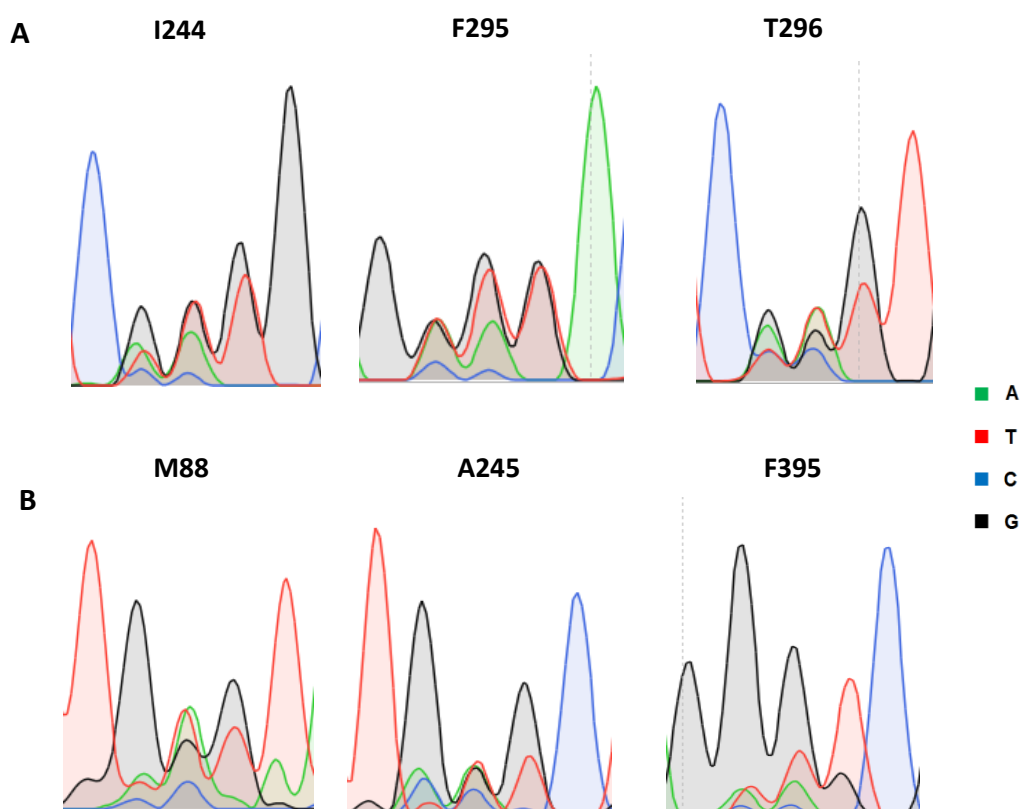


Figure 3.11. Selection of electropherograms corresponding to high quality site-saturation libraries (top row) and low quality libraries (bottom row) from the complete set prepared via the 22c-trick method in this study. The central three peaks correspond to the targeted codon in all cases and demonstrate the variation in library diversity obtained using this method.

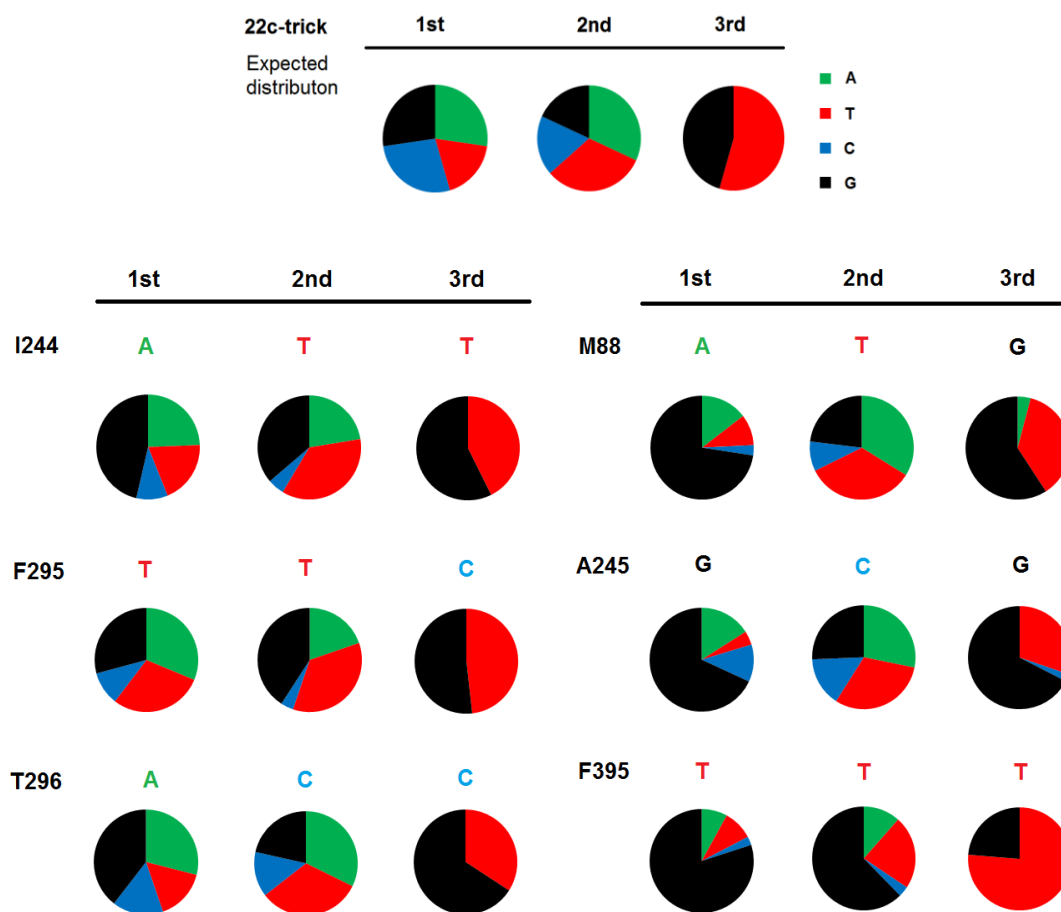


Figure 3.12. Obtained distribution of nucleotides for TxtE-BM3R mutant libraries represented as pie charts. The expected theoretical percentage distribution of nucleotide bases using the 22c-trick is shown at the top. Comparison with the distribution obtained for the high quality libraries (left) and low quality libraries (right) show the variation in library quality achieved using this method. Percentage base distribution of prepared libraries calculated from sequencing electropherograms. The wild type codon for the targeted site is shown above each pie chart.

It is apparent from inspection of the electropherograms corresponding to the lowest quality libraries (M88, A245 and F395) that there is a strong preference for guanine being introduced at the first base of the codon at up to 75 % of total nucleotides. This type of codon bias has been reported previously by several groups, and in particular the domination of codons bearing the GXG or GXT codon. In turn, this results in the libraries being dominated by variants bearing Gly, Asp, Val, Glu and Ala at the saturation site and all other amino acids being represented to a lesser degree. Also, the third nucleotide of the A245 and F395 libraries are largely biased towards the inclusion of G and A, respectively, which on closer inspection are in fact the wild-type codon. These effects have been deemed to be principally controlled by the wild-type DNA sequence and in some cases is

the result of incomplete *DpnI* digestion of the parent template. Furthermore, the F395 library has very little cytosine in the 2nd position and therefore overall this library was deemed the lowest quality. Despite numerous attempts to improve the quality of this library, these were not successful and so the F395 library was not selected for further investigation. The small amounts of unexpected A and C nucleotide in the third base of M88 and A245, respectively, is most likely a result of interference in the electrophoresis signal from sequencing of neighbouring bases.

The most balanced libraries were considered to be that of I244, F295 and T296. Direct comparison of the distribution pie charts (figure 3.12) of these libraries against the lower quality ones shows the distribution of bases is much more even, especially in the first position where a guanidine no longer dominates. At first glance there is very little C at the 2nd base of the I244 and F295 codons, however cytosine is only expected to form 18.2 % of bases at this position using the 22c-trick and therefore affects only 4 out of 22 codons. The T296 library has an even distribution of all 4 bases in position 1 and 2 of the codon which along with strong representation of G and T in position 3 makes it the highest quality library on the whole.

As observed here, site saturation libraries can vary hugely in their overall quality and this study reinforces the importance of quality control during library preparation. Optimisation may be necessary but in many of our cases the quality of the libraries could not be improved by either adjusting the T_a of the PCR experiment, adding DMSO to the PCR reaction or increasing the annealing time. Others have reported similar findings in their investigations, and in these cases it was suggested that the intrinsic properties of the primers, such as length, GC content and melting temperature, were some of the limiting factors affecting library quality.¹³⁵ In addition to this, the position and sequence of the targeted codon within the wild-type gene was also determined a key factor in library quality. This was also observed in our libraries as F295, T296 and W297 all produced high quality libraries in comparison to mutations located elsewhere in the gene. In some cases we could not produce library DNA at all. In addition to the F395 library, those involving A248, S177, E394 and T250 were withdrawn from our investigations as either the saturation mutagenesis experiment did not produce a PCR product or the libraries were considered too low a quality to be useful in screening experiments. We continued with the screening of the other 22 optimised libraries with L-Trp regardless of whether some codon bias was apparent as it has been widely accepted that it is not possible to acquire a perfect library because of the factors described above.

3.5 High-throughput screening of TxtE-BM3R mutants for improved activity with L-Trp **26**

From the plate heat maps of all mutant libraries, it is possible to compare the conversion of mutant assays against the wild-type control reactions (figure 3.13). Relative to the wild-type control assays, the M88, Y89, V63 and N293 mutant wells mostly showed a large decrease in activity compared to other libraries. This is perhaps not surprising for Y89 and N293, two key L-Trp binding residues located in the active site of the TxtE heme domain. For the Y89 library an average conversion of 14.9 % was calculated for the three control reactions, yet the best performing variant was only able to achieve 5.8 %. Similarly for N293, the best variant produced 4-NO₂-Trp **27** with only 13.4 % conversion compared to 25.9 % for TxtE-BM3R. Both libraries were determined to incorporate a decent range of amino acids at the targeted site by QQC therefore concluding that these two residues are extremely important for activity via the correct positioning of the substrate in the active site that cannot be achieved by any other amino acid. Conversely, the M88 library showed a single variant across the 84 mutant wells that converted L-Trp to the nitro product to a similar degree as the positive control reactions. At first glance it could be assumed that M88 is another residue crucial for the catalytic activity of TxtE-BM3R. However, inspection of the QQC for this library (figure 3.11) reveals imbalanced representation amino acids at the targeted site. In particular, Ser, Thr and Pro will rarely be incorporated as a result of minimal representation of CCX or ACX codons at the saturation site and therefore M88S/T/P may not appear at all within the library. Consequently the effect of M88 on TxtE activity would need to be investigated further, possibly by the three rare amino acids being substituted into the *txtE* gene in additional experiments.

Figure 3.13A shows the average conversion of all 84 mutants of each plate compared to the average conversion of the corresponding three TxtE-BM3R positive control reactions. Processing the data in this way ensures that the graphs provide a standardised analysis of all libraries as often the conversion obtained by the control reactions differed between plates. This variation in the control reaction values could be due to a number of factors including differences in AutoInduction media between batches, length of culture incubation, temperature differences within the laboratory and whether cell expression plates had to be defrosted or were used directly from overnight cultures. All saturation libraries showed a decrease in activity overall as it is likely that the number of productive mutations is far lower than those which are disruptive.

The degree to which each library is affected however is interesting to see. As discussed previously, saturation of M88 results in near abolishment of nitration activity, with mutation of Y89, V63 and N293 also resulting in a general large loss of activity. R59 shows an average low conversion across the plate compared to the control reactions and

within the same active site residues category, mutation at L241, A245, F295 and T296 similarly has a large effect on nitration. It could be expected that residues in the active site of L-Trp have the greatest effect on activity as they are most likely to affect substrate orientation or catalytic mechanism. These four residues lie close to the L-Trp substrate and could therefore either have a direct effect on substrate binding or an indirect effect by altering the overall shape of the active site. Of the remaining active site residues, mutation of M173, I244 and W297 did not affect activity as much as the previously discussed sites most likely as they do not lie particularly close to the substrate and so do not interact greatly with it.

In contrast, the three mutated sites which caused the least disruption to enzyme activity on the whole were D181, R34 and E391. These three residues lie outside of the active site, with D181 forming a part of the flexible F-G loop that is an important factor in determining nitration regioselectivity (section 1.6.2.1). Intriguingly, both R34 and E391 interact with the F-G loop by forming important hydrogen-bonding contacts with residues in this region of TxtE (figure 3.2). Figure 3.13B shows the best performing mutant in each library compared to the positive control reaction and three of the most active mutants across all libraries are modified at D181, R34 and E391, all proving to be at least 2-fold more active than the parent enzyme. As a comparison the best mutants in the Y89 and N293 libraries did not achieve even 50 % of the conversion of the parent underpinning the importance of these wild-type residues for L-Trp nitration.

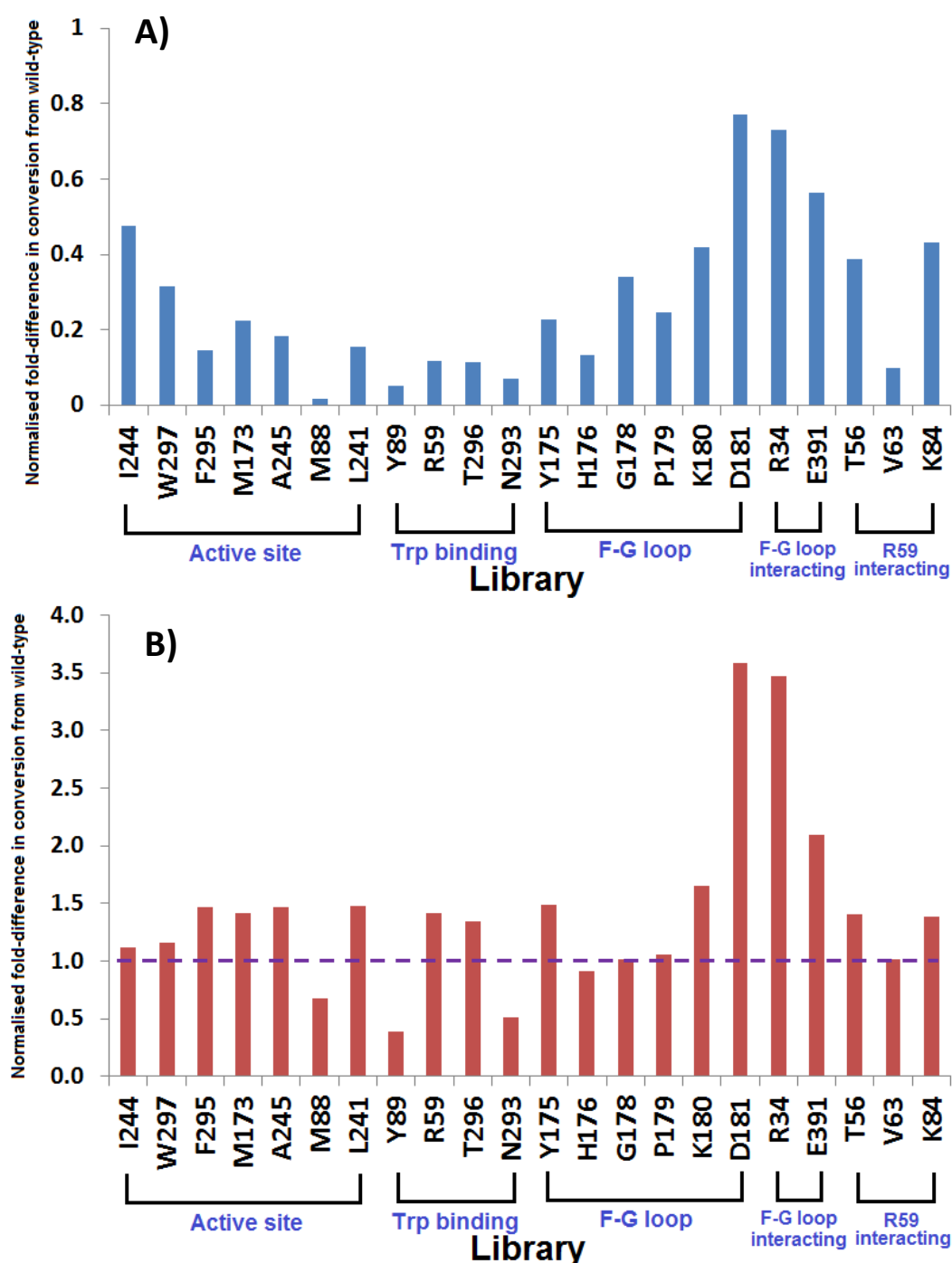


Figure 3.13. A) Bar chart of the average fold-difference in conversion for all 84 enzyme variants per library screening plate compared to the average of the three wild-type reactions for the corresponding plate which is represented as 1.0 (orange dashed line). B) Bar chart of the fold-difference in conversion for the best enzyme variant per library screening plate compared to the average of the three wild-type reactions for the corresponding plate which is represented as 1.0 (purple dashed line). Often the best mutant per plate exceeded that of the control reaction.

3.5.1 Selected positive variants

In order to analyse the variants in greater detail, 30 of the best performing variants from all libraries were selected for sequencing. A large majority, especially those that increased conversion by only 1.0-1.5-fold over the parent enzyme, were established as being the wild-type enzyme. The small variation in activity is most likely a result of gene expression differences and therefore variation in protein quantity in each well. Seven wells contained variants that did differ from wild-type enzyme (figure 3.14). The best variants were D181G, R34G and E391G, in which charged amino acids have been replaced by glycine, the simplest amino acid. Owing to its small size and lack of a side chain, glycine exhibits a unique degree of flexibility within protein structures not matched by any other amino acid. All three of these positions are within or near a dynamic region of TxtE, the F-G loop. The fact that charged amino acids, capable of forming hydrogen bonds, have now been replaced by a flexible non-bonding residue in all three cases implies that an even greater degree of motion has been introduced with this region. Closing of the F-G loop over the active site is believed to occur before catalysis, thus an increase in the degree of flexibility of this region could shift the loop equilibrium to the more catalytically relevant state, improving efficiency.

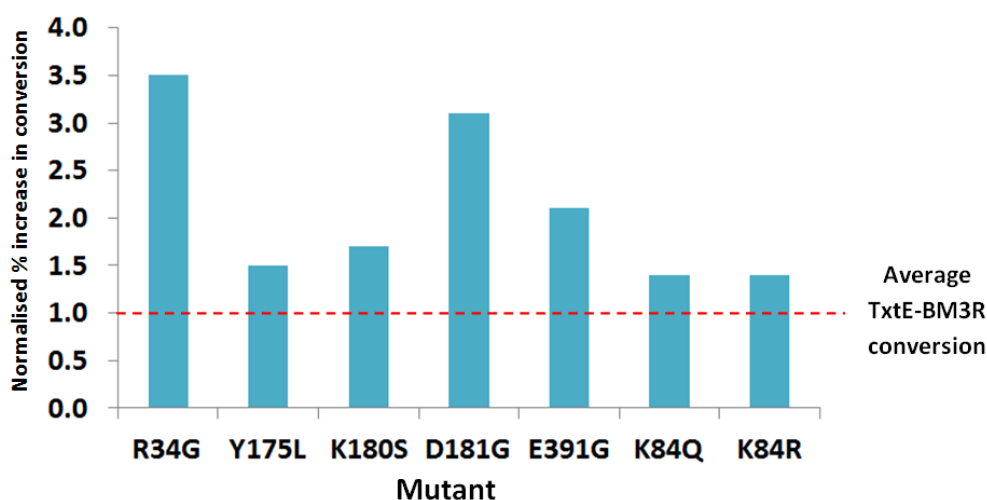


Figure 3.14. Comparison of the seven best mutants selected from the thirty that were picked from across all enzyme variant libraries. Comparisons are made against the average conversion of the three control reactions for the corresponding plate shown as being 1.0 (red dashed line). The specific mutation has been identified and enzymes which were established as being the wild-type protein were excluded.

Other changes in the F-G loop, Y175L and K180S, also produced an improvement in conversion. Again, the parent residues are comprised of either a large and rigid amino acid (Y) or a positively charged one (K), so a change to a smaller or less charged residue could be expected to decrease rigidity. Clearly further studies are required to obtain a better

understanding of the effects of the F-G loop mutations. As this region has previously been deemed too flexible for crystallographic studies, molecular dynamic simulations could be the key to better understanding the effects of these mutations.

A small increase in conversion of ~1.4-fold was observed in variants containing either an Arg or Glu residue in place of K84. K84R and K84Q would not be expected to differ majorly from the parent enzyme in terms structure but the Lys residue is proposed to interact with R59. From crystallographic studies, R59 had been shown to move through a 90° angle when L-Trp enters the active site, facilitating substrate binding.¹⁰² R59 then acts as the terminal binding residue, holding the substrate at the correct distance over the heme centre where the indole ring has the greatest potential to be affected by catalytic activity. The small improvements resulting from mutation at K84 could be because of a new interaction with R59, allowing it to more easily rotate upon substrate binding. Alternatively, since only a small increase is observed and the newly incorporated amino acid is very similar in terms of charge and size to Lys, the increase could simply be due to changes in gene expression as observed for the wild type-enzyme.

3.6 Conclusions

TxtE-BM3R mutant libraries were created and investigated for the improved conversion of L-Trp **26** to 4-nitro-L-Trp **27**. Initial studies were focussed on optimising the saturation mutagenesis techniques and the analytical methods by performing a validation experiment by generating a H176 mutant library. Previous studies have discovered a regioselectivity switch at this position in which mutation of H176 to Phe, Tyr or Trp mutant changed the site of nitration from the 4-position to the 5-position. We were successful in creating a H176 saturation library in TxtE-BM3R which confirmed that TxtE-BM3R(H176F/Y/W) variants also catalyse nitration at the 5-position of **26** producing **71**.

Generation of 26 further mutant libraries with residues targeted in the TxtE-active site, the F-G-loop and several external loop/helix residues were then attempted resulting in successful creation of 22. During the creation of these libraries we observed a large difference in the quality of the libraries, consistent with the findings of others. Analysis uncovered 7 mutations that resulted in the increased conversion of **26** to **27** compared to the wild-type enzyme. The improved variants were composed of Y175L, K180 and D181 in the F-G loop, and K84Q, K84R, R34G and E391G in positions external to the active site. Understanding the fundamental principles governing the increase in activity with the three best mutants (R34G, E391G and D181G) will require further investigations, particularly as these residues reside in dynamic regions of the TxtE heme domain.

The next step in further improving TxtE activity for **26** nitration at the 4-position would involve combining the individual mutations into one enzyme construct. The synergistic

effect of multiple mutations in a single protein has already been discussed and the potential improvement in TxtE activity by merging the mutations above will prove to be an exciting prospect.

A further use for the variant libraries will be a screen to test for altered substrate tolerance of TxtE. Identification of one such variant, capable of nitrating tryptamine is discussed in chapter 4.

Chapter 4: Identification of a TxtE-BM3R variant capable of tryptamine nitration

4. Introduction

The biosynthetic precision of enzymes can be detrimental to their usage in an industrial environment when trying to replicate enzyme activity on a non-natural substrate, e.g. a pharmaceutical building block.

In the case of TxtE, the substrate scope of the wild-type enzyme has previously been investigated and a strong preference for L-Trp **26** analogues containing both a carboxylic acid and an amine was found.¹⁰² The development of a panel of TxtE variants capable of nitrating a broader range of analogues, in particular to expand the substrate tolerance away from just amino acids, is much sought after and crucial to the employment of nitration biocatalysts in industry. We therefore set out to identify a TxtE-BM3R variant that can nitrate tryptamine **90**, which is structurally similar to **26** but lacks the carboxylic acid group.

Tryptamine **90** is a chemical moiety found in a vast range of pharmaceutical compounds and natural products (figure 4.1).¹³⁸ **90** utilised in biosynthetic pathways is derived from **26** via decarboxylation by tryptophan decarboxylase, one crucial step in the biosynthesis of the important indole alkaloid family of compounds which includes the neurotransmitter serotonin **119**,¹³⁹ the hormone melatonin¹⁴⁰ and the powerful hallucinogen psilocybin **120**.¹⁴¹ Additionally, strictosidine **122**, an intermediate in the biosynthesis of all monoterpene indole alkaloids, is derived from tryptamine **90** and secolagnin.¹³⁸ For these reasons, the ability to enzymatically nitrate tryptamine regiospecifically could prove useful for the production of novel drug or natural product analogues.

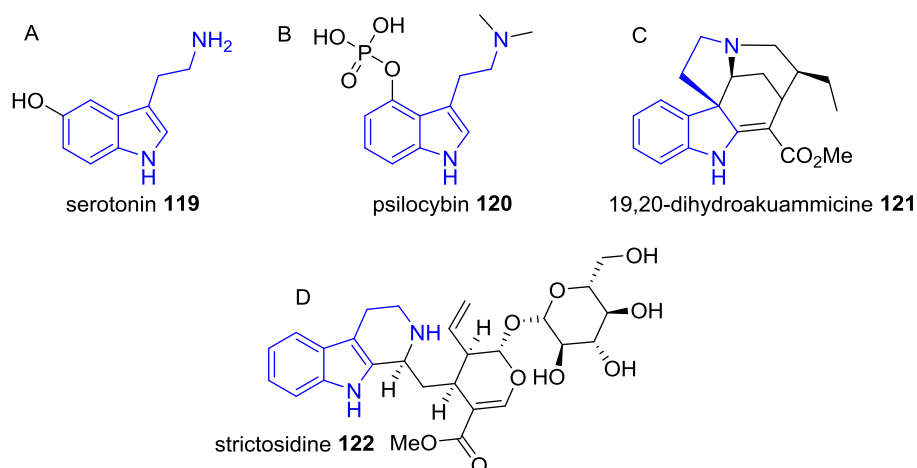


Figure 4.1. Structures of natural products that incorporate tryptamine in their biosyntheses. A) serotonin **119 B) psilocybin **120** C) 19,20-dihydroakuammicine **121** D) monoterpene indole alkaloid precursor, strictosidine **122**.**

4.1 Screening of R59 mutant library to identify tryptamine nitrating variant

The TxtE residue R59 plays a crucial role in binding and positioning the L-Trp **26** substrate and has been shown to hydrogen bond to the carboxylate of **26** and the heme propionate (figure 4.2).¹⁰² Rearrangement of the B-C loop, near which R59 resides, during opening and closing of the active site lid has been shown to be important for substrate binding in numerous CYPs.⁸⁹ It was reasoned that mutation of this charged and rather bulky residue to other amino acids could stabilise analogues of **26** with varied side chains, whether through providing different hydrogen bonding motifs or altering the spatial features in this region through incorporation of smaller amino acids. The R59 mutant library was therefore chosen to screen for variants capable of nitration of **90**.

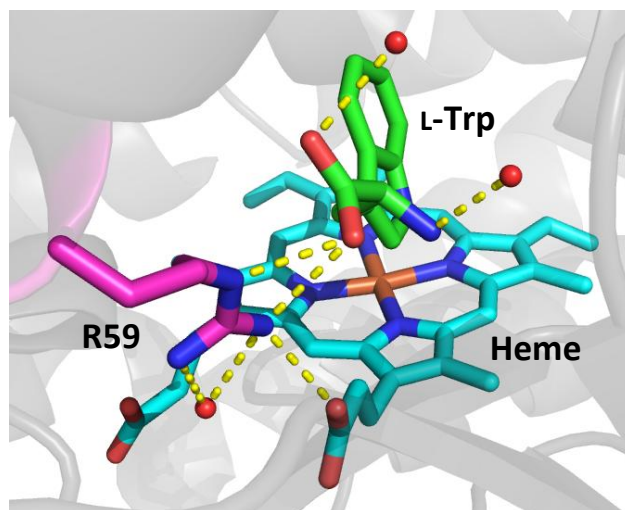
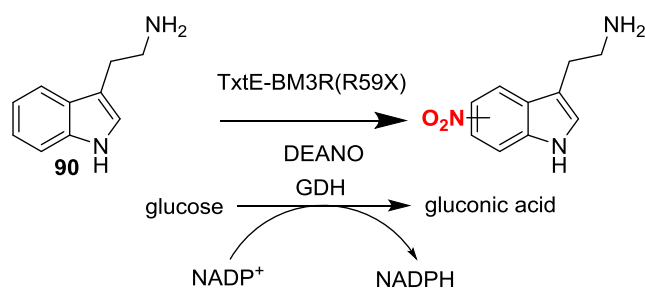


Figure 4.2. Hydrogen-bonding between R59 (magenta) and the carboxylic acid of L-Trp (green), the heme propionate (cyan) and active site waters (red spheres) in TxtE.

The screening procedure was conducted as described previously, except **90** was used as the substrate rather than **26** (figure 4.3A), and a similar LCMS method was used for analysis of the reaction plate (see experimental 7.4.4). The reaction was monitored across the whole screening reaction plate for m/z 206.2 corresponding to the $[M+H]^+$ ion of nitrotryptamine. The resulting plate heat map is shown in figure 4.3B.

A



B

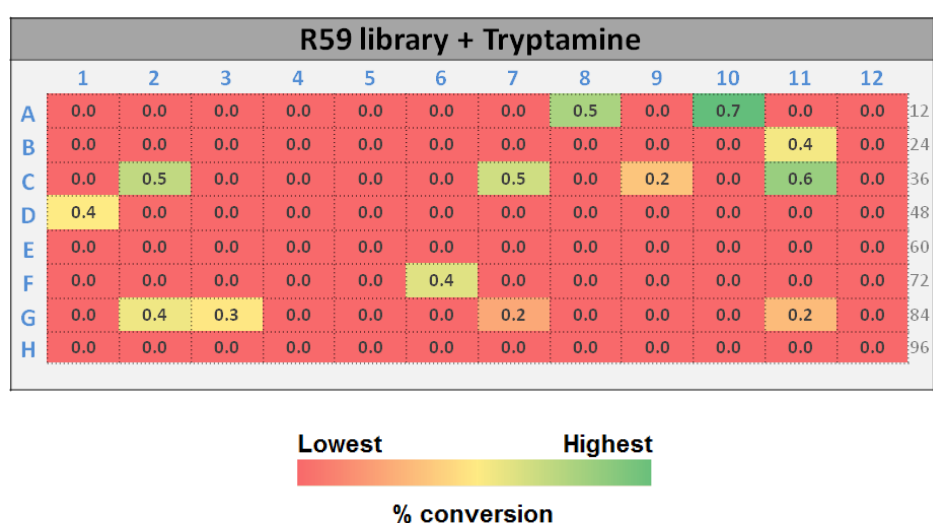


Figure 4.3. A) Library screening reaction for tryptamine nitration by TxE-BM3R(R59) variants. B) R59 mutant library screening plate heat map monitoring tryptamine nitration. Reaction monitored by LCMS for nitrotryptamine production and conversions assessed by UV-vis peak area.

Thirteen wells were identified that contained nitrotryptamine with conversions ranging from 0.2 – 0.7 %. As expected, the three wells containing wild-type enzyme (A12, B12 and C12) did not.

The four variants which produced the most nitrotryptamine were selected for sequencing, revealing them to be R59C (A8 and A10) and R59S (C7 and C11). Both cysteine and serine are similarly shaped and differ only in the presence of either the –SH or the smaller –OH functional group in cysteine (C) and serine (S), respectively.

4.2 Characterisation of TxtE-BM3R(R59C)

TxtE-BM3R(R59C) was selected for further investigations as it was the mutant with the highest conversion across the whole plate. This protein was purified from a 2 L scale culture, had the anticipated molecular weight (~111 kDa) by SDS-PAGE analysis and showed the characteristic UV absorbance at 420 nm indicating the presence of heme (figure 4.4).

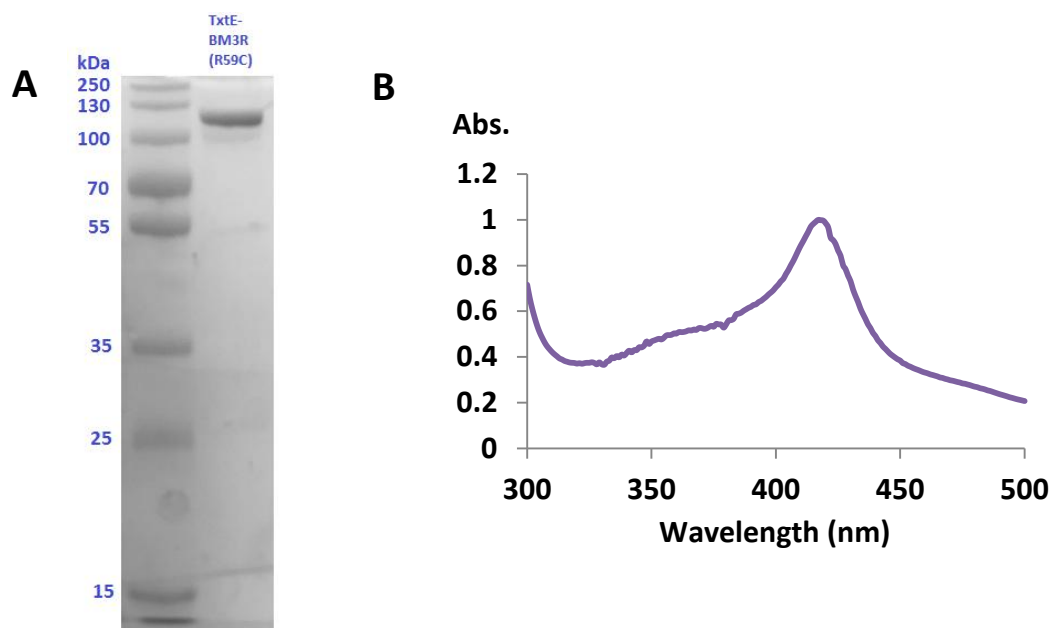


Figure 4.4. SDS-PAGE analysis of TxtE-BM3R(R59C) (112 kDa) (right-hand lane). Molecular weight marker is in the left-hand lane. B) UV-vis analysis of the TxtE-BM3R(R59C) domain with the peak at 420 nm showing the incorporation of heme cofactor in the enzyme active site.

The activity of TxtE-BM3R(R59C) was assessed by a small scale reaction using 10 μ M of purified protein and 2.0 mM of **90**. The reaction was quenched after incubation at 20°C for 4 hours and the resulting extracted ion chromatogram and mass spectra of the products from LC-MS analysis is shown in figure 4.5A. UHPLC-MS analysis is shown in figure 4.5B. A minor regioisomer elutes before the major nitro product and both peaks have an accurate mass of 206.0921 consistent with the formation of nitrotryptamine (expected $[M+H]^+$ 206.0924).

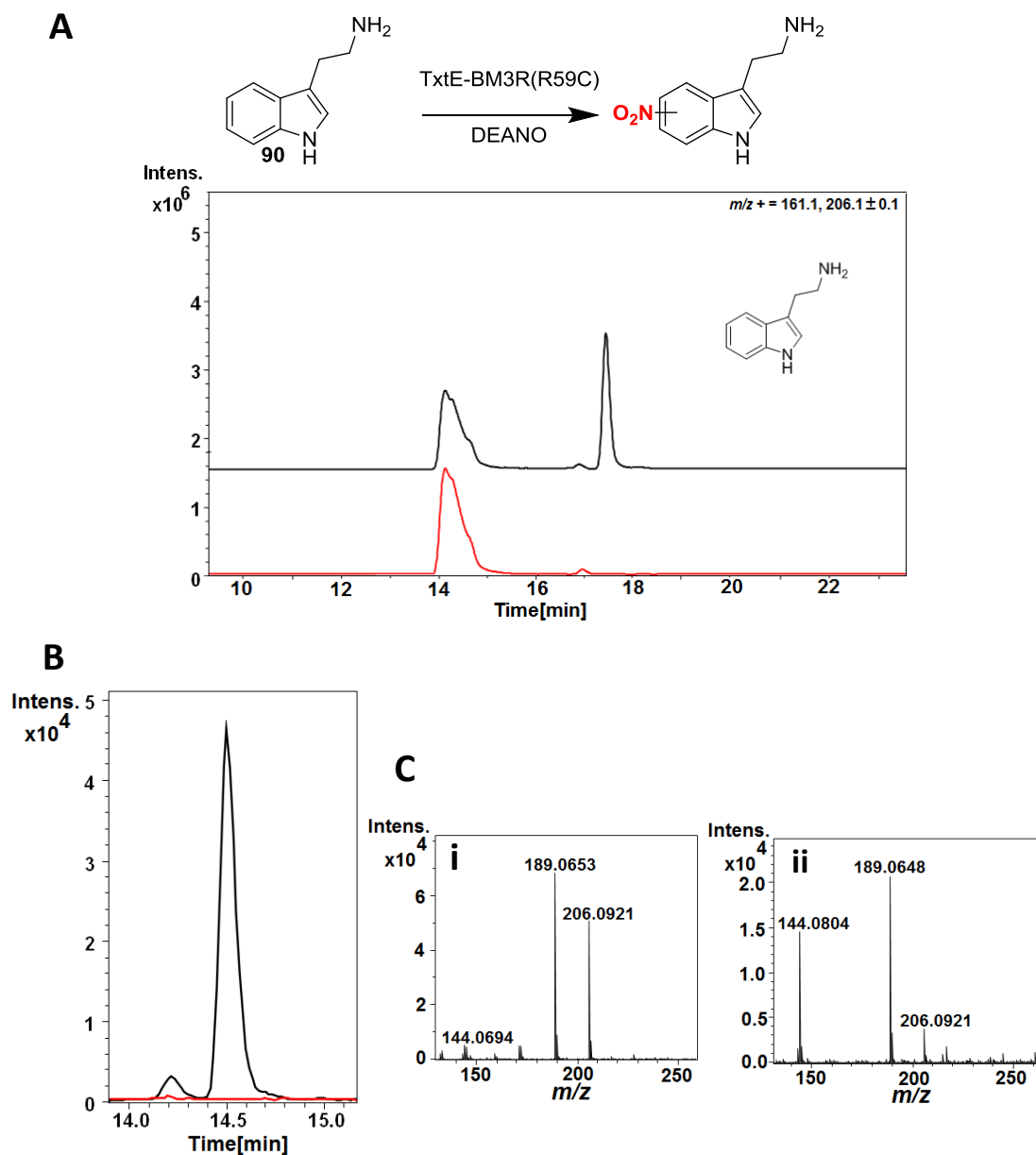


Figure 4.5. A) EIC at $m/z = 161.1$ and 206.1 corresponding to $[M+H]^+$ for 90 and nitrated 90, respectively, from LC-MS analysis of the TxtE-BM3R(R59C)-catalysed nitration of 90 reaction revealing one major product. B) EIC at $m/z = 206.1$ corresponding to nitrated 90 in UHPLC-ESI-Q-TOF MS analysis revealing major and minor products. C) i) Mass spectra for the $m/z = 206.1$ ion for the minor product ii) Mass spectra for the $m/z = 206.1$ ion for the major product.

4.3 Activity of TxtE-BM3R(R59C) with L-Trp

As a key L-Trp **26** hydrogen-bonding interaction had been lost in the new TxtE variant, we investigated whether nitration of **26** is still supported. Several assays were performed with substrate concentrations ranging from 0.5 mM to 2.0 mM. Only in assays using 2.0 mM substrate was a small peak corresponding to nitrotryptophan observed in the LCMS chromatogram, which was absent from the control reaction (figure 4.6). Since the wild-type enzyme produced nitrotryptophan at substrate concentrations as low as 0.1 mM, these results pointed towards a much weakened L-Trp binding interaction with TxtE-BM3R(R59C).

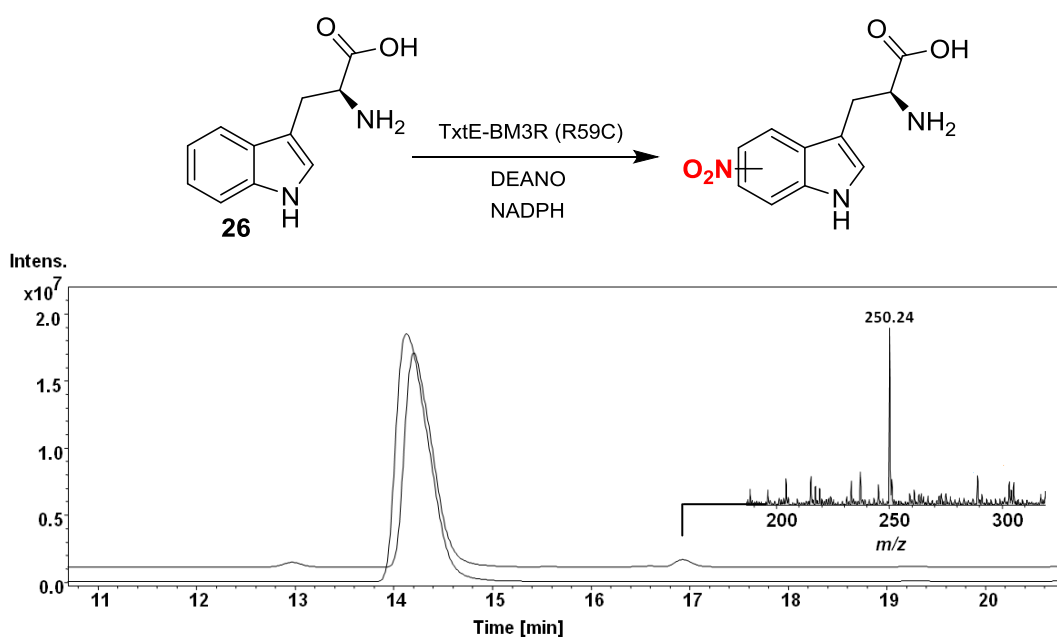


Figure 4.6. EICs at $m/z = 205.1$ and 250.1 corresponding to $[M+H]^+$ of **26** and nitrated **26**, respectively from LC-MS analysis of the TxtE-BM3R(R59C)-catalysed nitration reaction (top). A control reaction using boiled enzyme reagent is also shown (bottom). Inset: Mass spectrum of the $m/z = 250.1$ ion corresponding to $[M+H]^+$ for nitrated **26**.

4.4 Expression and activity of TxtE(R59C) heme domain

To check whether TxtE(R59C) is able to catalyse tryptamine nitration as a standalone enzyme, a stop codon was introduced at the terminus of the TxtE(R59C) gene allowing the ~ 48.5 kDa enzyme to be overproduced and purified without the BM3R reductase domain. Purification by Ni⁺-NTA chromatography yielded relatively pure enzyme by SDS-PAGE analysis (figure 4.7).

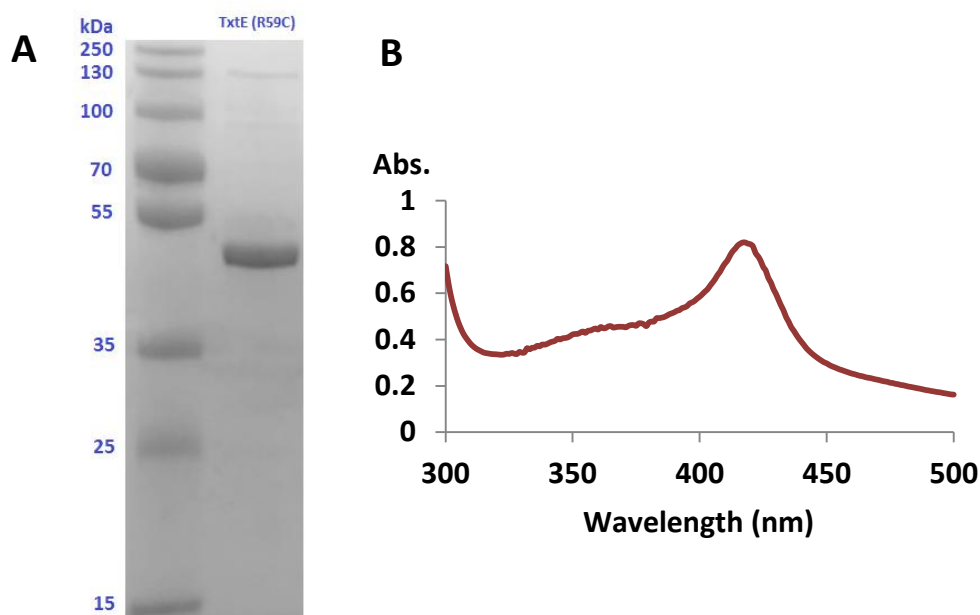


Figure 4.7. A) SDS-PAGE analysis of TxtE(R59C) (48.5 kDa) (right-hand side). Molecular weight marker is in the left-hand lane. B) UV-vis analysis of the TxtE(R59C) domain with the peak at 420 nm showing the incorporation of heme cofactor in the enzyme active site.

UV-vis analysis of TxtE(R59C) showed a prominent peak at 425 nm indicating the presence of heme cofactor (figure 4.7B). The overall incorporation of heme was comparable to that of the wild-type enzyme signifying that mutation of R59 to the much smaller cysteine amino acid did not have a large affect on the incorporation of this essential cofactor even though arginine is proposed to form a hydrogen bond to the heme propionate.

The catalytic activity of TxtE(R59C) with 2.0 mM tryptamine was reconstituted with ferredoxin (Fd), ferredoxin reductase (Fr) and NADPH. The activity of the standalone heme domain with the reductase partners from *Spinacia oleracea* was compared to that of the full length TxtE-BM3R(R59C) protein with NADPH only. The total NADPH and protein concentration of each variant was consistent throughout all assays including the control reactions. LCMS analysis confirmed that nitration only occurred in the presence of both reductase partners Fd and Fr as expected but less nitration occurred than when the reductase fusion protein was used (figure 4.8). This could be a result of higher efficiency and resultant turnover by the chimeric protein. However, as the Fd and Fr are purchased as soluble proteins, differences in stability and concentration could mean a decreased supply of electrons to the heme centre for the standalone enzyme.

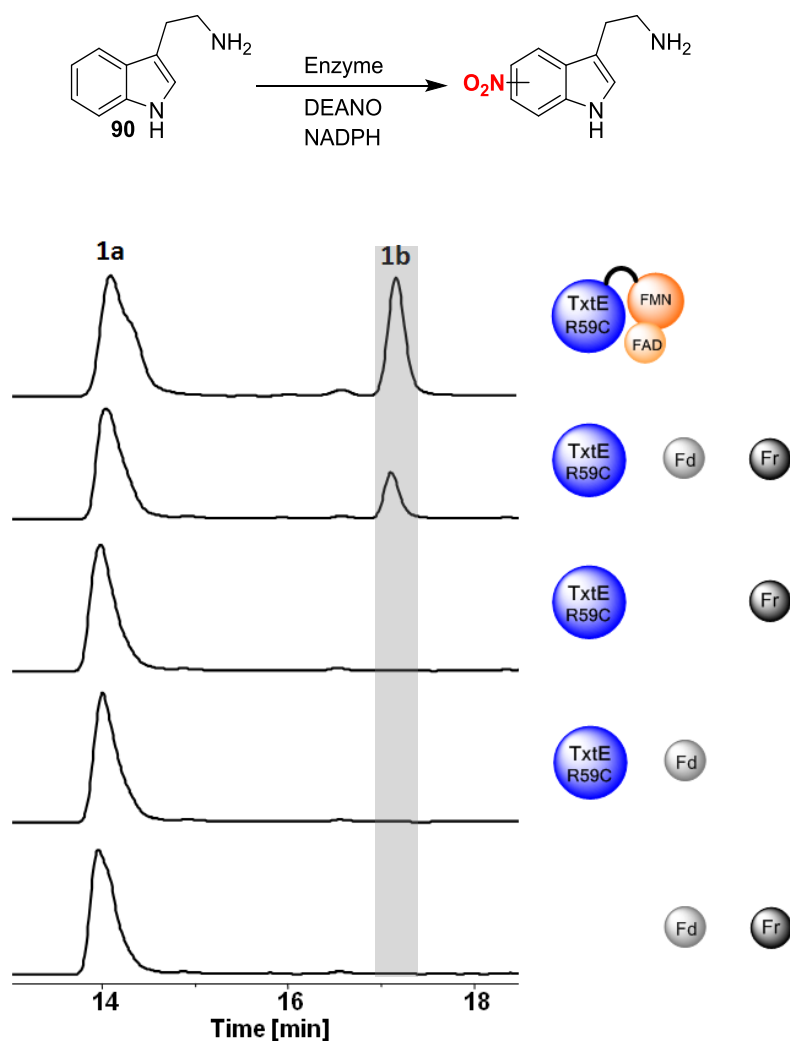


Figure 4.8. EICs at $m/z = 161$ and 206 corresponding to $[M+H]^+$ of **90** (left-hand peak) and nitrated **90** (highlighted, right-hand peak), respectively, from LCMS analyses of assays using TxtE-BM3R(R59C) or TxtE(R59C) coupled with ferredoxin (Fd) and ferredoxin reductase (Fr) or a combination of these (control reactions). NADPH (2.4 mM final) was consistent in all of the above assays.

4.5 UV-Vis analysis of substrate binding to TxtE-BM3R and the R59C variant

The effect of substrate binding to TxtE-BM3R and the R59C variant was investigated by UV-Vis analysis. Figure 4.9A compares the difference spectra of **26** binding to TxtE-BM3R and TxtE-BM3R(R59C). A type I spectrum, characterised by shift in the Soret peak from 420 nm to 390 nm, was observed for both proteins in the presence of **26**, indicating a switch in the heme iron from a low spin to high spin state upon substrate binding. However, the binding affinity of **26** is visibly reduced when binding to TxtE-BM3R(R59C) represented by a large decrease in height of the type I absorbance peaks at 390 nm and 420 nm relative to that with TxtE-BM3R as the receptor. Fig.4.9B shows the corresponding

binding spectra for **90**. In contrast, tryptamine produces type II spectra indicative of direct heme coordination and therefore inhibition of the CYP by preventing formation of the ferrous dioxygen species required to drive the catalytic cycle forward. This binding mode, in which the Soret peak shifts from 420 nm to 425-435 nm, is typical of primary amine interactions with CYPs. Compared to binding to TxtE-BM3R, the spectrum peaks at 417 nm and 431 nm are more prominent when **90** is titrated into TxtE-BM3R(R59C) which would indicate that the substrate is binding to the heme iron more strongly. A possible explanation for activity towards tryptamine despite inhibitory-type binding could be that **90** mainly binds to the heme iron but it is able to switch to a productive binding mode in which the catalytic cycle is free to proceed. Type II binding spectra could be a result of the proportion of productive binding being much smaller than the heme-binding mode, thus masking the type I spectra. The amine group of **90** can potentially hydrogen bond to several residues inside the TxtE active site including N293, T296 and multiple other amino acids via water –mediated interactions. When bound to one or more of these sites, molecular dioxygen would be free to bind to the iron and the various stages of catalysis would occur. Nitration of **90** would be the final step prior to regeneration of the resting state heme species and repeat of the catalytic cycle.

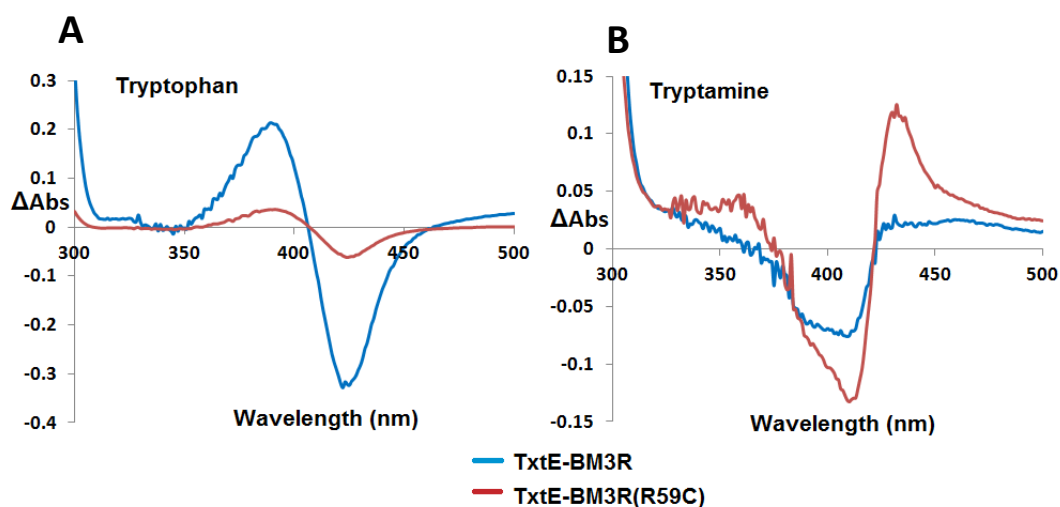


Figure 4.9. A) UV-Vis difference spectra comparing the binding of **26** to TxtE-BM3R (blue) and TxtE-BM3R(R59C) (red). B) UV-Vis difference spectra comparing the binding of **90** to TxtE-BM3R (blue) and TxtE-BM3R(R59C) (red). Final conc. of substrate added to each protein (8 μM) was 0.1 mM.

4.6 Molecular docking simulations of tryptamine binding in the TxtE-BM3R(R59C) active site

To investigate the binding preferences of **90** with TxtE-BM3R(R59C), crystallisation trials were attempted, however these were unsuccessful. A TxtE-BM3R(R59C) homology model was created using SWISS-MODEL¹⁴² with the substrate-bound wild-type TxtE structure (PDB: 4TPO) used as a template. Molecular docking simulations were then carried out with **90**, as the ammonium salt, docked into homology model using AutoDock Vina.¹⁴³ All residue side chains were set to rigid and the search was focused to a 10 Å grid box centred over the heme. The top two results equating to the two lowest energy binding modes are shown in figure 4.10.

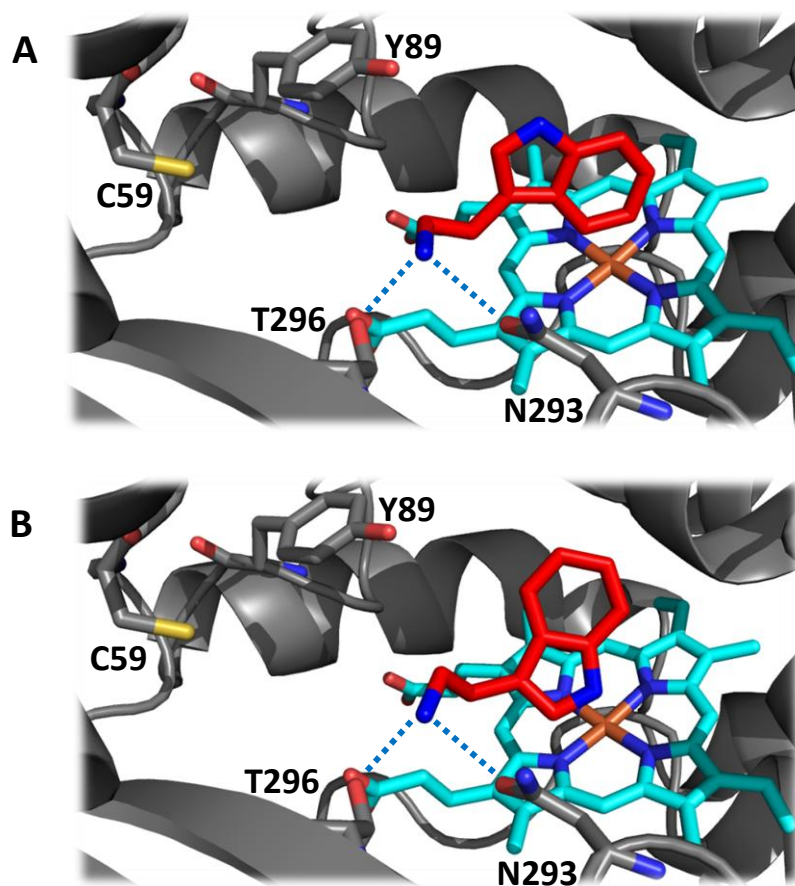


Figure 4.10. Molecular docking simulations of **90** in the modelled TxtE-BM3R(R59C) active site. A) The lowest energy binding mode showing the tryptamine (red sticks) indole moiety in a 'flipped' state with its N-H pointing towards the roof of the active site. B) The second lowest energy binding mode showing the tryptamine indole in an 'unflipped' state with its N-H pointing downwards. Substrate was docked as the ammonium form and in both cases hydrogen bonds to residues T296 and N293.

Both tryptamine binding modes mimic that of L-Trp modes in the wild-type structure observed during similar docking simulation experiments in terms of the indole ring.¹⁰¹ The lowest energy binding mode in figure 4.10A is similar to the ‘flipped’ state observed during molecular dynamics (MD) simulations in which F-G-loop conformations are taken into account.¹⁰⁷ The indole is positioned such that the N-H points towards the ceiling of the active, away from the heme iron. The docking mode in figure 4.10B is similar to the orientation that **26** adopts in the substrate-bound crystal structure. It is also representative of the open-lid state MD simulation binding mode in which the F-G loop is positioned in an arrangement permitting a solvent-accessible active site. In both cases the substrate primary amine is positioned such that it hydrogen bonds to residues N293 and T296, comparable to the interactions of the amine group of **26** in the substrate-bound structure. The evidence provided by these docking simulations support the proposition that **90** is able to assume a productive binding mode that facilitates nitration even though this particular substrate is likely to coordinate the heme iron a majority of the time. Note: Heme-binding interactions are omitted from the substrate docking and the F-G loop is in the open-lid state.

4.7 Determining the regioselectivity of tryptamine nitration

4.7.1 Scale-up of the tryptamine **90 nitration reaction**

To determine the regioselectivity of the TxtE-BM3R(R59C) catalysed nitration of tryptamine **90**, the enzymatic reaction was scaled-up to a 15 mL volume using 40 mg purified protein and 4.8 mg tryptamine. The reaction was stirred for 20 hours after which the solution changed from an orange/brown colour to a bright yellow colour. The product was purified by HPLC after removal of the precipitated protein, yielding a single pure nitrotryptamine isomer (1.1 mg, 18 %) which was characterised by LC-MS (figure 4.12) and NMR (figure 4.13). The aromatic region of the NMR spectrum contained a triplet, a singlet and a pair of doublets, suggestive of either 4- or 7-nitration. The chemical shifts of these peaks are similar to that of 4-nitrotryptophan **27** (section 2.6.2), however to be entirely sure of the regioselectivity, a 4-nitrotryptamine **124** standard was synthesised.

4.7.2 Synthesis of 4-nitrotryptamine **124 synthetic standard**

A previously reported method to synthesise tryptamine is via decarboxylation of tryptophans, however the direct nitration of tryptophan is impractical due to a lack of regioselectivity. As discussed in section 2.6.2, tryptophans can be made from the corresponding indole analogues via reaction with L-serine, AcOH and Ac₂O, followed by deprotection of the resulting acetylated product.

Another possibility is direct conversion of indole to tryptophan by tryptophan synthase (TrpS), the enzyme responsible for tryptophan biosynthesis in bacteria, fungi and plants, but not in animals.¹⁴⁴ Initial use of this enzyme in the synthesis of tryptophan analogues was hindered by low yields for many substrates, especially 4- and 7-substituted indoles.¹⁴⁵ More recently, TrpS β -subunits (TrpB) that have been engineered to function without the α -subunit (TrpA) responsible for formation of indole and glyceraldehydes-3-phosphate have become available.¹⁴⁶ Further work then focussed on engineering TrpB to accept a wider range of substituted indole including those with large or electron-withdrawing groups at the 4-position, providing enantiopure product in high yields. One such variant (*Pf2A6*) has been specifically engineered to utilise 4-nitroindole **123** as a substrate.¹⁴⁷ 4-nitro-L-tryptophan **27** was therefore prepared by L. Alkhalaf using TrpB(*Pf2A6*) according to the outlined procedure and kindly supplied as a pure product. The compound was decarboxylated in one step by heating to 154 °C for 2 hours in cyclohexan-1-one and cyclohexanol and purified to yield **124** (21 %).

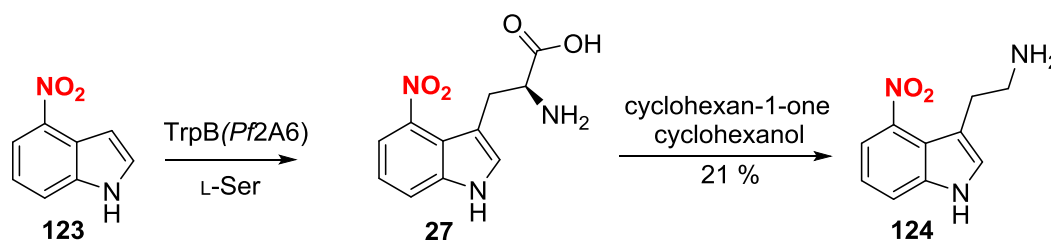


Figure 4.11. Route to **124** used in this study using an engineered TrpB(*Pf2A6*) variant with good activity for **123**, followed by a decarboxylation.

4.7.3 Comparison with the authentic standard

LC-MS analysis showed the retention times of the authentic standard and enzymatic product were identical, which was confirmed by co-injection of the samples (figure 4.12). Additionally, the ¹H NMR spectra were indistinguishable and showed one set of peaks when a combined spectrum obtained.

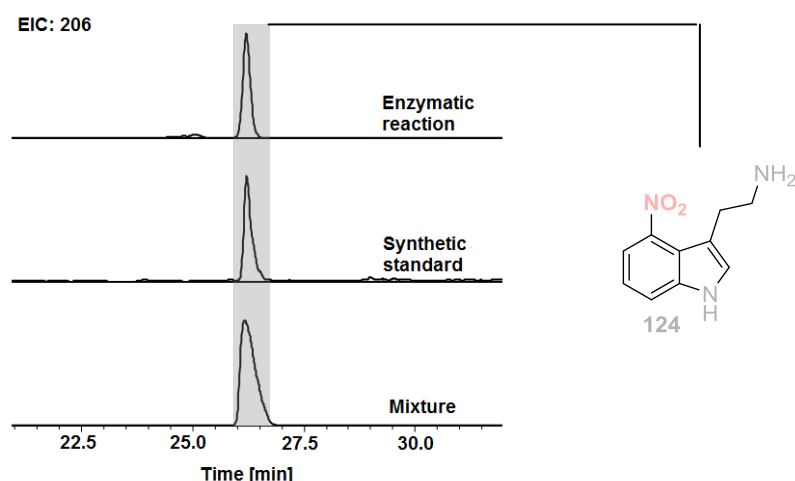


Figure 4.12. LCMS analysis of synthetic and enzymatic 124. EIC at $m/z = 206$ corresponding to $[M+H]^+$ of 124 (highlighted) for the enzymatic product isolated from the reaction, the synthetic 124 product and a mixture of the two solutions.

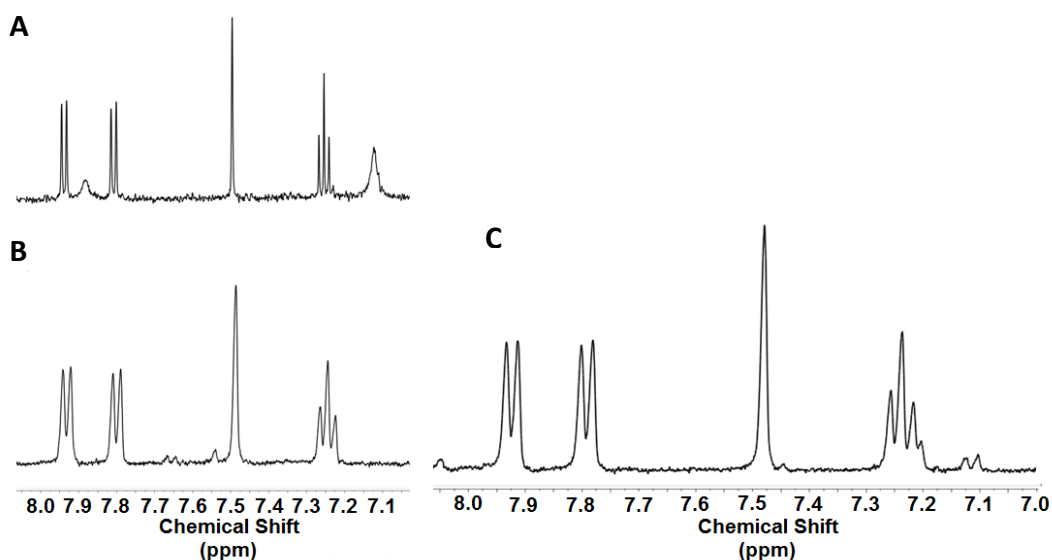


Figure 4.13. ^1H NMR spectroscopic analysis of synthetic and enzymatic 124 highlighting the aromatic region of A) Isolated product from reaction of TxtE-BM3R(R59C) with tryptamine B) synthetic 124. C) A mixture of enzymatic 124 isolated from the reaction and synthetic 124.

The results confirm that the major product following the nitration of tryptamine by TxtE-BM3R(R59C) is 4-nitrotryptamine **124**. There was an insufficient amount of the minor regioisomer to determine its regiochemistry. Retention of regioselectivity suggests the

indole of **90** and **26** is similarly orientated within the active site of TxtE(R59C) and wt-TxtE, respectively.

4.8 Conclusions

Screening of an R59 mutant library of TxtE-BM3R for the ability to nitrate tryptamine **90** identified two potential variants, TxtE-BM3R(R59C) and TxtE-BM3R(R59S). TxtE-BM3R(R59C) was then overproduced, purified and characterised by UV-Vis spectroscopy. Whilst **90** still appears to bind predominantly in a non-productive manner, the mutation must allow a shift towards productive binding that leads to product formation. Scale-up of the reaction and comparison of the product to a synthesised authentic standard confirmed the major nitrated product was the 4-nitro regioisomer **124**. In spite of the vast increase of the scale compared to analytical scale assays on a 0.1 mL reaction volume, the procedure remains simplistic and cost efficient largely owed to the use of the CYP-reductase fusion protein. This enzyme expresses well and is stable enough to not require any special handling techniques often associated with other classes of iron-containing proteins.

These results showcase the potential for developing nitration biocatalysts that tolerate substrates other than those containing amino acid functionality and the ability to support binding and reaction of compounds with fewer hydrogen bonding motifs. The R49C mutant can also provide a platform for future engineering efforts by acting as the parent enzyme in further rounds of mutagenesis focussed on activity with compounds containing useful synthetic handles. The variant may also provide the starting point for substrate walking experiments in which an enzyme is evolved over numerous rounds of engineering, screening for activity with substrates of increasing complexity and size. This approach has been adopted by Lewis *et.al* for the tryptophan halogenase, RebH, and was successful in evolving activity from tryptamine to the large natural product yohimbine.¹⁴⁸ The first step of this progression is presented in Chapter 5 and builds upon the work presented here by investigating the current substrate tolerance of the R59C mutant.

Chapter 5: Substrate scope and mechanistic investigation of TxtE-BM3R(R59C)

5.1. Substrate scope of TxtE-BM3R(R59C)

A panel of substrates was selected for investigating the substrate tolerance of the TxtE-BM3R(R59C). All assays were conducted as described in section 2.3 by incubation of TxtE-BM3R(R59C) with the substrate, DEANO and NADPH for 4 hours. A negative control was carried out in each case using heat denatured TxtE-BM3R(R59C). Substrate binding was investigated by UV-Vis spectroscopy as described in section 2.2, but will not be discussed here as spectra type assignment was inconclusive.

A number of 5-substituted tryptamine analogues were chosen for testing as this moiety is present in wide range of natural products, most notably serotonin and melatonin as described previously (section 4.1), but also in the structurally related bufotenine¹⁴⁹ **125** and the alkaloid physostigmine¹⁵⁰ **126** (figure 5.1). Tryptoline **136** was selected as it is also an important scaffold in several natural products including pinoline **127** (figure 5.1.C) and the related compounds eleagine and tetrahydroharmine¹⁴⁸ (not shown). Tryptamine analogues containing benzofuran or benzothiophene were tested with TxtE-BM3R(R59C) to primarily investigate the nitration mechanism. However, both structures are biomedically important,^{33,151} and the ability to nitrate these scaffolds is attractive from an industrial perspective. Most notable is the benzofuran-containing natural product griseofulvin **128**, an antifungal from *Penicillium* sp. first discovered in 1939 and still in use today (Fig.5.1D).¹⁵²

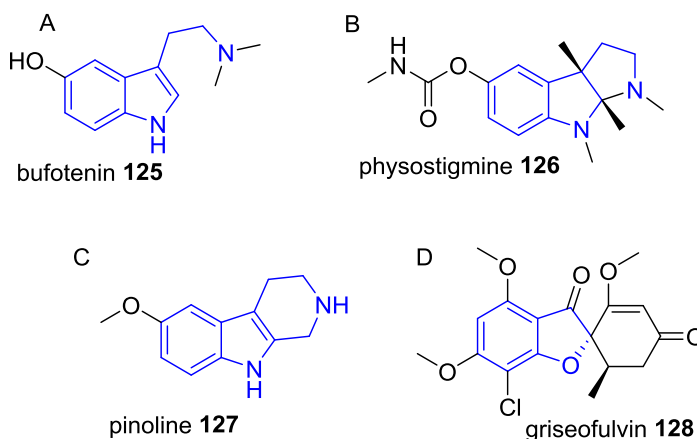


Figure 5.1. A) Structure of bufotenin 125 B) Structure of physostigmine 126 C) Structure of pinoline 127 D) Structure of griseofulvin 128.

5.1.1 Modification to the 5-position of tryptamine

The 5-substituted tryptamine analogues (5-fluoro- **129**, 5-methyl- **130** and 5-methoxytryptamine **131**) were selected to investigate whether both indole ring electronics and substrate sterics influenced the degree of nitration displayed by TxtE-BM3R(R59C). Previously, wt-TxtE has been shown to accept a number of 5-substituted L-Trp analogues. As can be seen in figure 5.2, a nitrated product was observed for all the 5-substituted substrates tested with TxtE-BM3R(R59C), although total conversion was lower than that for tryptamine **90** (figure 4.5A). The major product in each case is proposed to be the 4-nitro regioisomer since this is the site of nitration of **90**. Furthermore, the regioselectivity of wt-TxtE has been shown to be unaffected by substitution at the 5-position.

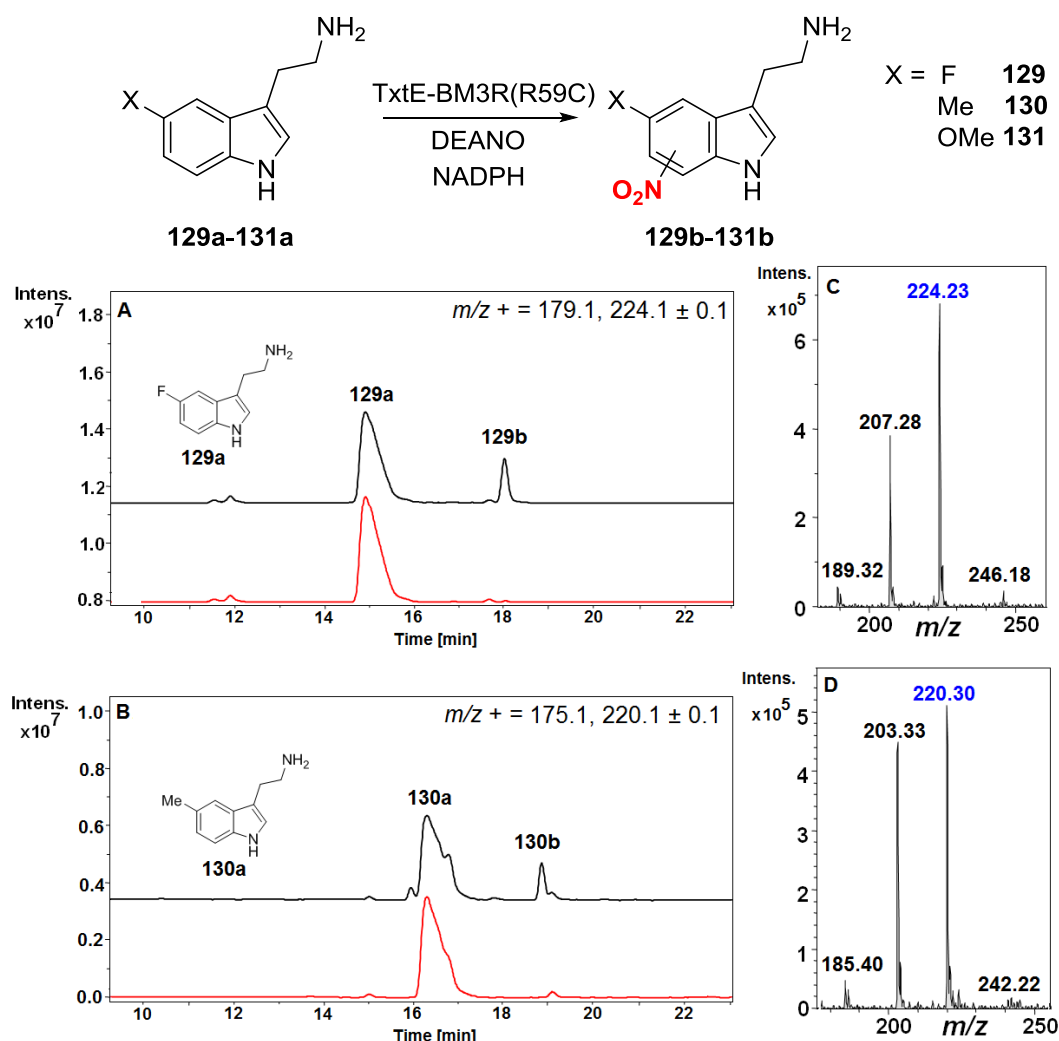


Figure 5.2. EICs at A) $m/z = 179.1$ and 224.1 corresponding to $[M+H]^+$ for **129a** and **129b**, respectively, and B) $m/z = 175.1$ and 220.1 corresponding to $[M+H]^+$ for **130a** and **130b**, respectively, from LCMS analysis of nitration assays containing **129a** or **130a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). C) Mass spectrum of the $m/z = 224$ ion corresponding to $[M+H]^+$ of **129b**. D) Mass spectrum of $m/z = 220$ ion corresponding to $[M+H]^+$ of **130b**.

Reaction with the 5-methoxy analogue **131** also led to production of a second nitrated product (figure 5.3A), which is presumably an alternative regioisomer. Interestingly, a small amount of the major regioisomer was also present in the control reaction. The methoxy-substituent on this analogue increases the electron density of the indole ring making this particular substrate more susceptible to attack by peroxynitrite species formed from the non-catalysed reaction of NO with oxygen.¹²²

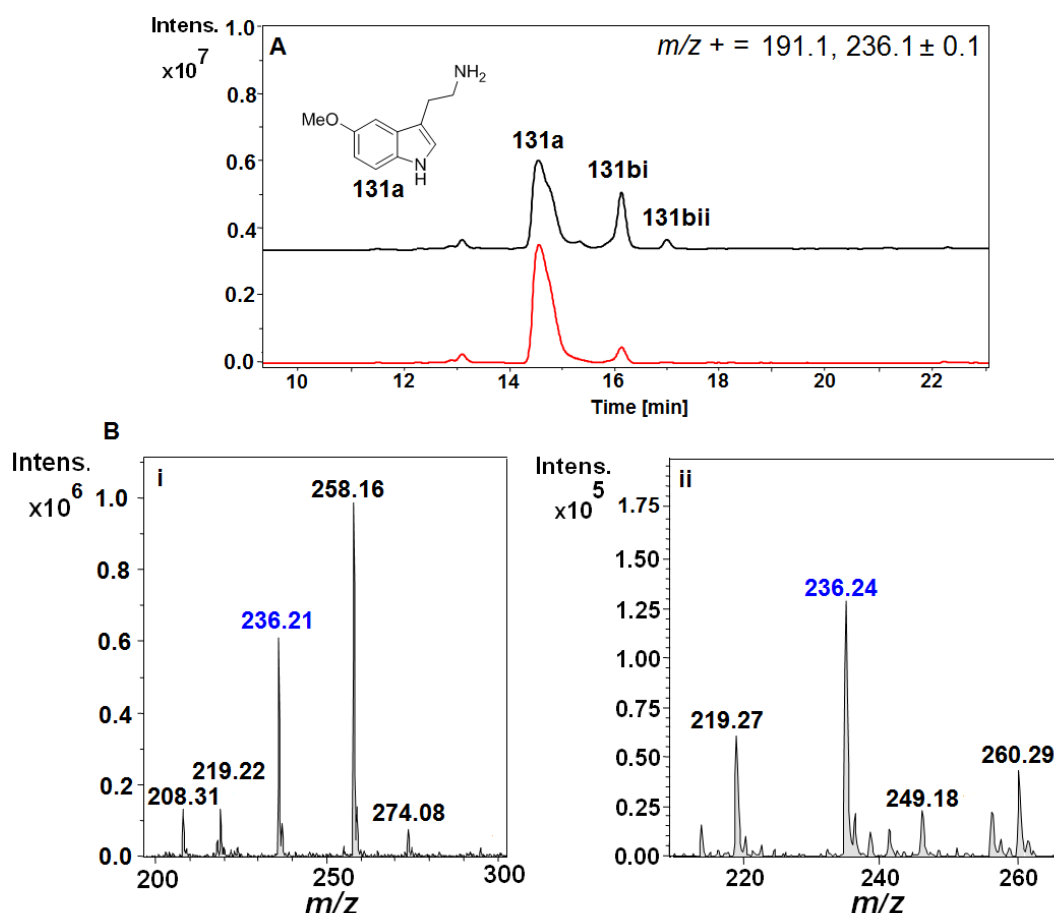


Figure 5.3. A) EIC at $m/z = 191.1$ and 236.1 corresponding to $[M+H]^+$ for **131a** and **131b**, respectively, from LCMS analysis of nitration assays containing **131a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). B) Mass spectra of peaks with $m/z = 236$ at i) $t = 16.1$ min and ii) $t = 16.9$ min, corresponding to $[M+H]^+$ of regioisomers **131bi** and **131bii**, respectively.

5.1.2 Modification of the amino acid of tryptamine

Substrates **83**, **86**, **91**, and **133** had all previously been investigated for activity with wt-TxtE but were found to not undergo nitration in our investigations. **91** demonstrated type II inhibitor-like binding with the wild-type enzyme similar to tryptamine and so nitration also by the R59C mutant may also be plausible. Substrates **86** and **133** have both been shown to demonstrate type I binding with the wt-enzyme however they are most likely orientated or positioned incorrectly for productive chemistry to occur.¹⁰²

Interestingly, substrates **91** and **86**, were successfully nitrated by TxtE-BM3R(R59C) despite not being accepted by the wild-type enzyme. The reactions with each of these substrates produced a single product as determined by LCMS analysis. Substrates **132-134** were not nitrated even though these compounds contain either a hydroxyl or amine group within their structures (data not shown).

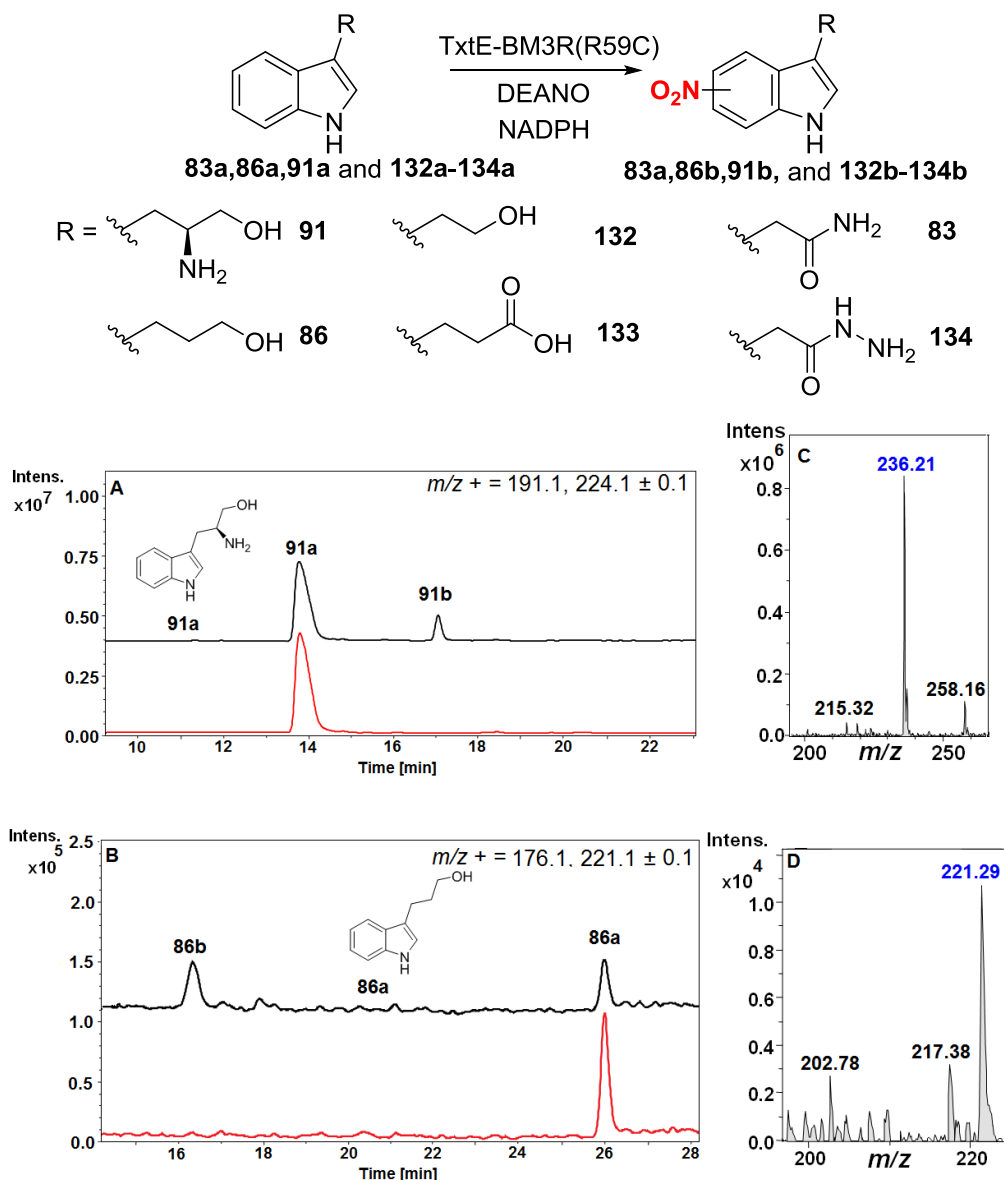


Figure 5.4. EICs at A) $m/z = 191.1$ and 236.1 corresponding to $[\text{M}+\text{H}]^+$ for **91a** and **91b**, respectively, and B) $m/z = 176.1$ and 221.1 corresponding to $[\text{M}+\text{H}]^+$ for **86a** and **86b**, respectively, from LCMS analysis of nitration assays containing **91a** or **86a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). C) Mass spectrum of the $m/z = 224$ ion corresponding to $[\text{M}+\text{H}]^+$ of **91b**. D) Mass spectrum of $m/z = 220$ ion corresponding to $[\text{M}+\text{H}]^+$ of **86b**.

5.1.3 Tricyclic analogues

Also included within the panel are tricyclic compounds **135** and **136**, two substrates not previously tested for activity with wt-TxtE. The constrained tryptamine analogue,

tryptoline (**135**), is a building block widespread in pharmaceuticals and natural products with derivatives displaying activity as potent monoamine oxidase inhibitors (MAOIs).¹⁴⁸ No nitrated derivatives could be identified following attempted nitration of **136** (data not shown). Nitrated products could be observed however, following TxtE-BM3R catalysed nitration of **135**, although conversion was low. Interestingly, two regioisomers were identified, suggesting a loss of regioselectivity for this substrate. This provides the first example of a substrate accepted by TxtE that has a constrained structure, the production of nitrotryptoline opens up the possibility of producing nitrated analogues of several closely related natural products (Figure 5.1). In related studies by Lewis *et al.*, similar compounds were successfully chlorinated by RebH variants that initially accepted tryptamine and tryptoline as substrates.¹⁴⁸

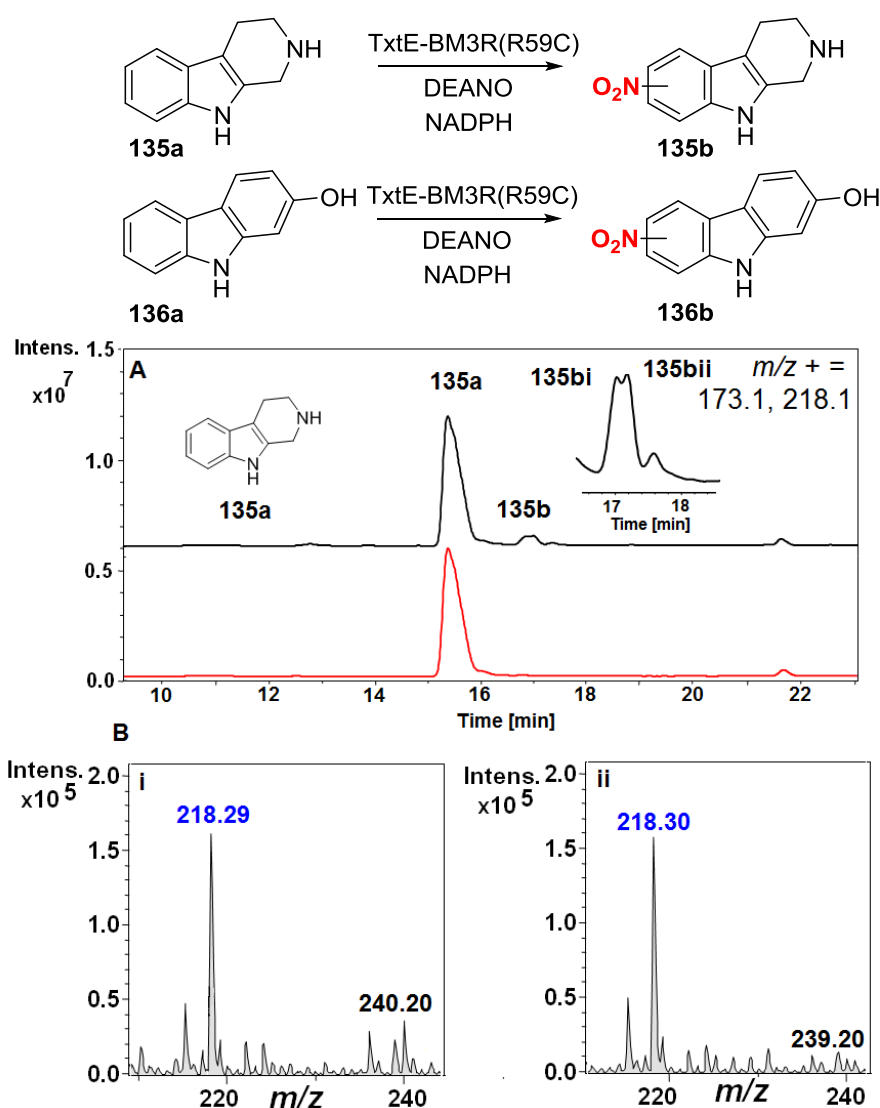


Figure 5.5. A) EIC at $m/z = 173.1$ and 218.1 corresponding to $[M+H]^+$ for **135a** and **135b**, respectively, from LCMS analysis of nitration assays containing **135a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). B) Mass spectra of peaks with $m/z = 218$ at i) $t = 17.0$ min and ii) $t = 17.2$ min, corresponding to $[M+H]^+$ of regioisomers **135bi** and **135bii**.

5.1.4 Modification of the indole nitrogen to probe the mechanism of nitration

As discussed in section 1.6.1, the mechanism for the TxtE-catalysed reaction has been proposed to proceed either via a nitro radical or a nitro cation. A nitro radical based reaction would require a hydrogen abstraction from the indole N-H. It has previously been shown that whilst wt-TxtE will bind an *N*-methylated L-Trp analogue **89**, no nitration occurs.¹⁰² Similarly, a benzofuran analogue of L-Trp (**140**) was also not an accepted substrate (D. Withall, MChem project). These results would suggest abstraction of the indole N-H is involved in the mechanism as removal of this labile proton inhibits TxtE activity. We therefore set out to repeat these experiments with TxtE-BM3R(R59C) and the corresponding tryptamine analogues as well as a further benzothiophene analogue **138**, which was synthesised from the readily available amino acid derivative **88**.¹⁵³ **88** has previously been shown to not undergo nitration with wt-TxtE either (section 1.6.4).¹⁰² Whilst **138** and **139** were not accepted as substrates of TxtE, a small amount of nitrated **137** could be identified in the LC-MS chromatogram (figure 5.6).

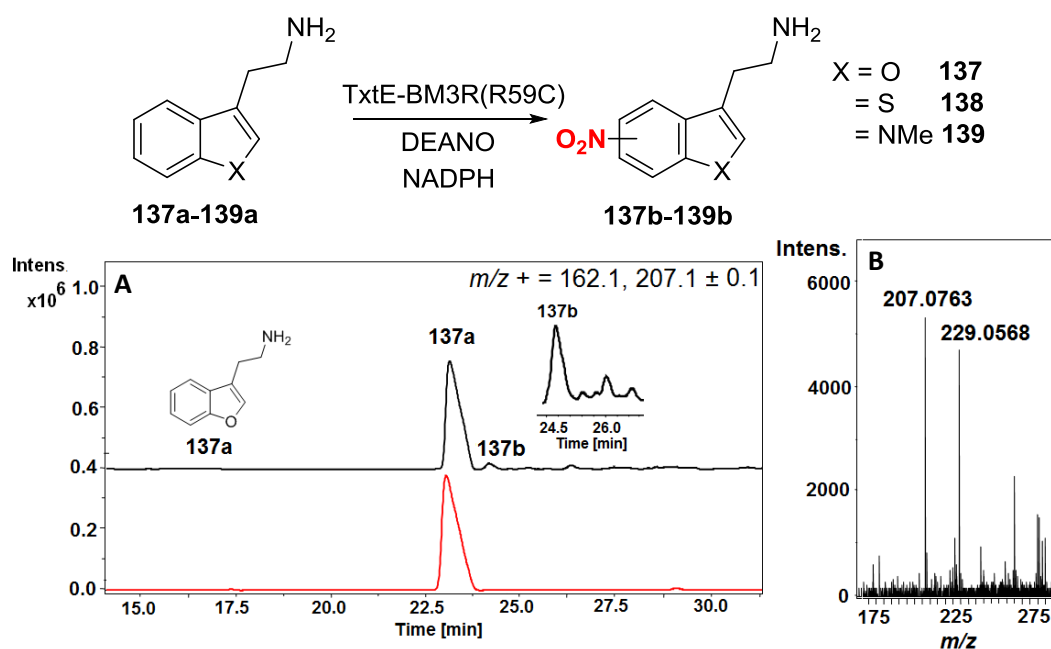


Figure 5.6. A) EIC at $m/z = 162.1$ and 207.1 corresponding to $[M+H]^+$ for **137a** and **137b**, respectively, from LCMS analysis of nitration assays containing **137a** and either TxtE-BM3R(R59C) (top trace) or heat-denatured TxtE-BM3R(R59C) (bottom trace). B) Mass spectrum of peak with $m/z = 207$, corresponding to $[M+H]^+$ of **137b**.

The products **129b**, **130b**, **131b**, **91b**, **135bi**, **135ii** were all characterised by HR-MS and masses correspond well with the calculated values (Appendix III).

5.2 Mechanistic analysis of 2-(benzofuran-3-yl)ethanamine 137 nitration by TxtE-BM3R(R59C)

Figure 5.7A. shows the two possible mechanisms for TxtE-catalysed nitration of L-Trp dependent on whether there is homolytic or heterolytic cleavage of the O-O bond of the heme-bound peroxynitrite species (figure 5.7B).

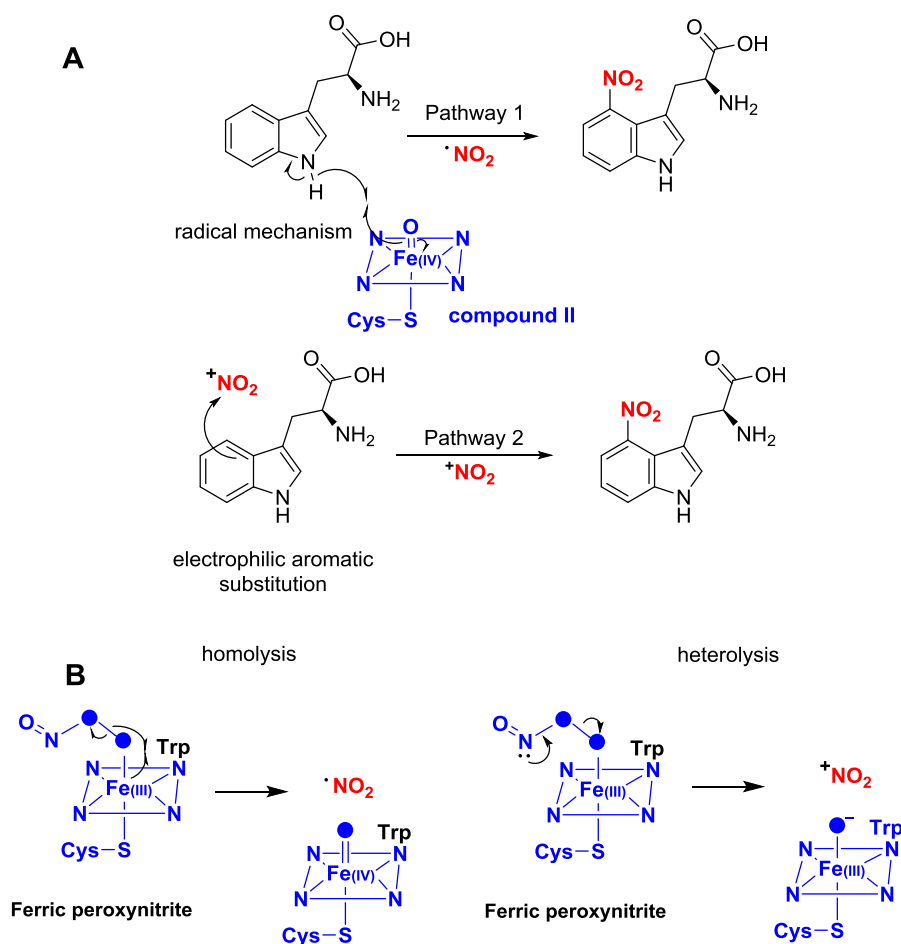


Figure 5.7. A) The two proposed pathways for indole aromatic nitration catalysed by TxtE-BM3R. Pathway 1 involves radical hydrogen abstraction of the indole N-H followed by delocalisation of the unpaired electron to the 4-position where reaction with the nitro radical can occur. B) Pathway 2 mimics classical electrophilic aromatic substitution where a nitronium cation directly reacts at the 4-position of indole.

B) Proposed peroxynitrite O-O bond cleavage mechanisms for the heme-bound species. Left: Homolytic cleavage of the bond resulting in release of a nitronium radical. Right: Heterolytic cleavage gives a nitronium cation.

Pathway 1 relies on homolytic cleavage and generation of $\cdot\text{NO}_2$. Hydrogen abstraction at the indole N-H leaves an unpaired electron on the indole nitrogen. This electron could then be delocalised within the indole ring and positioned at C-4 to react with $\cdot\text{NO}_2$ resulting in the observed 4-nitration.

Pathway 2 resembles classical electrophilic aromatic substitution. In this case, heterolytic cleavage of the O-O bond gives a nitro cation (NO_2^+) which would directly attack the 4-position of L-Trp. Elimination of the proton by a basic active site water, would result in the observed 4-nitrated product.

The fact that abstraction of hydrogen is not possible for **137** means nitration of this substrate is consistent with pathway 2. A similar mechanism would be expected for the benzothiophene analogue **138**. **138** was tested with TxtE-BM3R(R59C) but did not react even when the concentration of enzyme was increased or the reaction time was lengthened.

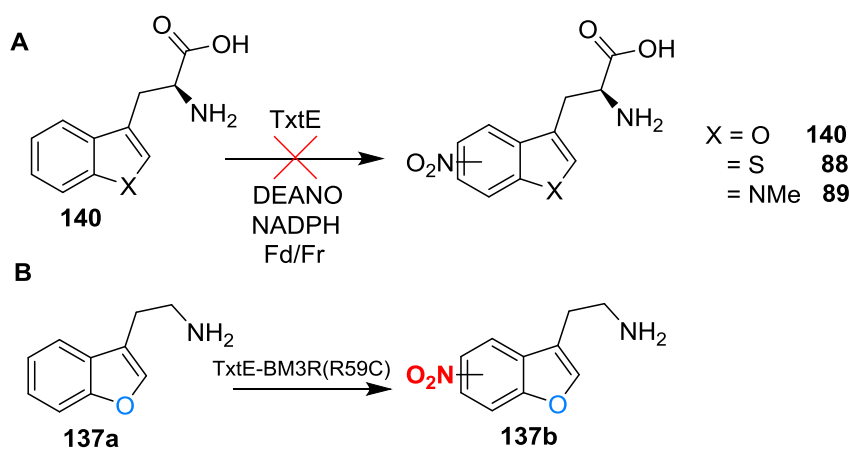


Figure 5.8. A) Tryptophan analogues modified at the indole N-H that do not undergo nitration when incubated with wt-TxtE B) Reaction of 137 with TxtE-BM3R(R59C) produces a single nitro product.

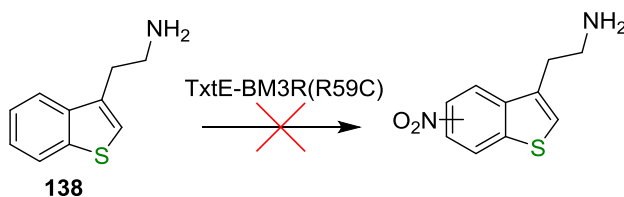


Figure 5.9. Reaction of 138 with TxtE-BM3R(R59C) also did not produce nitrated product as determined in these studies.

A possible explanation for this is that benzothiophene is less reactive in electrophilic aromatic substitution than benzofuran, and even more so than indole. The reason for this is that the sulphur atom of benzothiophene has a lone pair of electrons in a 3p orbital that overlaps less effectively with the 2p orbitals of the carbons. The N and O heteroatoms of indole and benzofuran on the other hand have their lone pair located in a 2p orbital that overlaps much more efficiently with those of the ring carbons during conjugation and are therefore more effective at donating their electrons into the benzene ring. Additionally, the 2p orbital containing the benzofuran oxygen lone pair does not overlap with the carbon 2p

orbitals as well as those of the nitrogen in indole. Hence the order of increasing reactivity towards electrophilic aromatic substitution is: benzothiophene, benzofuran, indole.

This correlates with the observed reactivity pattern for nitration of the three heterocycles by TxtE-BM3R(R59C).

5.3 Analysis of *N*-methyl-L-tryptamine binding to TxtE-BM3R(R59C)

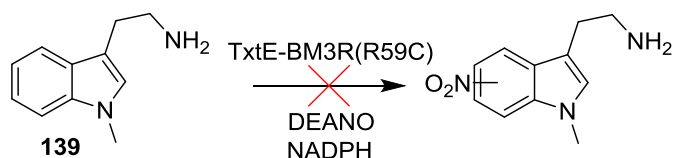


Figure 5.10. Incubation of the tryptamine analogue **139 with TxtE-BM3R(R59C) also does not result in the formation of nitro product.**

Similar to **137** and **138** *N*-methylated analogue **139** does not have an indole N-H. The lack of nitrated **139** after incubation with TxtE-BM3R(R59C) can be rationalised by analysing molecular docking simulations with the two lowest energy binding modes from the top nine docking results shown in figure 5.12. **139** is shown to bind similar to L-Trp in the substrate-bound crystal structure (Figure 1.15) such as the primary amine hydrogen bonding to N293 and T296, and the orientation of the indole core. However, the presence of a methyl group on the indole nitrogen seems to impart greater steric bulk over the heme cofactor than the natural substrate. This could potentially hinder the formation of heme-bound catalytic species during the reaction or even block dioxygen from coordinating to the heme in the first instance.

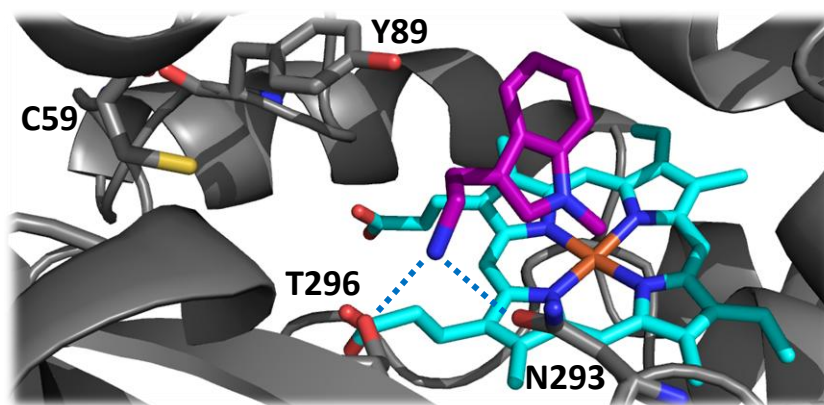


Figure 5.11. Molecular docking simulations of **139** (purple) binding in the modelled TxtE(R59C) active site. Binding mode showing the substrate in an 'unflipped' state. Substrate was docked as the ammonium form and in both cases this group hydrogen bonds to residues T296 and N293. Heme is shown in cyan.

5.4 Conclusions

Having discovered a TxtE variant capable of nitrating **90**, this chapter focused on investigating the substrate scope of this mutant. Several substituted tryptamine analogues were found to be nitrated by TxtE-BM3R(R59C) along with β -aminoindole-3-propanol **91** and the even simpler indole-3-propanol **86**, two substrates that were not accepted by the wild-type enzyme. Additionally, the first examples of enzymatic nitration of a tricyclic compound and an analogue that does not contain the core indole structure were identified using tryptoline **135** and 2-(benzofuran-3-yl)ethanamine **137**, respectively. The successful nitration of **137** suggests a classical electrophilic aromatic nitration involving a nitronium cation as the reactive species as opposed to a radical mechanism involving abstraction of a labile hydrogen. Future work would develop upon these experiments and determine whether one mechanism is consistent for all substrates or whether different mechanisms can be utilised.

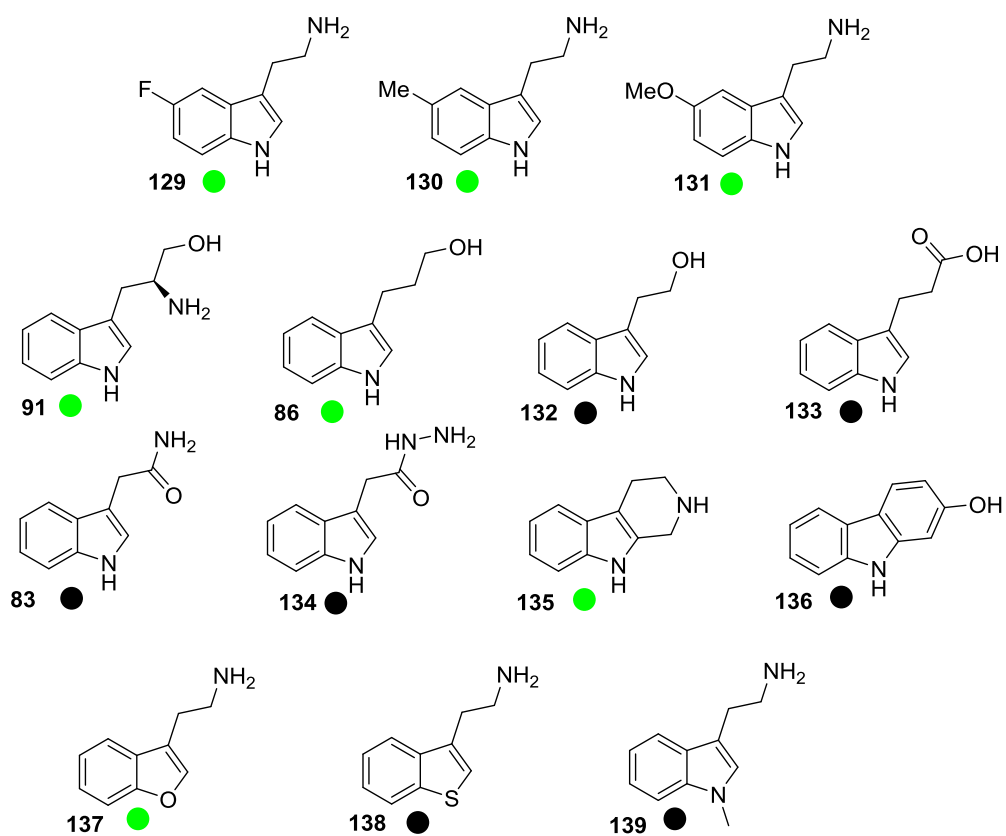


Figure 5.12. Substrate tolerance of TxtE-BM3R(R59C). Structures of all substrates tested for nitration with TxtE-BM3R.

Chapter 6: Conclusions and future work

6.1 Conclusions

6.1.1 Optimising TxtE for high-throughput or large-scale nitration

One of the major milestones associated with this project involved the preparation of self-sufficient TxtE-reductase fusion proteins. Initial work focussed on fusing to RhFRED, a naturally occurring CYP-reductase partner from *Rhodococcus* sp. NCIMB 9784 which had previously been used to reconstitute the activity of several biosynthetic CYPs. However, activity assays with **26** demonstrated a very low activity with this. TxtE-BM3R was consequently investigated and activity assays showed this fusion protein was far superior to TxtE-RhFRED. The reaction conditions were optimised to allow TxtE-BM3R to be used on a 100 mL scale utilising a NADPH regeneration system for the nitration of 4-methyltryptophan. The isolation of 8 mg of purified product was 10-times greater than previous attempts and was sufficient to use NMR characterisation to confirm the site of nitration as the 5-position by comparison to a synthetic 5-nitro-4-methyl-L-tryptophan¹⁰² standard. The reaction conditions demonstrate a plausible way in which TxtE could be employed on an industrially relevant scale. Furthermore, TxtE-BM3R was used to test several indole containing substrates in high-throughput experiments in 96-well plates validating experimental methods for future experiments.

6.1.3 Creation of mutant libraries and screening of variants for improved L-Trp nitration

With a functional TxtE-reductase fusion in hand, efforts were focussed on engineering this enzyme to accept a broader range of substrates. With the aid of Dr Pamela J. Thomas (GSK) the TxtE heme domain was analysed for potential targets for mutation. In total, 15 residues were selected in the TxtE active site, 7 residues in the flexible F-G-loop were also chosen, as well as 5 residues deemed to interact with important loops or helices within the heme domain. Site-saturation mutagenesis was conducted at these positions, in an attempt to include all possible amino acid substitutions at each of the sites. Codon redundancy was reduced via the use of the 22c-trick method¹³⁵ producing a more even distribution of amino acids at the mutagenesis site compared with traditional methods. An initial validation screen was carried out using H176, as mutations at this site have previously been discovered to affect nitration regioselectivity. Site-saturation mutagenesis followed by overproduction and reaction of individual mutants for nitration of **26** confirmed H176F/Y/W variants catalyse 5-nitration as determined by high-throughput LCMS analysis. Translating this approach to the remaining libraries uncovered several variants that had improved nitration activity of **26**, several of which were selected for sequencing

analysis. This identified two mutations, which conferred increased levels of L-Trp nitration activity of TxtE. Both residues interact with the F-G-loop, thus mutation to glycine is proposed to remove important hydrogen-bonding motifs, potentially altering the dynamics of this already highly flexible region of TxtE. These mutants merit further investigation into their potential effects on TxtE catalysis either by protein crystallography studies or the use of molecular dynamic simulations.

6.1.4 Identification and substrate tolerance of tryptamine nitration variant

A broadened substrate scope of TxtE is of significant importance if it is to be developed into an industrial biocatalyst. Tryptamine **90** was selected as a target compound for nitration by TxtE as it is a key precursor to several classes of natural products and pharmaceutical molecules. The R59 library was selected for screening the activity with **90** as this residue provides hydrogen-bonding interactions with the carboxylic acid of L-Trp that is lacking in this substrate. Screening results identified two mutations capable of facilitating nitration of **90**, R59C and R59S, with the former showing a slightly higher conversion to product. TxtE-BM3R(R59C) was purified and characterised by UV-Visible spectroscopy for its ability to bind both **90** and **26**. These experiments revealed **90** still binds mainly as an inhibitor with type II binding characteristics and that binding of **26** is severely diminished compared to binding with the wild-type enzyme. Nitration of **90** was retained by the standalone TxtE(R59C) heme domain in which the activity was reconstituted with Fd and Fr. Molecular docking simulations shed some light on the productive conformations that **90** can bind in the R59C active site and alluded to enhanced hydrogen bonding to T296 and N293 by the substrate terminal amine. Scale-up of the enzymatic reaction and subsequent isolation of the product yielded 1.1 mg of nitrated **90**. Synthesis of a 4-nitrotryptamine **124** standard enabled comparison by LCMS and NMR analysis and confirmed the site of nitration to be the 4-position.

6.1.5 Mechanistic investigations into TxtE-catalysed nitration

A nitration mechanism based on the radical abstraction of a proton from the labile indole N-H to drive the TxtE catalytic cycle was previously proposed based the inability to nitrate *N*-methyltryptophan **89** (Dr S. Barry, unpublished) or benzofuran analogue of L-Trp (Dr D. Withall). To further probe this question, **137**, **138** and **139** were tested as substrates for TxtE-BM3R. **137** was successfully nitrated, albeit with low conversion. This would suggest that radical abstraction of a proton from a heteroatom is not required for catalysis to occur and that loss of the proton from the site of nitration in a classical electrophilic

aromatic substitution mechanism is possible. Benzothiophene **138** did not react with TxtE-BM3R(R59C) possibly as a result of this aromatic ring not being activated enough towards nitration due to poor sulphur orbital overlap with the ring carbons. Alternatively, for **139** and its tryptophan counterpart **89**, it is likely that introduction of the bulky methyl-substituent hinders substrate binding to a degree large enough to prevent reactivity. It should be noted that the indole N-H of **26** also plays an important role in anchoring the aromatic ring in the active site in a productive orientation by hydrogen bonding to N293. With **137** being deficient in such a bonding motif it is likely that substrates of TxtE and variants can adopt multiple productive binding modes.

The results discussed here would imply that the O-O bond of the heme-bound ferric peroxynitrite is cleaved via a heterolytic mechanism, releasing a nitronium cation as the reactive species required for electrophilic substitution rather than a nitrogen dioxide radical resulting from homolytic cleavage.

6.2 Future work

6.2.1 Characterisation of the products of the TxtE-BM3R(R59C) variant-catalysed nitration reactions

To confirm the regioselectivity of the nitration reactions carried out by TxtE-BM3R(R59C) the reactions would need to be carried out on a preparative scale to allow for NMR characterisation. A limiting factor however is the low conversion to nitrated product for many substrates as demonstrated by LCMS analysis and the difficulty associated with isolating such small amounts of compound. To avoid the requirement for large quantities of biocatalyst and reagents, a solution to this problem could be the incorporation of the R34G and D181G mutations into TxtE-BM3R(R59C). These two individual mutations have earlier been identified as improving conversion of L-Trp to nitro-L-Trp as determined during high-throughput screening of TxtE-BM3R variants. Conversions of over 3-times that of the parent enzyme were observed for both of these mutants. If a similar effect is observed with TxtE-BM3R(R59C) then this would facilitate larger scale nitration.

6.2.2 Investigate mutant libraries for activity with other non-natural substrates

The availability of the large numbers of TxtE-BM3R variants created during these studies is beneficial for investigating substrates that are not accepted by the wild-type enzyme or the R59C mutant. The promising results arising from only testing the R59 library with **90** suggests discovering TxtR-BM3R variants capable of nitrating compounds such as indole-

3-propanoic acid **133** and 2-hydroxycarbazole **136** could be possible. Furthermore the discovery that the nitration of benzofuran-containing analogues is not limited by TxtE reaction mechanism means that finding variants capable of nitrating other benzofuran derivatives.

6.2.3 Engineer TxtE-BM3R(R59C) for further broadening of substrate scope

Another potential route to creating TxtE-BM3R variants with broader substrate tolerance or increased activity would be to further engineer the enzyme via random mutagenesis. Thus far the TxtE heme domain mutations have targeted focussed regions within the structure such as the active site and F-G-loop. This evolution approach has previously been used to increase the substrate scope of CYP102A1 (P450_{BM3})^{113, 114} with great success via techniques such as error-prone PCR. A viable course includes beginning from the R59C variant and using directed evolution to sequentially engineer TxtE to accept larger and more complex substrates (substrate walking) as has previously been demonstrated for the tryptophan chlorination enzyme RebH.¹⁴⁸ Alternatively, preparing R59C double mutants via site-saturation mutagenesis of neighbouring residues could potentially provide variants with broadened substrate scope.

6.2.4 Investigate RufO, a tyrosine nitrating CYP from Rufomycin biosynthesis

As discussed in section 1.6.3 nitrating CYP, RufO, has been discovered to be involved in the biosynthesis of rufomycin by catalysing the 3-nitration of tyrosine **72**. Analysis of the protein amino acid sequence and comparison with that of TxtE could assist in identifying key substrate recognition motifs and residues capable of supporting Tyr nitration rather Trp. Furthermore, identification of the corresponding substrate carboxylic acid binding residue in RufO (R59 in TxtE) could present an engineering opportunity to produce a RufO variant capable of tyramine nitration.

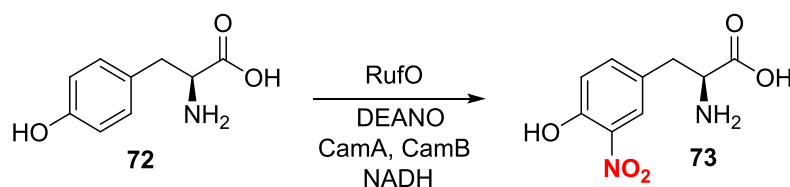


Figure 6.1: The 3-nitration of tyrosine catalysed by the CYP enzyme RufO using the nitric oxide donor, diethylamine NONOate (DEANO), putidaredoxin (CamA) and putidaredoxin reductase (CamB) redox system and NADH.

6.2.5 Investigate the biosynthesis of nitro-tetraene-containing natural products

A newly discovered gene cluster containing both a CYP with homology to TxtE and a NOS has been suggested to be involved in the biosynthesis of a lajollamycin analogue by Dr. E. de Los Santos and Dr. D. M. Roberts. Lajollamycin **16** is a nitro-tetraene spiro- β -lactone- γ -lactam first isolated from *Streptomyces nodosus*.²⁴ The formation of the nitro-polyene unit is of particular interest as it presents an enzymatic route to nitrated linear chains rather than the aromatic substrates of previously discovered CYPs of this class. Investigations are currently underway regarding the study of these enzymes from a structural and mechanistic viewpoint. Future endeavours involving the gene shuffling of these enzymes with TxtE could potentially prove a powerful tool for accessing substrate space not viable by either protein alone.

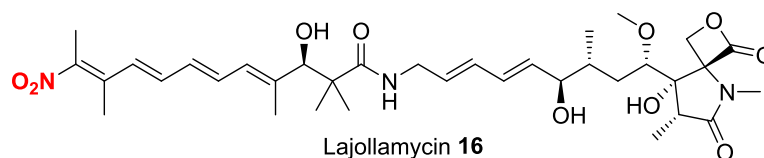


Figure 6.2: The structure of lajollamycin with the nitro group on the polyene unit shown in red.

Chapter 7: Experimental procedures

7.1 General

7.1.1 Chemicals and reagents

All chemicals and reagents were purchased from Sigma-Aldrich, Fisher Scientific, Alfa Aesar, VWR International, Acros Organics or Fluorochem and were used as supplied unless stated otherwise. Solvents were sourced from either Sigma-Aldrich or Fisher Scientific. Thin-layer chromatography was performed on silica-coated aluminium from Merck (TLC silica gel 60 F₂₅₄) and visualised using UV fluorescence and/or ninhydrin stain. Silica column chromatography was performed using silicon dioxide from Sigma-Aldrich (40-63 μm , 60 Å pore size). NMR solvents were purchased from Sigma-Aldrich or Cambridge Isotope Laboratories.

7.1.2 Instrumentation

NMR spectra were obtained on either a Bruker Avance III 400 instrument for ¹H-NMR at 400 MHz and ¹³C-NMR at 100 MHz, or a Bruker Avance III 500 instrument equipped with a DCH cryogenic probe for ¹H-NMR at 500 MHz and ¹³C-NMR at 125 MHz. Spectra were recorded at 25 °C in D₂O, DCl in D₂O or MeOD. Chemical shifts are reported as δ values in ppm relative to the internal D₂O peak (4.80 ppm (¹H)) or CD₃OD peaks (3.34 ppm (1H) and 49.8 ppm (¹³C)). Coupling constants, *J*, are reported in Hz, with multiplicities quoted as singlet (s), doublet (d), triplet (t), quartet (q), doublet of doublets (dd), doublet of triplets (dt), or multiplet (m). Low resolution mass spectra were measured on an Agilent 6130 ESI-Quad instrument. Low resolution LCMS analyses were performed on a Bruker AmaZon ESI spectrometer connected to an Agilent 1200 HPLC instrument fitted with an Agilent Eclipse Plus C18 column (5 μm , 4.6 x 150 mm). UHPLC-MS analyses were performed on a Bruker MaXis Impact ESI-TOF-MS connected to a Dionex 3000 RS UHPLC instrument fitted with an Agilent Zorbax Eclipse Plus C18 column (100 x 2.1 mm, 1.8 μm). HRMS, 500 MHz ¹H NMR and 125 MHz ¹³C NMR were obtained via the departmental Mass Spectrometry and NMR facilities run by Dr Lijiang Song and Dr Ivan Prokes, respectively. UV-visible analyses were carried out on a PerkinElmer Lambda 35 instrument at 25 °C. LC-MS analysis in 96-well plates (NUNC) were carried out on a Waters ZQ instrument using an Acquity UPLC CSH C18 column (50mm x 2.1mm i.d. 1.7 μm) column and the flow rate was 1.0 ml min⁻¹. Mass-directed autopreparation (MDAP) was carried out on a Waters QDA instrument using an Xselect CSH C18 column (150mm x 30mm i.d. 5 μm) column and the flow rate was 40 ml min⁻¹.

Polymerase Chain Reaction experiments were performed in either a Bio-Rad T100 thermocycler or an Eppendorf Mastercycler Gradient. DNA agarose gel electrophoresis was performed in a Bio-Rad Wide Mini-Sub Cell GT tank connected to a Bio-Rad Basic

PowerPac power supply. PCR products were visualised on UVP UV transilluminating imaging system. SDS-PAGE analyses were conducted in a Bio-Rad Mini-PROTEAN Tetra Vertical Electrophoresis Cell connected to a Bio-Rad Basic PowerPac power supply. DNA and protein concentration was measured using a Thermo Scientific NanoDrop Lite spectrometer. Cell growth and protein expression was conducted in New Brunswick Scientific Innova44 incubators and cells were lysed using a Constant Systems cell disruptor. Cell optical density was assessed using a Thermo BioMate3 spectrometer at 600 nm. Centrifugation was achieved using a Hitachi himac CR22N high-speed centrifuge equipped with either a R9A or R20A2 rotor.

7.1.3 Plasmids, bacterial strains and enzymes

Plasmid pSB13 containing the gene for *txtE* from *Streptomyces turgidiscabies* Car8 in pET-151 was previously prepared by Dr Sarah Barry. The gene encoding the reductase domain, *RhFRED*, from *Rhodococcus* sp. NCIMB 9784 was purchased from Epoch Life Science, TX, USA, as plasmid pUC-*rhfred*. The gene encoding TxtE-BM3R, codon-optimised for *E.coli*, plasmid pET28-(a)+-*txtE_bm3r*, was purchased from GenScript, NJ, USA.

Competent cells used in biological experiments were either Invitrogen One Shot TOP10 and BL21 Star (DE3) or New England Biolabs DH5 α and BL21(DE3). Restriction enzymes were purchased from either ThermoFisher Scientific or New England Biolabs.

7.1.4 Biological reagents and media

Miller's Luria-Bertani (LB) media was purchased from Fisher Scientific and prepared to 25 g/L in deionised water. Solid LB-agar media was prepared using 25 g/L LB broth and 15 g/L bacto agar in deionised water. Terrific broth (TB) was purchased from VWR and prepared by adding 50 g of powder to 990 mL deionised water and 10 mL glycerol prior to autoclaving. Ampicillin and kanamycin stock solutions were prepared to 100 mg/mL and 50 mg/L, respectively, and filter sterilised prior to being stored at -20° C until use. Isopropyl β -D-1-thiogalactopyranoside (IPTG) purchased from VWR was prepared as a 0.5 M stock solution in deionised water, filter sterilised and stored at -20° C until use.

7.2 Biological procedures

7.2.1 Preparation of pET151_ *txtE*_ *rhfred*

The vector, pCloneEZ-*RhFRED*, containing the reductase domain, the 32 amino acid polypeptide linker and 20 bp overlap regions was double digested with PvuII and AclI for 15 mins. The 1083 bp *RhFRED* fragment was separated from the vector by agarose gel electrophoresis, then excised from the resulting gel and purified using the GeneJET gel extraction kit (Thermo Scientific). Previously constructed pSB13 was digested with Eco53kI, and purified similarly to *RhFRED*, to yield the cleaved backbone. The gene fragment encoding RhFRED and linker was then inserted into pSB13 via Gibson assembly using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB), with incubation at 50 °C for 1 hour in a thermocycler, to afford pRS01 (Fig.7.1). The stop codon and the nine nucleotides between the end of the *txtE* gene and the Eco53KI restriction site were deleted using the Q5 Site-Directed Mutagenesis Kit (NEB) (table 7.1, 7.2 and 7.3) using forward primer 5' -TCCACTCATATGCGGCTGGC-3' and reverse primer 5' -GCGGAGGCTGAGCGGCAG-3' to give the correct construct pRS012. The sequence of the cloned gene was confirmed by DNA sequencing (GATC).

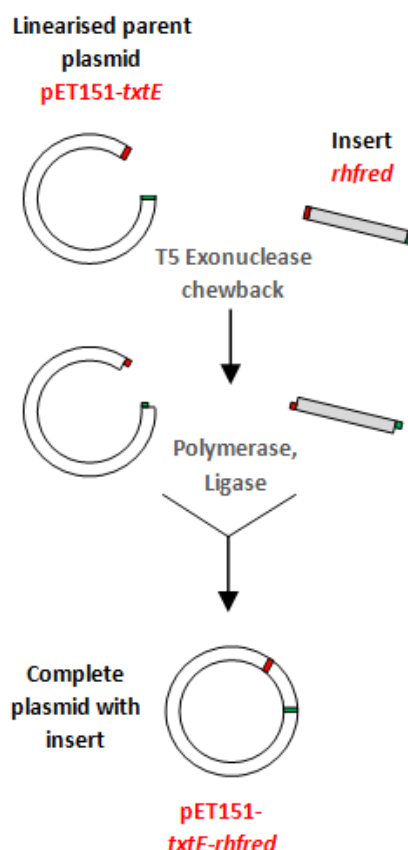


Figure 7.1. General protocol for NEBuilder HiFi DNA Assembly kit

	<i>Per 25 μL reaction (μL)</i>	<i>Final conc.</i>
Q5 Hot Start High Fidelity 2X master Mix	12.5	1X
10 μ M Forward Primer	1.25	0.5 μ M
10 μ M Reverse Primer	1.25	0.5 μ M
Template DNA (10-25 ng/ μ L)	1.0	10-25 ng
Nuclease-free water	9.0	-

Table 7.1. Components and quantities of the PCR reactions according to the Q5-Site-Directed Mutagenesis protocol (NEB)

<i>Step</i>	<i>Temp ($^{\circ}$C)</i>	<i>Time</i>
Initial denaturation	98	30 s
25 cycles	98	10 s
Final extension	71	30 s
	72	8 min 30 s
	72	2 min
Hold	4	∞

Table 7.2. PCR cycling conditions for the Q5-Site-Directed mutagenesis protocol (NEB). Reactions carried out in 0.2 mL flat top PCR tubes.

	<i>Volume (μL)</i>	<i>Final conc.</i>
PCR product	1	
2X KLD reaction buffer	5	1X
10X KLD enzyme mix	1	1X
Nuclease-free water	3	

Table 7.3. Components and quantities of KLD treatment mixture containing kinase, ligase, DpnI. Reactions incubated at room temperature for 5 mins in 0.2 mL PCR tubes.

<i>Protein</i>	<i>Molecular weight (Da)</i>	<i>Calculated ϵ (dm³ / mol .cm)</i>
TxtE	44751	46410
TxtE-RhFRED	84862	74830
TxtE-BM3R	111166	119180

Table 7.4. Calculated protein molecular weights and predicted extinction coefficients for TxtE, TxtE-RhFRED and TxtE-BM3R used for the calculation of protein yield after purification.

7.2.2 General procedure for protein overproduction

A general procedure for the expression of TxtE, TxtE-RhFRED and TxtE-BM3R is described as follows: the corresponding plasmid was used to transform *E.coli* BL21Star (DE3). A single transformant from antibiotic-supplemented LB-agar plates was used to inoculate LB medium (10 ml) supplemented with ampicillin (100 $\mu\text{g ml}^{-1}$) for TxtE-RhFRED and TxtE or kanamycin (50 $\mu\text{g/ml}$) for TxtE-BM3R and was incubated at 30 °C overnight with shaking at 220 rpm. This overnight culture (5-10 ml) was used to inoculate LB or TB medium (1 L) supplemented with antibiotic and was incubated at 37 °C with shaking at 220 rpm until the optical density at 600 nm reached 0.6-0.8. At this point the cultures were cooled before isopropyl- β -D-thiogalactopyranoside (IPTG) (0.1 mM), 5-aminolevulinic acid hydrochloride (5-ALA) (1 mM), iron (III) chloride (5 mg/L) and thiamine (1 mM) were added to the culture for expression of the N-terminal His₆-translational fusion protein. The culture was incubated overnight at 15 °C with shaking at 220 rpm. Cells from the overnight incubation were harvested by centrifugation (4000 x g, 30 min) and resuspended in 10 mL Buffer A (table 7.5.) before addition of phenylmethylsulphonyl fluoride (PMSF, 1 mM final conc.). The cells were lysed by cell disruption and cell debris was pelleted by centrifugation (11,000 x g, 30 min). The supernatant was further filtered through a 0.2 μm syringe filter before being applied to a 1 mL HiTrap™ HP affinity column (Nickel Sepharose High Performance, GE Healthcare) equilibrated with Buffer A. Unbound proteins were eluted with Buffer A (10 ml) and His₆-TxtE-variant was eluted using 5, 3, 3, and 3 ml of buffers B1-B4, respectively (table 7.5.). The orange fractions were pooled and concentrated to ~ 300 μl (Ultra centrifugal device 50 kDa membrane) before being diluted with Buffer C and concentrated again. This concentration procedure was repeated twice more before the purity of the collected

fractions was analysed by SDS-PAGE (10%), and flash frozen as aliquots in liquid nitrogen for storage at -80 °C until further use.

TxtE was purified similarly to above, except: 5-ALA, FeCl₃ and thiamine were not added to the expression cultures, and a 30 kDa membrane ultra centrifugal device was used for protein concentration.

<i>Process</i>	<i>Buffer</i>	<i>Base</i>	<i>NaCl</i>	<i>Imidazole</i>	<i>Glycerol (%)</i>
Lysis	A	Tris-HCl, pH8	100	20	0
Elution	B1	Tris-HCl, pH8	100	50	0
	B2	Tris-HCl, pH8	100	100	0
	B3	Tris-HCl, pH8	100	200	0
	B4	Tris-HCl, pH8	100	300	0
Concentration	C	Tris-HCl, pH8	100	0	10

Table 7.5. Components and quantities of buffers used during the purification of proteins

7.2.3 Protein analysis and characterisation

SDS-PAGE analysis of TxtE, TxtE-RhFRED and TxtE-BM3R was conducted using a 10% acrylamide/bis-acrylamide gel (Table 7.6). The loading gel was allowed to set for 30 minutes prior to the addition of the stacking gel plus comb. This was set for at least 30 minutes before conducting the analysis. The electrophoresis tank was filled with running buffer (2 M glycine, 1% SDS, 250 mM Tris-HCl, pH 8) and the set gel was appropriately placed inside. Protein samples (10-20 µL) were prepared firstly by the addition of loading dye (2-4 µL, 5X stock contains 250 mM Tris-HCl, pH 6.8, 10% SDS, 30% (v/v) Glycerol, 10 mM DTT, 0.05% (w/v) Bromophenol Blue). The samples were heated to 80 °C for 5 minutes before being applied to the SDS-PAGE gel lane alongside protein ladder (3 µL, Thermo Scientific Page Ruler Plus Prestained). The gel was run at 200 V for 35 minutes and visualised using InstantBlue Protein Stain (Expedeon).

<i>Component</i>	<i>10 % Loading gel (mL)</i>	<i>Stacking gel (mL)</i>
H ₂ O	4.0	1.4
30 % acrylamide/bis-acrylamide	3.3	0.33
1.5 M Tris-HCl, pH 8.8	2.5	-
1.0 M Tris-HCl, pH 6.8	-	0.25
10 % sodium dodecyl sulphate	0.1	0.02
10 % ammonium persulfate	0.1	0.02
TEMED	0.004	0.002

Table 7.6. Components and quantities of 10 % acrylamide/bis-acrylamide SDS-PAGE analysis gel. TEMED = tetramethylethylenediamine

The presence of heme-containing protein in the solution was confirmed by UV-visible spectroscopy. A solution of the purified protein (10-20 μ L) was added to a cuvette containing Tris-HCl (25 mM, pH 8.0) and a scan was run between 200 and 500 nm using a blank cuvette containing Tris-HCl (25 mM, pH 8.0) as a reference. Absorbance at 420 nm confirmed the presence of heme cofactor in the protein solution.

Total protein concentration was calculated using a Nanodrop Lite instrument with the absorbance units set to 1 mg/mL. Using the Beer-Lambert law, $A = \epsilon lc$, where A is the measured absorbance, l is the path length (1 cm), and ϵ is the predicted extinction coefficients, the protein concentration, c , was calculated for each protein prior to further experiments.

7.2.4 Overproduction of TxtE-BM3R in 96-well plates

In preparation for high-throughput screening of TxtE-BM3R mutants by LCMS analysis, TxtE-BM3R was overproduced in 96-deep well plates:

A single transformant was used to inoculate 100 mL of LB media containing 50 μ g/mL kanamycin. The culture was incubated at 30 °C with shaking at 220 rpm for 16 hours. 1 mL of Overnight Express Autoinduction media (Novagen) containing 50 μ g/mL kanamycin was aliquoted into columns 1-9 of several 96-deep well plates using a Multidrop Combi Reagent dispenser (Thermo Scientific) leaving columns 10-12 for buffer control reactions. To each well, 10 μ L of the LB culture was added via the Biomek FX liquid handling

system. The plates were covered with breathable membranes and incubated at 37° C with shaking at 800 rpm. Once the cultures had reached an $OD_{600nm} = 1.0-1.5$, the temperature was lowered to 20° C and incubation was continued for a further 20 hours. The plates were centrifuged (4000 x g, 20 mins) and the media discarded. The cell pellet containing 96 well plates were covered with a plastic seal and stored at -80° C until further use.

7.2.5 Generation of TxtE-BM3R mutant libraries

Each selected position in TxtE-BM3R was randomised in one step using the Q5-site directed mutagenesis protocol (New England Biolabs). The three mutagenic forward primers encoding NDT, VHG and TGG (table.7.7.) at the site of saturation were mixed in a ratio of 12:9:1, respectively, to a final concentration of 10 μ M. The PCR reaction mixture (50 μ L final volume) contained 25 μ L Q5 hot start polymerase, 2.5 μ L forward primer mixture (10 μ M), 2.5 μ L reverse primer (10 μ M), 2 μ L template DNA (10 ng/ μ L) and 18 μ L nuclease-free water. The PCR reaction mixture for each library (25 μ L) was aliquoted into 2 separate PCR tubes (12.5 each) except for the R59 mixture which was prepared to double the reaction scale (50 μ L) and split into 4 x 12.5 μ L reactions accounting for the greater number of annealing temperatures required (table 7.7.). The PCR program consisted of an initial denaturisation step at 98° C for 30 s, followed by 25 cycles of 98° C for 10 s, annealing at the given temperatures for 30 s, elongation at 72° C for 4 min 30 s, followed by a final elongation step at 72° C for 2 min. The PCR reaction mixtures were combined and 1 μ L was added to 5 μ L of KLD reaction buffer, 1 μ L of 10X KLD enzyme mix and 3 μ L nuclease-free water. The mixture was incubated for 5 min at room temperature before being transformed with 5-alpha competent *E.coli* (New England Biolabs). The next day the transformants were rinsed from the agar plate (50 μ g/mL kanamycin) using 1 mL LB broth and plasmid DNA was extracted using a DNA miniprep kit (Qiagen). Randomisation at the site of saturation was confirmed for each library by sequencing.

<i>Residue</i>	<i>Forward primers 5' - 3'</i>	<i>Reverse primer 5' - 3'</i>	<i>T_a</i>
M88	GAACGTGGTT ND TATACCGATC GAACGTGGTT VHG TATACCGATC GAACGTGGTT TGG TATACCGATC	TTGGTCCAACGGTGAAAG	61.0 59.0
M173	GAACACCATT ND TACCTACCACA GAACACCATT VHG ACCTACCACA GAACACCATT TGG ACCTACCACA	ACACGCGGGATCAGAAAG	64.8 62.0
I244	GCTGGTGTTC ND TGCGCTGTTTG GCTGGTGTTC VHG GGCGCTGTTTG GCTGGTGTTC TGG GGCGCTGTTTG	GCCAGTTGGTGAACGGTC	66.5 62.0
A245	GGTGTTCATT ND TCTGTTTGCGC GGTGTTCATT VHG CTGTTTGCGC GGTGTTCATT TGG CTGTTTGCGC	AGCGCCAGTTGGTGAACG	67.9 64.8
A248	TGCGCTGTT ND TCCGACCACCC TGCGCTGTT VHG CCGACCACCC TGCGCTGTT TGG CCGACCACCC	ATGAACACCAGCGCCAGTTGGTGAACG	69.0 64.8
F295	GAGCAACCAG ND TACCTGGCGTG GAGCAACCAG VHG ACCTGGCGTG GAGCAACCAG TGG ACCTGGCGTG	GCGTTGTAACGCAGCACC	69.0 64.8
W297	CCAGTTCACC ND TCGTGTGGCGG CCAGTTCACC VHG CGTGTGGCGG CCAGTTCACC TGG CGTGTGGCGG	TTGCTCGCGTTGTAACGCAGCAC	67.9 62.5
F395	AAACCTGGAG ND TCGTAGCCTGC AAACCTGGAG VHG CGTAGCCTGC AAACCTGGAG TGG CGTAGCCTGC	TCGTTCCAAATCGGCTCGC	68.6 62.0
R59	GACCGCGGAC ND TGGTACCGAGG GACCGCGGAC VHG GGTACCGAGG GACCGCGGAC TGG GGTACCGAGG	AGACGCGGGTCCTCAGACCCG	72.0 70.2 65.4 61.0
Y175	CATTATGACC ND TACAGCGGTC CATTATGACC VHG CACAGCGGTC CATTATGACC TGG CACAGCGGTC	GTGTTACACGCGGGATC	64.8 63.4
H176	TATGACCTAC ND TAGCGGTCCGA TATGACCTAC VHG AGCGGTCCGA TATGACCTAC TGG AGCGGTCCGA	ATGGTGTTCACACGCGGG	70.2 68.0
S177	GACCTACCAC ND TGGTCCGAAGG GACCTACCAC VHG GGTCCGAAGG GACCTACCAC TGG GGTCCGAAGG	ATAATGGTGTTCACACGC	70.2 61.0
G178	CTACCACAGC ND TCCGAAGGATC CTACCACAGC VHG CCGAAGGATC CTACCACAGC TGG CCGAAGGATC	GTCATAATGGTGTTCACAC	70.2 61.0
P179	CCACAGCGGT ND TAAAGGATCAGC CCACAGCGGT VHG AAGGATCAGC CCACAGCGGT TGG AAGGATCAGC	TAGGTCATAATGGTGTTCACACGCG	70.2 68.0
	CAGCGGTCCG ND TGATCAGCCGG	TGGTAGGTCATAATGGTGTTCACACGCG	72.0

K180	CAGCGGTCCG VH GGATCAGCCGG CAGCGGTCCG TGG GATCAGCCGG		70.2
D181	CGGTCCGAAG ND TAGCCGGTTA CGGTCCGAAG VH GAGCCGGTTA CGGTCCGAAG TGG CAGCCGGTTA	CTGTGGTAGGTCATAATGGTGTTCACAC	72.0 70.2
R34	CTGGGTGA ND TCTGAACGCGT CTGGGTGA VH GCTGAACGCGT CTGGGTGA TGG CTGAACGCGT	TGAACCGGGCGACGTTGC	72.0 68.0
E391	GATTTGGAAC ND TAACTGGAGTTTCGTAGCCTG GATTTGGAAC VH GAACCTGGAGTTTCGTAGCCTG GATTTGGAAC TGG AACCTGGAGTTTCGTAGCCTG	GGCTCGCCCGCGGTACGG	72.0 68.0
V63	TGGTACCGAG ND TCTGGCGGCGA TGGTACCGAG VH GCTGGCGGCGA TGGTACCGAG TGG CTGGCGGCGA	CGGTCCGCGGTGAGACGC	72.0 65.4
T56	CCCGCGTCTG ND TGCGGACCGTG CCCGCGTCTG VH GCGGACCGTG CCCGCGTCTG TGG GCGGACCGTG	TCCTTCAGACCCCGCGCAATC	72.0 65.4
K84	CCGTTGGAC ND TAACTGGTTA CCGTTGGAC VH GAACGTGGTTA CCGTTGGAC TGG AACGTGGTTA	TGAAAGATGTTGTCCGGCG	65.4 61.0
E394	CGAAAACCTG ND TTTCGTAGCC CGAAAACCTG VH GTTTCGTAGCC CGAAAACCTG TGG TTCGTAGCC	TTCCAAATCGGCTCGCCC	68.0 61.8
T296	CAACCAGTTC ND TGGCGTGTGGCGG CAACCAGTTC VH TGGCGTGTGGCGG CAACCAGTTC TGG TGGCGTGTGGCGG	CTCGCGTTGTAACGCAGC	68.0 65.4
Y89	CGTGGTTATG ND TACCGATCCGCCGC CGTGGTTATG VH GACCGATCCGCCGC CGTGGTTATG TGG ACCGATCCGCCGC	TTCTTGGTCCAACGGTGA	65.4 61.0
N293	CAACGCGAGC ND TAGTTCACCTGG CAACGCGAGC VH GAGTTCACCTGG CAACGCGAGC TGG CAGTTCACCTGG	TAACGCAGACCTCGTTC	65.4 61.8
L241	CCAAGTGGCG ND TGTGTTTCATTG CCAAGTGGCG VH GTGTTTCATTG CCAAGTGGCG TGG GTGTTTCATTG	TGAACGGTCTGCTCCGGG	72.0 68.0
T250	GTTTGCGCCG ND TACCCGGGTA GTTTGCGCCG VH GACCCGGGTA GTTTGCGCCG TGG ACCCGGGTA	AGCGCAATGAACACCAGC	72.0 68.0

Table 7.7. List of forward and reverse primers used for TxtE-BM3R site-saturation mutant library generation by the 22c-trick method. The annealing temperatures used for each PCR reaction are also noted.

<i>Amino acid</i>	<i>Encoding base</i>		
	1 st	2 nd	3 rd
NDT			
Ile	A	T	T
Asn	A	A	T
Ser	A	G	T
Gly	G	G	T
Asp	G	A	T
Val	G	T	T
Arg	C	G	T
His	C	A	T
Leu	C	T	T
Phe	T	T	T
Tyr	T	A	T
Cys	T	G	T
VHG			
Met	A	T	G
Thr	A	A	G
Lys	A	G	G
Glu	G	G	G
Ala	G	A	G
Val	G	T	G
Gln	C	G	G
Pro	C	A	G
Leu	C	T	G
TGG			
Trp	T	G	G

Table 7.8. Theoretical base distribution when using NDT, VHG and TGG and the mutagenic codons according to the 22c-trick method.

	First (%)	Second (%)	Third (%)
A	27.3	31.8	---
T	18.2	31.8	54.5
C	27.3	18.2	---
G	27.3	18.2	45.5

Table 7.9. Percentage distribution of bases in the first, second and third codon of the mutagenic region (NDT, VHG and TGG) of primers when using the 22c-trick method.

7.2.6 Preparation of TxtE-BM3R_R59C variants

The following three TxtE-BM3R variants were prepared using the Q5-site directed mutagenesis kit:

A) The linker region between TxtE(R59C) and the BM3R reductase domain was shortened from 27 residues to 14 residues by deletion of 39 bases using the forward primer 5'-ACCGAACAGAGCGCGAAGAAAG-3' and reverse primer 5'-ACGCAGGCTCAGCGGCAG-3'. An annealing temperature, T_a , of 71 °C was used.

B) Residue 176 in TxtE-BM3R was mutated from His to Trp using forward primer 5'-TATGACCTACTGGAGCGGTCCGAAGGATC-3' and reverse primer 5'-ATGGTGTTCACACGCGGG-3'. An annealing temperature, T_a , of 66 °C was used.

C) A stop codon was inserted between the TxtE(R59C) heme domain and the BM3R reductase domain primer 5'-GAGCCTGCGTTGAAAAAGCAAGAAAATCCCGC-3' and reverse primer 5'-AGCGGCAGGCTACGCAGG-3'. An annealing temperature, T_a , of 72 °C was used.

7.3. UV-vis characterisation of heme-containing proteins

7.3.1 UV-Vis spectroscopic analysis of TxtE

The TxtE resting state Fe-(III)-H₂O complex has a λ_{max} of 417 nm and was observed by simply adding TxtE (8 μ M) to a cuvette containing Tris-HCl (25 mM, pH 8.0). The Fe(II) complex was generated by adding the reducing agent sodium dithionite (5 μ l, 1 M) to the cuvette, where the λ_{max} was found to be blue-shifted to 404 nm. The TxtE Fe(III)-NO complex was detected by adding a solution of DEANO (5 μ l, 100 mM) to a fresh solution of the protein. The resulting mixture exhibited a λ_{max} of 431 nm.

7.3.2 Qualitative binding assay – determining protein-substrate binding type

The protein was added to a cuvette (A) containing Tris-HCl (25 mM, pH 8.0) to a concentration of 8 μ M. A spectrum was recorded between 300 nm and 500 nm to provide a reference. L-Trp or tryptamine was then added to the cuvette (B) to a final concentration of 2 mM. Spectrum A was subtracted from spectrum B to provide the difference spectrum.

7.3.3 General procedure for the determination of protein-substrate binding constants

In order to determine the binding constant of substrates to TxtE and variants, the following general procedure was performed:

A solution of the protein was prepared to 8 μ M in Tris-HCl (25 mM, pH 8.0). A solution of the substrate (0.5-1.0 μ l titrations, 10-50 mM stock) was added stepwise to the cuvette (final concentration 0-1 mM) and a spectrum was recorded. This spectrum was subtracted from a reference spectrum in which the corresponding volume of buffer was added. This procedure was repeated until the total volume of substrate solution added to the cuvette was no greater than 5 % of the volume of the protein solution. The experiment was conducted in at least duplicate for each substrate tested.

The absorbance at λ_{\min} was subtracted from the absorbance λ_{\max} and plotted against the concentration of substrate. The data was fitted to the Michaelis-Menten equation using OriginPro 9.1, and the dissociation constants were calculated from the resulting binding curves.

This general procedure was followed for UV-Visible analysis of TxtE-BM3R and TxtE-BM3R(R59C) with L-tryptophan and tryptamine as the substrates.

7.3.4 Qualitative binding assay of TxtE-BM3R and TxtE-BM3R_R59C with 2-(benzofuran-3-yl)ethanamine as the substrate

The protein, either TxtE-BM3R or TxtE-BM3R_R59C, was added to a cuvette (A) containing Tris-HCl (25 mM, pH 8.0) to a concentration of 8 μ M. A spectrum was recorded between 300 nm and 500 nm to provide a reference. A solution of 2-(benzofuran-3-yl)ethanamine (1 μ L, 20 mM) was then added to the cuvette (B) giving a final substrate concentration of 0.1 mM. Spectrum A was subtracted from spectrum B to provide the difference spectrum for the substrate with each protein.

7.4 In vitro assays, optimisation and screening methods

7.4.1 TxtE-catalysed nitration of L-Trp and substrate analogues

TxtE reaction uses ferredoxin and ferredoxin reductase redox partner proteins from *Spinachia oleracea*. In a 2 mL eppendorf tube 10 μ M enzyme in 25 mM Tris-HCl, pH 8.0, was incubated with 0.5 mM substrate, ferredoxin (2 μ L), ferredoxin reductase (2 μ L), 2.4 mM NADPH and 4.0 mM DEANO in a final volume of 100 μ L. The reaction was incubated at 20 °C for 4 hours before being quenched with methanol (100 μ L). The mixture was filtered using 0.2 μ M microspin filter tubes (Corning) and analysed by LC-MS (table.7.12). The appropriate control reactions were also conducted.

7.4.2 TxtE-reductase fusion protein catalysed nitration of L-tryptophan using purified protein

In general, to an 2 ml eppendorf tube containing L-Trp (0.5 mM) in Tris-HCl (25 mM, pH 8.0), TxtE-reductase variant (1.5 μ M), NADPH (1 mM) and 2-(*N,N*-diethylamino)-diazonolate 2-oxide (DEANO) (0.5 mM) were added. The reaction was incubated at room temperature for 2 hours before HCl (5 μ L, 3 M) was added to the solution to precipitate the protein. A reaction using boiled enzyme was also carried out as a control assay. The solution was filtered using 0.2 μ M microspin filter tubes (Corning) and analysed by LC-MS (table.7.10).

7.4.3 TxtE-BM3R reaction buffer and NADPH regeneration screen

A single 96-well plate containing TxtE-BM3R-expressing cell pellets was removed from the -80 °C freezer and defrosted at 20 °C for 30 minutes. Several different lysis buffers were prepared each containing 1mg/mL lysozyme, 1 mg/mL polymixin B and 0.5 U/mL benzonase nuclease diluted in either Tris-HCl at pH 7, 8, 9 or 10, or potassium phosphate buffer at pH 6, 7 or 8. Each buffer was added to at least 6 wells of the 96-well plate and incubated at 20 °C for 2 hours with shaking at 800 rpm. The plate was centrifuged (4000 x g, 10 mins) to remove cell debris then 240 μ L of each well solution was transferred to the corresponding well of a fresh 96-deep well plate. For each buffer type, 3 wells contained 300 μ L of 100 mM Tris-HCl (pH8), 0.1 mM NADP⁺, 3.5 mM glucose and 0.5 units glucose dehydrogenase and the final 3 wells contained NADPH (300 μ L, 4.0 mM). To each well of the plate a solution of 4-methyl-DL-tryptophan (21 μ L, 35 mM) was added followed by DEANO (21 μ L, 92 mM) to initiate the reaction, and incubated at 25 °C with shaking at 800 rpm for 20 hours. Once reaction was complete, 100 μ L of the mixture was transferred to a fresh 96-deep well plate containing 300 μ L MeOH in each well. The plate was shaken at 800 rpm for 10 minutes before being centrifuged to pellet precipitated

proteins. 200 μ L of the supernatant was transferred to a NUNC 96-shallow well plate for analysis by LCMS (table 7.10).

7.4.4 High-throughput screening and analysis of TxtE-BM3R activity assays with L-tryptophan substrate analogues

A single 96-well plate containing TxtE-BM3R-expressing cell pellets was removed from the -80°C freezer and defrosted at 20°C for 30 minutes. Lysis buffer (300 μ L) containing 1mg/mL lysozyme, 1 mg/mL polymixin B and 0.5 U/mL benzonase nuclease in Tris-HCl, pH 8, was added to each well and the plate was incubated at 20°C with shaking at 800 rpm for 2 hours. The plate was centrifuged (4000 x g, 10 mins) to remove cell debris then 240 μ L of each well solution was transferred to the corresponding well of a fresh 96-deep well plate containing 300 μ L of 100 mM Tris-HCl (pH8), 0.1 mM NADP⁺, 3.5 mM glucose and 0.5 units glucose dehydrogenase. A solution of substrate **26** (21 μ L) prepared as a 35 mM stock was added to 3 wells in columns 1-9 (TxtE-BM3R expressing wells) and one well in column 10-12 (control wells). The reaction was started by the addition of DEANO (21 μ L, 92 mM) and incubated at 25°C with shaking at 800 rpm for 20 hours. Once reaction was complete, 100 μ L of the mixture was transferred to a fresh 96-deep well plate containing 300 μ L MeOH in each well. The plate was shaken at 800 rpm for 10 minutes before being centrifuged to pellet precipitated proteins. 200 μ L of the supernatant was transferred to a NUNC 96-shallow well plate for analysis by LCMS (table 7.10).

<i>Time</i>	<i>% A (water + 0.1 % formic acid)</i>	<i>% B (acetonitrile + 0.1 % formic acid)</i>	<i>Flow rate (mL/min)</i>
0	97	3	1.000
3	85	15	1.000
4	5	95	1.000
5	97	3	1.000
6	97	3	1.000

Table 7.10. LCMS analysis conditions used during the optimisation and substrate screening experiments for TxtE-BM3R and variants.

7.4.5 Scale-up of the nitration of 4-methyl-DL-tryptophan by TxtE-BM3R using clarified cell lysate

The general protein overproduction procedure described in section 7.2.1 was followed. The cells were collected by centrifugation (4000 x g, 20 mins) and resuspended in 100 mM Tris-HCl, pH8. Cell paste yield varied between batches from 10 g/L to 15 g/L and were

resuspended in 4X w/v of the reaction buffer (i.e. 30 g cells in 120 mL buffer). This solution was passed through muslin cloth and lysed by microfluidisation three times. The resulting solution was centrifuged (11,000 x g, 30 mins) and 90 mL of the supernatant was transferred to a 100 mL EasyMax Chemical Reactor (Mettler Toledo). A solution of 4-methyl-DL-tryptophan (2 mL, 250 mM in 0.5 M NaOH) and NADPH (5 mL, 200 mM in reaction buffer) were added with stirring at 300 rpm. The mixture was incubated at 20 °C for 5 minutes before the reaction was initiated by the addition of 2-(N,N-Diethylamino)-diazonolate 2-oxide sodium salt hydrate (DEANO, 0.65mL, 1M in 10 mM NaOH). Stirring was continued for a further 20 hours. The reaction was quenched by the addition of 300 mL MeOH and filtered to remove precipitated proteins. The solution was concentrated to 80 mL *in vacuo* and freeze-dried. The resulting residue was resuspended in 5 mL water, 3 mL MeOH and 2 mL 0.1 M HCl before being purified by a Waters Mass-Directed Auto Preparation system (MDAP, table 7.11) using a mass detection parameter of 264.09. The pure collected fractions were concentrated *in vacuo* to give nitrated 4-methyl-DL-tryptophan.

<i>Time</i>	<i>% A (water + 0.1 % formic acid)</i>	<i>% B (methanol + 0.1 % formic acid)</i>	<i>Flow rate (mL/min)</i>
0	100	0	40
1	100	0	40
20	75	25	40
20.5	1	99	40
25	1	99	40

Table 7.11. Mass-Directed Auto Preparation conditions for the purification of 4-nitro-L-tryptophan from the scaled-up reaction in cell lysate

7.4.6 High-throughput screening and analysis of the TxtE-BM3R mutant libraries

For each TxtE-BM3R mutant library, the site-randomised DNA mixture was used to transform *E. coli* BL21 (DE3) competent cells (Invitrogen). For each individual well in columns 1-11 of a 96-deep well plate, 88 colonies from each transformation were used to inoculate 100 µL LB broth with kanamycin (50 µg/mL). In column 12, 3 wells containing the culture broth were inoculated with TxtE-BM3R, 3 wells were inoculated with *E. coli* cells transformed with pET28a(+) plasmid and 2 wells were left empty for buffer control reactions. Each plate was incubated at 30 °C overnight with shaking at 220 rpm. The next day 10 µL of each well culture was transferred to the corresponding well in a 96-deep well plate containing 1 mL Overnight Express Autoinduction media (Novagen) containing 50

µg/mL kanamycin. The plates were sealed and incubated at 37 °C with shaking at 800 rpm. Once the cultures had reached an OD_{600nm} = 1.0-1.5 the temperature was lowered to 20 °C and the plates were incubated for a further 20 hours. The plates were centrifuged (4000 x g) for 10 mins, the media was discarded and the plates containing cell pellets were stored at -80 °C until required for reaction screening.

The 96-deep well plates containing the cell pellets for each library were removed from the freezer and thawed at room temperature for 30 mins. Lysis buffer (300 µL) containing 1 mg/mL lysozyme, 1mg/mL polymixin B and 10 unit/mL benzonase nuclease in 25 mM Tris-HCl, pH 8.0 was added to each well and the plate was incubated at 20 °C for 2-3 hours with shaking at 800 rpm. When lysis was complete the plate was centrifuged (4000 x g) for 10 mins to remove cell debris and 240 µL of each well solution was transferred to the well of a fresh 96-well plate, each containing 330 µL 25 mM Tris-HCl, pH 8.0, with 1.2 mM tryptamine, 0.1 mM NADP⁺, 3.5 mM glucose and 0.5 units glucose dehydrogenase. The reaction was started by the addition of 2-(N,N-Diethylamino)-diazene 2-oxide sodium salt hydrate (DEANO) to a final concentration of 2.4 mM. The plate was sealed and incubated at 25 °C with shaking at 800 rpm for 20 hrs. The reaction was terminated by the addition of 1 mL methanol and the plate was shaken at 800 rpm at 20 °C for 10 mins before being centrifuged (4000 x g). 200 µL of the supernatant was transferred to a 96-shallow well plate for analysis by LCMS (table.7.10).

7.4.7 In vitro activity assays of TxtE-BM3R(R59C) with L-tryptophan and tryptamine

To 2 ml eppendorf tubes containing either tryptamine or L-tryptophan (0.5 - 2.0 mM) in Tris-HCl (25 mM, pH 8.0), TxtE(R59C) (1.5 µM), NADPH (1 mM) and 2-(N,N-diethylamino)-diazene 2-oxide (DEANO) (0.5 mM) were added. The reaction was incubated at room temperature for 2 hours before HCl (5 µl, 3 M) was added to the solution to precipitate the protein. A reaction using boiled enzyme was also carried out as a control assay for each substrate. The solutions were filtered using 0.2 µm microspin filter tubes (Corning) and analysed by LC-MS (table.7.12).

7.4.8 In vitro activity assays with TxtE(R59C) and redox partner proteins

As 7.4.1 except TxtE is replaced with TxtE(R59C)

7.4.9 In vitro screening and analysis of TxtE-BM3R(R59C) with tryptamine substrate analogues

The ability of TxtE-BM3R_R59C to nitrate several tryptamine analogues was tested by preparing the following reactions in 2 mL eppendorf tubes: 10 μ M enzyme in 25 mM Tris-HCl, pH 8.0, containing 2.0 mM substrate x-x, 2.4 mM NADPH and 4.0 mM DEANO in a final volume of 100 μ L. The reaction was incubated at 20° C for 4 hours before being quenched using 100 μ L methanol. The mixture was filtered using 0.2 μ M microspin filter tubes (Corning) and analysed by LC-MS (table.7.12).

<i>Time</i>	<i>% A (water + 0.1 % formic acid)</i>	<i>% B (methanol + 0.1 % formic acid)</i>	<i>Flow rate (mL/min)</i>
0	100	0	1.000
5	100	0	1.000
30	0	100	1.000
35	0	100	1.000
40	100	0	1.000

Table 7.12. LCMS analysis conditions for the screening of tryptamine substrate analogues with TxtE-BM3R_R59C variant

7.4.10 Scale-up of the tryptamine nitration reaction with purified TxtE-BM3R_R59C protein

The general protein overproduction procedure described in section 7.2.1 was followed. Cells from the overnight incubation were harvested by centrifugation (4000 x g, 30 min) and resuspended in 30 mL Buffer A (table 7.5) before addition of phenylmethylsulphonyl fluoride (PMSF, 1 mM final conc.). The cells were lysed by cell disruption and cell debris was pelleted by centrifugation (11,000 x g, 30 min). The supernatant was further filtered through a 0.2 μ m syringe filter before being applied to a 5 mL HiTrap™ HP affinity column (Nickel Sepharose High Performance, GE Healthcare) equilibrated with Buffer A. Unbound proteins were eluted with Buffer A (30 ml) and His₆-TxtE-variant was eluted using 15, 10, 10, and 10 ml of buffers B1-B4, respectively. The orange fractions were pooled and concentrated to 1 ml (Ultra centrifugal device 50 kDa membrane) before being diluted with 20 mL Reaction Buffer (25 mM Tris-HCl, pH 8 and 100 mM NaCl) and concentrated again. Concentration was repeated twice more to ensure complete removal of imidazole from the protein solution. The TxtE-BM3R_R59C protein solution (1 mL, 48 mg/mL) was diluted with a further 12 mL Reaction Buffer (RB) in a 50 mL falcon tube, followed by the addition of tryptamine (0.6 mL, 50 mM in 10% DMSO in reaction buffer) and NADPH (0.72 mL, 50 mM in RB). The mixture was stirred at 20° C for 5 mins before DEANO (0.6 mL, 0.1 M in 10 mM NaOH) was added to start the reaction and the tube was sealed. After 20 hours the reaction was quenched with 10 % HCl in MeOH (15 mL),

stirred for 30 mins and filtered to remove precipitated proteins. The filtrate was concentrated *in vacuo* and freeze-dried to remove the remaining liquid. The residue was dissolved in 2 mL H₂O and purified by HPLC (table. 7.13) to give a pure nitrotryptamine isomer (1.1 mg, 18 %).

Time (min)	% A (water + 0.1 % formic acid)	% B (methanol + 0.1 % formic acid)	Flow rate (mL/min)
0	100	0	9.000
5	100	0	9.000
34	0	100	9.000
38	0	100	9.000
40	100	0	9.000

Table 7.13. HPLC elution conditions used for the purification of nitrotryptamine from the scaled-up reaction using purified TxtE-BM3R_R59C protein.

7.5 Computational methods

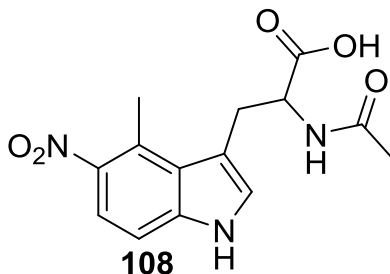
7.5.1 Molecular docking studies

A homology model incorporating the R59C mutation in TxtE was created using SWISS-MODEL where the wild-type crystal structure with L-tryptophan bound (PDB: 4TPO) was used as a template. The heme cofactor was modelled into the resulting structure by aligning with the wild-type structure in PYMOL to give the final model in which the substrate could be docked. Substrate geometry optimisation and force field minimisation was performed using Avogadro. Tryptamine was set to 3 rotational bonds around the ethylamine chain in AutoDock Vina and molecular docking simulations were carried out in a cubic grid box of 10 angstroms in each axis, centered over the heme iron atom. The obtained binding modes were analysed in PYMOL.

7.6 Chemical synthesis

7.61 Synthesis of 4-methyl-5-nitro-L-tryptophan

N-acetyl-4-methyl-5-nitro-D,L-tryptophan **108**



Synthesis was adapted from a previously published procedure. L-serine (21 mg, 0.2 mmol) was added to a solution of 4-methyl-5-nitro-1H-indole (17.6 mg, 0.1 mmol) in acetic acid (0.24 mL) and acetic anhydride (0.9 mmol). The reaction mixture was stirred for 2 h at 73 °C before being cooled to room temperature. The solution was diluted with diethyl ether (2 mL) and adjusted to pH 11 using 30 % sodium hydroxide solution. The solution was further diluted with diethyl ether (3 mL) before the layers were separated. The ether layer was further extracted with 30 % sodium hydroxide (3 x 2 mL). The aqueous solutions were combined and neutralised to pH 7 using concentrated HCl before being concentrated *in vacuo*. The ether layer was dried over magnesium sulphate and the solvent removed *in vacuo* to recover the unreacted indole. The crude *N*-acetylated product **2** was purified by HPLC (19 mg, 0.063 mmol, 63 %) using a C₁₈ column (Agilent, Zorbax C-18,) using the method in table 7.14.

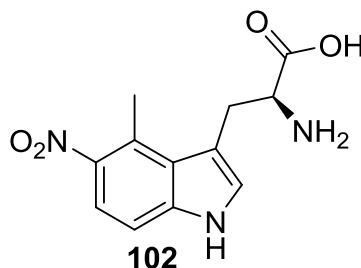
Time (min)	% A (water + 0.1 % formic acid)	% B (methanol + 0.1 % formic acid)	Flow rate (mL/min)
0	100	0	9.000
5	100	0	9.000
24	0	100	9.000
28	0	100	9.000
30	100	0	9.000

Table 7.14. HPLC eleution conditions used for the purification of *N*-acetyl-4-methyl-5-nitro-D,L-tryptophan

¹H NMR (400 MHz, D₂O) δ: 1.81 (s, 3H), 2.78 (s, 3H), 3.10 (m, 1H), 3.48 (dd, 1H, *J* = 15 Hz, *J* = 5.0 Hz), 4.45 (dd, 1H, *J* = 9.0 Hz, *J* = 5.2 Hz), 7.20 (s, 1H), 7.28 (d, 1H, *J* = 9.0 Hz), 7.67 (d, 1H, *J* = 9.0 Hz)

m/z (ESI): calculated [C₁₄H₁₄N₃O₅]: 304.0939 Found: 304.0942

4-methyl-5-nitro-L-tryptophan 102

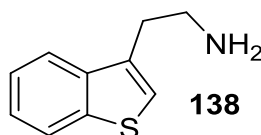


Purified compound **2** (15 mg, 0.05 mmol) was dissolved in borate buffer solution (50 mM, 0.75 mL) and a solution of acylase (12.5 mg, Amano Inc.) in borate buffer (1.2 mL) was added. Cobalt (II) chloride hexahydrate was added to a final concentration of 0.125 mM. The solution was incubated at 37 °C with shaking at 180 rpm for 24 hrs. The reaction was quenched by the addition of 10 % hydrochloric acid until the solution was pH 5, and filtered to remove precipitated proteins. The filtrate was washed with ethyl acetate (3 x 3 mL) and the aqueous layer was purified by ion-exchange chromatography using DOWEX[®] Resin (H⁺ form, 50W X8, 200-400 mesh). The resin was prepared by firstly washing with deionised water before being activated using hydrochloric acid (2 M). The resin was washed again until the flow-through was pH 5. The reaction mixture was applied to the resin column and washed with water (100 mL). The product was eluted with 10 % ammonia in methanol (20 mL) and fractions which showed UV activity and responded positively to a ninhydrin test were pooled. The solvent was removed *in vacuo* to afford the L-enantiomer of the disubstituted tryptophan **3** (3 mg, 0.011 mmol, 22 %).

¹H NMR (400 MHz, D₂O) δ : 2.80 (s, 3H), 3.12 (dd, 1H, $J = 15.0$ Hz, $J = 9.5$ Hz), 3.52 (dd, 1H, $J = 15.0$ Hz, $J = 5.5$ Hz), 3.74 (dd, 1H, $J = 9.5$ Hz, $J = 5.5$ Hz), 7.29 (s, 1H), 7.32 (d, 1H, $J = 9.0$ Hz), 7.71 (d, 1H, $J = 9.0$ Hz)

m/z (ESI⁺): calculated [C₁₂H₁₄N₃O₄]⁺: 264.0979 Found: 264.0970

7.6.2 2-(benzo[b]thiophen-3-yl)ethanamine 138



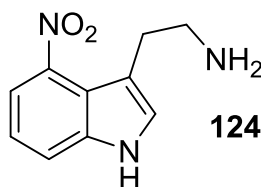
To a stirred solution of 1 % 2-cyclohexan-1-one in cyclohexanol (0.6 mL), 3-(benzothien-3-yl)-D-alanine (90 mg, 0.41 mmol) was added and the suspension was heated to 154 °C under reflux. After 2 hours the reaction was cooled to room temperature and diluted with chloroform (1 mL). The dark brown solution was passed through a silica pad and then washed with water (2 x 2 mL). The organic layer was concentrated *in vacuo* to give the desired product as a pale brown solid (22.6 mg, 0.13 mmol, 32 %).

¹H NMR (400 MHz, MeOD) δ: 2.96 (s, 2H), 3.23 (s, 2H), 7.22 (s, 1H), 7.29 (m, 2H), 7.74 (s, 1H), 7.79 (d, 1H)

¹³C NMR (125 MHz, D₂O) δ: 32.30, 42.20, 122.81, 123.79, 124.04, 125.27, 125.59

m/z (ESI⁺): calculated [C₁₀H₁₂NS]⁺: 178.0690 Found: 178.0689

7.6.3 4-nitrotryptamine 124



Method adapted from Ref: 153. To a stirred solution of 1 % 2-cyclohexan-1-one in cyclohexanol (0.6 mL), 4-nitro-L-tryptophan (41 mg, 0.16 mmol) was added and the suspension was heated to 154 °C under reflux. After 2 hours the reaction was cooled to room temperature and diluted with chloroform (1 mL). The solution was passed through a silica pad and washed with water (2 x 2 mL). The organic layer was concentrated *in vacuo* and further diluted with methanol (1 mL) before being purified by C18-reversed phase HPLC (table 7.15) to give the 4-nitrotryptamine product (7 mg, 0.034 mmol, 21 %).

¹H NMR (400 MHz, D₂O) δ: 3.19 (m, 2H), 3.26 (m, 2H), 7.32 (t, 1H, *J* = 8.0 Hz), 7.56 (s, 1H), 7.87 (d, 1H, *J* = 8.0 Hz), 8.01 (d, 1H, *J* = 8.0 Hz)

¹³C NMR (125 MHz, D₂O) δ: 25.18, 40.85, 108.46, 118.10, 118.47, 119.32, 121.77, 130.37, 139.38, 141.39

m/z (ESI⁺): calculated [C₁₀H₁₂N₃O₂]⁺: 206.0924 Found: 206.0924

Time (min)	% A (<i>water + 0.1 % formic acid</i>)	% B (<i>methanol + 0.1 % formic acid</i>)	Flow rate (mL/min)
0	100	5	15.000
5	100	5	15.000
10	40	60	15.000
45	20	80	15.000
47	5	95	15.000
50	5	95	15.000
52	95	5	15.000

Table 7.15. HPLC eleution conditions used for the purification of synthetic nitro standard 124.

References:

1. A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, *Nat. Rev. Drug Discov.*, 2015, **14**, 111-129.
2. D. J. Newman, G. M. Cragg, *J. Nat. Prod.*, 2007, **70**, 461-477.
3. C. D. Campbell, J. C. Vederas, *Biopolymers*, 2010, **93**, 755-763.
4. V. De Luca, A. J. Cutler, *Plant Physiol.*, 1987, **85**, 1099-1102.
5. W. P. Hammes, F. C. Neuhaus, *Antimicrob. Agents. Chemother.*, 1974, **6**, 722-728.
6. J. Staunton, K. J. Weissman, *Nat. Prod. Rep.*, 2001, **18**, 380-416.
7. D. Schwarzer, R. Finking, M. A. Marahiel, *Nat. Prod. Rep.*, 2003, **20**, 275-287.
8. T. Robinson, *Science*, 1974, **184**, 430-435.
9. M. Baunach, J. Franke, C. Hertweck, *Angew. Chem. Int. Ed.*, 2014, **54**, 2604-2626.
10. K-S. Ju, R. E. Parales, *Microbiol. Mol. Biol. Rev.*, 2010, **74**, 250-272.
11. R. Winkler, C. Hertweck, *ChemBioChem*, 2007, **8**, 973-977.
12. R. M. Smith, D. A. Joslyn, O. M. Gruhzit, I. W. McLean, Jr, M. A. Penner, J. Ehrlich, *J. Bacteriol.*, 1948, **55**, 425-448.
13. F. Washizu, H. Umezawa, N. Sugiyama, *J. Antibiot. Ser. A*, 1954, **7**, 60.
14. E. Rössner, A. Zeeck, W. König, *Angew. Chem.*, 1990, **102**, 84-85.
15. K. Arima, H. Imanaka, M. Kousaka, A. Fukuda, G. Tamura, *Agric. Biol. Chem.*, 1964, **28**, 575-576.
16. G. T. Carter, J. A. Nietzsche, J. J. Goodman, M. J. Torrey, T. S. Dunne, D. B. Borders, R. T. Testa, *J. Antibiotics*, 1987, **40**, 233-236.
17. X.-k. Wang, Y.-r. Zhao, T.-f. Zhao, S. Lai, C.-t. Che, *Phytochemistry*, 1993, **35**, 263.
18. R. R. King, C. H. Lawrence, M. C. Clark, L. A. Calhoun, *J. Chem. Soc. Chem. Commun.*, 1989, 849-850.
19. R. Raju, A. M. Piggott, L. X. Barrientos Diaz, Z. Khalil, R. J. Capon, *Org. Lett.*, 2010, **12**, 5158-5160.
20. C. Carter, W. McChesney, *Nature*, 1949, **164**, 575-576.
21. K. Doxtader, M. Alexander, *J. Bacteriol.*, 1966, **91**, 1186-1191.
22. M. Vanhaelen, R. Vanhaelen-Fastre, P. But, J.L. Vanherweghem, *Lancet*, 1994, **343**, 174.

23. M. Lang, P. Spiteller, V. Hellwig, W. Steglich, *Angew. Chem.*, 2001, **40**, 1704-1705.
24. R. R. Manam, S. Teisan, D. J. White, B. Nicholson, J. Grodberg, S. T. C. Neuteboom, K. S. Lam, D. A. Mosca, G. K. Lloyd, B. C. M. Potts, *J. Nat. Prod.*, 2005, **68**, 240-243.
25. S. Kawai, G. Kawabata, A. Kobayashi, K. Kawazu, *Agric. Biol. Chem.*, 1989, **53**, 1127-1133.
26. S. M. Barry, G. L. Challis, *ACS Catal.*, 2013, **3**, 2362-2370.
27. J. Lee, H. Zhao, *Angew. Chem. Int. Ed.*, 2006, **45**, 622-625.
28. R. Winkler, C. Hertweck, *Angew. Chem. Int. Ed.*, 2005, **44**, 4083-4090.
29. M. Simurdiak, J. Lee, H. Zhao, *ChemBioChem*, 2006, **7**, 1169-1172.
30. S. V. Kryatov, E. V. Rybak-Akimova, S. Schindler *Chem. Rev.*, 2005, **105**, 2175-2226.
31. G. T. Carter, J. A. Nietsche, J. J. Goodman, M. J. Torrey, T. S. Dunne, M. M. Siegel, D. B. Borders, *J. Chem. Soc. Chem. Commun.*, 1989, 1271-1273.
32. S. M. Barry, J. A. Kers, E. G. Johnson, L. Song, P. R. Aston, B. Patel, S. B. Krasnoff, B. R. Crane, D. M. Gibson, R. Loria, G. L. Challis, *Nat. Chem. Biol.*, 2012, **8**, 814-816.
33. J. D. Maya, B. K. Cassels, P. Iturriaga-Vásquez, J. Ferreira, M. Faúndez, N. Galanti, A. Ferreira, A. Morello, *Comparative Biochemistry and Physiology, Part A*, 2007, **146**, 601-620.
34. W. Brumfitt, J. M. T. Hamilton-Miller, *J. Antimicrob. Chemother.*, 1998, **42**, 363- 371.
35. B. S. Hall, S. R. Wilkinson, *Antimicrob. Agents Chemother.*, 2012, **56**, 115-123.
36. J. P. O'Keefe, K. A. Troc, K. D. Thomson, *Antimicrob. Agents Chemother.*, 1982, **22**, 426-430.
37. A. A. Voak, V. Gobalakrishnapillai, K. Seifert, E. Balczo, L. Hu, B. S. Hall, S. R. Wilkinson, *J. Biol. Chem.*, 2013, **288**, 28466-28476.
38. S. R. Wilkinson, M. C. Taylor, D. Horn, J. M. Kelly, I. Cheeseman, *Proc. Natl. Acad. Sci. U.S.A.*, 2008, **105**, 5022-5027.
39. H. Pennemann, S. Forster, J. Kinkel, V. Hessel, H. Löwe, L. Wu, *Org. Process Res. Dev.*, 2005, **9**, 188 -192.
40. H. H. Bieber, W. F. Schurig, *Ind. Eng. Chem.*, 1957, **49**, 832-837.
41. W. E. Noland, L. R. Smith, D. C. Johnson, *J. Org. Chem.*, 1963, **28**, 2262-2266.

42. J. B. Hester, Jr. *J. Org. Chem.*, 1964, **29**, 1158-1160.
43. R. M. Roberts, H. P. Browder Jr., K. A. Kobe, *J. Am. Chem. Soc.*, 1959, **81**, 1165–1167.
44. X. Zhang, S. Li, *Nat. Prod. Rep.*, 2017, **34**, 1061-1089.
45. J Ma, Z. Wang , H. Huang , M. Luo, D. Zuo, B. Wang, A. Sun, Y.C. Cheng, C. Zhang, J. Ju, *Angew. Chem., Int. Ed. Engl.*, 2011, **50**, 7797-802.
46. C. T. Walsh, H. Chen, T. A. Keating, B. K. Hubbard, H. C. Losey, L. Luo, C. G. Marshall, D. A. Miller, H. M. Patel, *Curr. Opin. Chem. Biol.*, 2001, **5**, 525-534.
47. S. Uhlmann, R. D. Süssmuth, M. J. Cryle, *ACS Chem. Biol.*, 2013, **8**, 2586–2596.
48. K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch, M. J. Cryle, *Nature*, 2015, **521**, 105-111.
49. B. Lauer, R. Russwurm, C. Bormann, *Eur. J. Biochem.*, 2000, **267**, 1698–170.
50. J. D. Rudolf, C.Y. Chang, M. Ma, B. Shen, *Nat. Prod. Rep.*, 2017, **34**, 1141-1172.
51. M. Katagiri, B. N. Ganguli, I. C. Gunsalus, *J. Biol. Chem.*, 1968, **243**, 3543-3546.
52. A. D. N. Vaz, D. F. McGinnity, M. J. Coon, *Proc. Natl. Acad. Sci. U.S.A.*, 1998, **95**, 3555–3560.
53. M. Sono, M. P. Roach, E. D. Coulter, J. H. Dawson, *Chem. Rev.*, 1996, **96**, 2841-2887.
54. R. Fasan, *ACS Catal.*, 2012, **2**, 647–666.
55. D. W Nebert, D. W Russell, *Lancet*, 2002, **360**, 1155-1162.
56. T. L. Poulos, B. C. Finzel, G. C. Wagner, J. Kraut, *J. Biol. Chem.*, 1985, **260**, 16122-16130.
57. Y. Xue, D. Wilson, L. Zhao, D. H. Sherman, *Chem. Biol.*, 1998, **5**, 661-667.
58. M. V. Mendes, N. Anton, J. F. Martin, J. F. Aparicio, *Biochem., J.*, 2005, **386**, 57–62.
59. W. Ding, W. Deng, M.-C. Tang, Q. Zhang, G.-L. Tang, Y.-R. Bi, W. Liu, *Mol. BioSyst.*, 2010, **6**, 1071–1081.
60. Y. Anzai, S. Li, M. R. Chaulagain, K. Kinoshita, F. Kato, J. Montgomery, D. H. Sherman, *Chem. Biol.*, 2008, **15**, 950-959.
61. J. C. Carlson, S.-Y. Li, S. S. Gunatilleke, Y. Anzai, D. A. Burr, L. M. Podust, D.H. Sherman, *Nat. Chem.*, 2011, **3**, 628–633.

62. C. J. Schulze, W. M. Bray, F. Loganzo, M-H. Lam, T. Szal, A. Villalobos, F. E. Koehn, R. G. Linington, *J. Nat. Prod.*, 2014, **77** , 2570–2574.
63. L. Duan, G. Jogl, D. E. Cane, *J. Am. Chem. Soc.*, 2016, **138**, 12678–12689.
64. M. Makino, H. Sugimoto, Y. Shiro, S. Asamizu, H. Onaka, S. Nagano, *Proc. Natl. Acad. Sci. U.S.A.*, 2007 104, 11591-11596.
65. Q. Cheng, D. C. Lamb, S. L. Kelly, L. Lei, F. P. Guengerich, *J. Am. Chem. Soc.*, 2010, **132**, 15173–15175.
66. Y. Xie, Q. Li, Y. Song, J. Ma, J. Ju, *ChemBioChem*, 2014, **15**, 1183–1189.
67. A.W. Munro, H. M. Girvan, A. E. Mason, A. J. Dunford, K. J. McLean, *Trends Biochem. Sci.*, 2013, **38**, 140-150.
68. A. W. Munro, H. M. Girvan, K. J. McLean, *Biochim. Biophys. Acta*, 2007, **1770**, 345– 359.
69. A. W. Munro, H. M. Girvan, K. J. McLean, *Nat. Prod. Rep.*, 2007, **24**, 585–609.
70. K.-Y. Choi, E. O. Jung, D.-H. Jung, B. P. Pandey, H. Yun, H.-Y. Park, R. J. Kazlauskas, B.-G. Kim, *FEBS J.*, 2012, **279**, 1650–1662.
71. L. O. Narhi, A. J. Fulco, *J. Biol. Chem.* 1986, **261**, 7160-7169.
72. M. A. Noble, C. S. Miles, S. K. Chapman, D. A. Lysek, A. C. MacKay, G. A. Reid, R. P. Hanzlik, A. W. Munro, *J. Biol. Chem.* 1999, **262**, 6683– 6690.
73. A. W. Munro, S. Daff, J. R. Coggins, J. G. Lindsay, S. K. Chapman, *Eur. J. Biochem.* 1996, **239**, 403–409.
74. G. A. Roberts, G. Grogan, A. Greter, S. L. Flitsch, N. J Turner, *J. Bacteriol.*, 2002, **184**, 3898–3908
75. E. O'Reilly, M. Corbett, S. Hussain, P. P. Kelly, D. Richardson, S. L. Flitsch, N. J. Turner, *Catal. Sci. Technol.*, 2013, **3**, 1490-1492.
76. V. R. Dodhia, A. Fantuzzi, G. Gilardi, *J. Biol. Inorg. Chem.*, 2006, **11**, 903–916.
77. D. Degregorio, S. D'Avino, S. Castrignanò, G. Di Nardo, S. J. Sadeghi, G. Catucci, G. Gilardi, *Front. Pharmacol.* 2017, **8**, 1-13.
78. S. Li, L. M. Podust, D. H. Sherman, *J. Am. Chem. Soc.* 2007, **129**, 12940-12941.
79. W. Zhang, Y. Liu, J. Yan, S. Cao, F. Bai, Y. Yang, S. Huang, L. Yao, Y. Anzai, F. Kato, L. M. Podust, D. H. Sherman, S. Li, *J. Am. Chem. Soc.* 2014, **136**, 3640–3646.
80. M. Nodate, M. Kubota, N. Misawa, *Appl. Microbiol. Biotechnol.*, 2006, **71**, 455–462.

81. D. Harris, G. Loew, L. Waskell, *J. Am. Chem. Soc.*, 1998, **120**, 4308-4318.
82. I. G. Denisov, T. M. Makris, S. G. Sligar, I. Schlichting, *Chem. Rev.*, 2005, **105**, 2253-2277.
83. H. Sezutsu, G. Le Goff and R. Feyereisen, *Philos. Trans. R. Soc., B*, 2013, **368**, 20120428.
84. C. A. Hasemann, R. G. Kurumbail, S. S. Boddupalli, J. A. Peterson, J. Deisenhofer, *Structure*, 1995, **3**, 41-62.
85. Y. Yoshigae, U. M. Kent, P. F. Hollenberg, *Biochemistry*, 2013, **52**, 4636–4647.
86. K. G. Ravichandran, S. S. Boddupalli, C. A. Hasemann, J. A. Peterson and J. Deisenhofer, *Science*, 1993, **261**, 731–736.
87. E. E. Scott, Y. A. He, M. R. Wester, M. A. White, C. C. Chin, J. R. Halpert, E. F. Johnson, C. D. Stout, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 13196–13201.
88. H. Li, T. L. Poulos, *Nat. Struct. Biol.*, 1997, **4**, 140–146.
89. T. L. Poulos, *Proc. Natl. Acad. Sci. U.S.A.*, 2003, **100**, 13121-13122.
90. B. Zhao, F. P. Guengerich, A. Bellamine, D. C. Lamb, M. Izumikawa, L. Lei, L. M. Podust, M. Sundaramoorthy, J. A. Kalaitzis, L. M. Reddy, S. L. Kelly, B. S. Moore, D. Stec, M. Voehler, J. R. Falck, T. Shimada, M. R. Waterman, *J. Biol. Chem.*, 2005, **280**, 11599–11607.
91. D. H. Sherman, S. Li, L. V. Yermalitskaya, Y. Kim, J. A. Smith, M. R. Waterman, L. M. Podust, *J. Biol. Chem.*, 2006, **281**, 26289–26297.
92. B. A. Fry and R. Loria, *Physiol. Mol. Plant. Pathol.* 2002, **60**, 1-8.
93. R. R King, C. H. Lawrence, L. A. Calhoun, *L.A. Agric. Food. Chem.* 1992, **40**, 834–837.
94. J. A. Kers, K. D. Cameron, M. V. Joshi, R. A. Bukhalid, J. E. Morello, M. J. Wach, D. M. Gibson and R. Loria, *Mol. Microbiol.* 2005, **55**, 1025-1033.
95. J. A. Kers, M. J. Wach, S. B. Krasnoff, J. Widom, K. D. Cameron, R. A. Bukhalid, D. M. Gibson, B. R. Crane and R. Loria, *Nature*, 2004, **429**, 79- 82.
96. W. K. Alderton, C. E. Cooper and R. G. Knowles, *Biochem. J.* 2001, **357**, 593–615.
97. F. G. Healy, M. Wach, S. B. Krasnoff, D. M. Gibson and R. Loria, *Mol. Microbiol.* 2000, **38**, 794-804.
98. M. Strieker, A. Tanović, M. A. Marahiel, *Curr. Opin. Chem. Biol.*, 2010, **20**, 234–240.
99. F. G. Healy, S. B. Krasnoff, M. Wach, D. M. Gibson, R. Loria, *J. Bacteriol.* 2002, **184**, 2019–2029.

100. S. M. Barry, J. A Kers, E. G. Johnson, L. Song, P. R. Aston, B. Patel, S. B. Krasnoff, B. R. Crane, D. M. Gibson, R. Loria, G. L. Challis, *Nat. Chem. Biol.* 2012, **8**, 814-816.
101. F. Yu, M. Li, C. Xu, Z. Wang, H. Zhou, M. Yang, Y. Chen, L. Tang, J. He, *PLoS ONE*, 2013, **8**, e81526
102. S. C. Dodani, J. K. B. Cahn, T. Heinisch, S. Brinkmann-Chen, J. A. McIntosh, F. H. Arnold, *ChemBioChem.*, 2014, **15**, 2259 – 2267.
103. G. Zocher, M. E. A. Richter, U. Mueller and C. Hertweck, *J. Am. Chem. Soc.*, 2011, **133**, 2292-2302.
104. D. Leys, C. G. Mowat, K. J. McLean, A. Richmond, S. K. Chapman, M. D. Walkinshaw, A. W. Munro, *J. Biol. Chem.*, 2003, **278**, 5141-5147.
105. J. G. McCoy, H. D. Johnson, S. Singh, C. A. Bingham, I-K. Lei, J. S. Thorson, G. N. Phillips Jr, *Proteins: Struct.,Funct.Bioinf.*, 2009, **74**, 50-60.
106. K. G. Ravichandran, S. S Boddupalli, C. A Hasermann, J. A Peterson, J. Deisenhofer, *Science*, 1993, **261**, 731-736.
107. S. C. Dodani, G. Kiss, J. K. B. Cahn, Y. Su1, V. S. Pande and F. H. Arnold, *Nat. Chem.* 2016, **8**, 419-425.
108. H. Tomita, Y. Katsuyama, H. Minami, Y. Ohnishi, *J. Biol. Chem.*, 2017, **292**, 15859-15869
109. J. D. Bloom, F. H. Arnold, *Proc.Natl. Acad. Sci.U.S.A.*, 2009, **106**, 9995–10000.
110. J. D Bloom, M. M Meyer, P. Meinhold, C. R Otey, D. MacMillan, F. H Arnold, *Curr. Opin. Chem. Biol*, 2005, **15**, 447–452.
111. E. O. McCullum, B. A. Williams, J. Zhang, J. C. Chaput, *Methods Mol Biol.*, 2010, **634**,103-109.
112. S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult, *Chem. Biol.*, 2005, **12**, 45-54.
113. E. T. Farinas, U. Schwaneberg, A. Glieder, F. H. Arnold, *Adv. Synth. Catal.* 2001, **343**, 601-606
114. T. Kubo, M. W. Peters, P. Meinhold, F. H. Arnold., *Chem. –Eur. J.*, 2006, **12**, 1216-1220.
115. M. Landwehr, L. Hochrein, C. R. Otey, A. Kasrayan, Jan-E. Bäckvall, F. H. Arnold, *J. Am. Chem. Soc.*, 2006, **128**, 6058-6059
116. J. C. Lewis, S. Bastian, C. S. Bennett, Y. Fu, Y. Mitsuda, M. M. Chen, W. A. Greenberg, C. H. Wong, F. H. Arnold. *Proc. Natl. Acad. Sci.*, 2009, **106**, 16550-16555.
117. R. Singh, M. Bordeaux, R. Fasan, *ACS Catal.*, 2014, **4**, 546–552.

118. X. Ren, J. A. O'Hanlon, M. Morris, J. Robertson, L. L. Wong, *ACS Catal.*, 2016, **6**, 6833–6837.
119. P. S. Coelho, E. M. Brustad, A. Kannan, F. H. Arnold, *Science*, 2013, **339**, 307–310.
120. A. Seifert, S. Vomund, K. Grohmann, S. Kriening, V. B. Urlacher, S. Laschat, J. Pleiss, *ChemBioChem*, 2009, **10**, 853 – 861.
121. S. M. Barry and G. L. Challis, *Methods enzymol.* 2012, **516**, 171–194.
122. R. Radi, *Acc. Chem. Res.*, 2013, **46**, 550–559.
123. A. Robin, G. A. Roberts, J. Kisch, F. Sabbadin, G. Grogan, N. Bruce, N. J. Turner, S L. Flitsch, *Chem. Commun.*, 2009, 2478–2480.
124. D. G. Gibson, L. Young, R. Y. Chuang, J. C. Venter, C. A. Hutchison, H. O. Smith, *Nat. Methods*. 2009, **6**, 343–345.
125. R. Zuo, Y. Zhang, J. C. Huguet-Tapia, M. Mehta, E. Dedic, S. D. Bruner, R. Loria, Y. Ding, *Biotechnol. J.* 2016, **11**, 624–632.
126. Y. Lu, L. Mei, *J Ind. Microbiol. Biotechnol.* 2007, **34**, 47–53.
127. G. Blaser, J. M. Sanderson, A. S. Batsanov, J. A. K. Howard, *Tetrahedron Lett.*, 2008, **49**, 2795–2798.
128. E. G. Johnson, S. B. Krasnoff, D. R. D. Bignell, W-C. Chung, T. Tao, R. J. Parry, R. Loria, D. M. Gibson, *Mol. Microbiol.*, 2009, **73**, 409–418.
128. J. A. McIntosh, C. C. Farwell, F. H Arnold, *Curr. Opin. Chem. Biol.*, 2014, **19**, 126–134.
129. M. T. Reetz, D. Kahakeaw, R. Lohmer, *ChemBioChem.*, 2008, **9**, 1797–1804.
130. M. T Reetz, J. D. Carballeira, *Nat. Protoc.*, 2007, **2**, 891–903.
131. M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya, M. Bocola, *J. Am. Chem. Soc.*, 2010, **132**, 9144–9152.
132. R. M. P. Siloto, R. J. Weselake, *Biocatal. Agric. Biotechnol.*, 2012, **1**, 181–189.
133. M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chem. Int. Ed.*, 2005, **44**, 4192 –4196.
134. D. J. Bougioukou, S. Kille, A. Taglieber, M. T. Reetz, *Adv. Synth. Catal.*, 2009, **351**, 3287–3305.
135. S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z-G. Zhang, D. J. Opperman, M. T. Reetz, J. P Acevedo, *ACS Synth. Biol.* 2013, **2**, 83–92.
136. Q. D. Bosley, M. Ostermeier, *Biomol. Eng.*, 2005, **22**, 57–61.
137. W. M. Patrick, A. E. Firth, J. M. Blackburn, *Protein Eng.*, **16**, 451–457.

138. M. El-Sayed, R. Verpoorte, *Phytochem. Rev.*, 2007, **6**, 277–305.
139. S. N. Young, M. Leyton, *Pharmacol. Biochem. Behav.*, 2002, **71**, 857–865.
140. A. Brzezinski, *N. Engl. J. Med.*, 1997, **336**, 186–195.
141. J. Fricke, F. Blei, D. Hoffmeister, *Angew. Chem. Int. Ed.*, 2017, **56**, 12352–12355.
142. T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, *Nucleic Acids Res.*, 2003, **31**, 3381–3385.
143. Oleg Trott and Arthur J. Olson, *J. Comput. Chem.*, 2010, **31**, 455–461.
144. M. F. Dunn, D. Niks, H. Ngo, T. R.M. Barends, I. Schlichting, *Trends Biochem. Sci.*, 2008, **33**, 254–264.
145. R. J. M. Goss, P. L. A. Newill, *Chem. Commun.*, 2006, 4924–4925.
146. J. Murciano-Calles, D. K. Romney, S. Brinkmann-Chen, A. R. Buller, F. H. Arnold. *Angew. Chem.*, 2016, **55**, 11577–11581.
147. D. K. Romney, J. Murciano-Calles, J. E. Wehrmüller, F. H. Arnold, *J. Am. Chem. Soc.*, 2017, **139**, 10769–10776.
148. J. T. Payne, C. B. Poor, J. C. Lewis, *Angew. Chem. Int. Ed. Engl.*, 2015, **54**, 4226–4230.
149. T. Lyttle, D. Goldstein, J. Gartz, *J. Psychoact. Drugs*, 1996, **28**, 267–290.
150. J. Liu, T. Ng, Z. Rui, O. Ad, W. Zhang, *Angew. Chem. Int. Ed.*, 2014, **53**, 136–139.
151. R. S. Keri, K. Chand, S. Budagumpi, S. Balappa Somappa, S. A Patil, B. M. Nagaraja, *Eur. J. Med. Chem.*, 2017, **138**, 1002–1033.
152. C. M. Harris, J. S. Roberson, T. M. Harris, *J. Am. Chem. Soc.*, 1976, **98**, 5380–5386.
153. H. Mitsunori, E. Yutaka, O. Yasutomo, I. Toahiaki, A. Seiichi, *Chem. Lett.*, 1986, **15**, 893–896.
154. R. Zuo, Y. Zhang, J. C. Hackett, R. Loria, S. D Bruner, Y. Ding, *Sci. Rep.*, 2017, **842**, 1–9.

Appendix:

TxtE-BM3R mutant library analysis

I: Screening plate maps

All following plate maps monitoring for 4-NO₂-Trp conversion:

M88X

Conversion NRS301-31												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0	0.0	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.0
B	0.3	0.5	0.3	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.3	9.2
C	0.0	0.2	0.0	0.2	0.0	0.0	0.0	0.0	0.4	0.3	0.0	7.7
D	0.0	0.0	0.0	6.3	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0
E	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0
F	0.0	0.0	0.3	0.0	0.0	0.2	0.2	0.2	0.0	0.2	0.0	0.0
G	0.0	0.0	0.1	0.0	0.0	0.2	0.0	0.2	0.0	0.4	0.0	0.0
H	0.0	0.0	0.0	0.4	0.0	0.0	0.2	0.0	0.0	0.3	0.0	0.0

T296X

Conversion NRS301-32												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.1	0.6	2.0	1.0	1.2	0.7	14.3	0.2	0.8	0.5	0.7	14.2
B	1.3	1.3	0.5	0.3	0.3	1.3	5.0	0.5	0.6	1.1	1.0	16.1
C	16.3	0.7	0.7	0.0	0.5	9.0	0.3	0.3	13.8	0.0	0.4	16.2
D	0.4	0.4	0.0	1.5	0.5	1.3	0.3	0.3	0.0	1.7	1.7	0.4
E	0.0	0.4	0.5	0.7	0.9	1.0	8.5	0.0	0.0	0.3	0.7	0.3
F	0.2	0.5	0.3	0.6	0.2	0.3	0.0	0.5	0.3	0.5	0.3	0.3
G	1.7	0.2	1.2	0.2	0.9	0.3	0.0	20.8	0.7	0.7	0.3	0.0
H	0.4	0.3	0.3	0.5	0.2	0.4	13.3	0.5	0.4	0.0	0.4	0.0

W297

Conversion NRS301-33													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	1.3	8.1	7.8	16.2	0.0	3.1	14.1	0.2	3.7	0.2	0.5	14.9	12
B	4.8	8.6	0.0	3.7	0.3	8.0	2.5	6.1	1.4	4.3	6.8	15.8	24
C	6.8	9.9	2.4	5.8	2.3	2.9	4.0	0.3	2.7	1.0	3.5	11.5	36
D	2.9	6.3	9.7	8.8	14.8	0.8	7.4	10.9	8.0	3.5	0.3	0.3	48
E	2.1	0.2	1.7	0.4	9.8	0.9	12.0	2.3	4.6	0.3	3.4	0.3	60
F	1.6	1.4	3.0	0.4	0.0	3.9	0.4	2.5	4.0	10.0	7.5	0.1	72
G	4.7	0.2	0.2	0.5	6.6	5.3	0.2	8.6	0.4	9.0	2.2	0.1	84
H	0.2	3.6	3.9	0.6	9.6	10.0	4.3	0.0	4.0	4.6	3.8	0.1	96

F295

Conversion NRS301-34													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	0.4	2.1	8.9	1.6	0.0	1.1	5.7	2.1	2.0	18.5	4.8	13.6	12
B	0.2	4.6	0.0	20.1	4.9	1.7	1.2	1.2	0.9	0.5	0.8	16.1	24
C	0.5	0.5	0.2	0.5	0.3	1.8	0.4	4.9	1.8	0.5	0.0	11.5	36
D	0.0	0.1	0.0	0.3	1.4	6.7	5.4	0.7	0.0	1.8	0.5	0.2	48
E	0.2	0.3	8.2	0.5	0.6	0.7	0.4	1.8	0.1	0.5	0.8	0.1	60
F	3.1	0.2	0.1	1.2	2.0	1.0	1.5	0.1	0.8	1.7	0.5	0.1	72
G	0.0	0.1	0.0	4.7	0.1	0.8	0.2	0.2	5.8	0.4	0.8	0.0	84
H	1.0	0.2	0.3	1.8	1.7	0.1	1.1	0.1	8.1	1.5	0.9	0.1	96

M173

Conversion NRS301-35												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.5	0.1	4.4	0.1	0.3	0.1	0.1	0.1	0.3	0.4	4.3	3.8
B	0.3	0.2	0.2	0.5	0.1	0.0	1.1	0.4	0.7	1.2	0.4	3.4
C	0.0	0.8	0.3	4.1	0.2	0.2	0.1	0.3	0.2	0.3	0.1	2.2
D	0.1	0.3	0.1	3.8	0.4	0.4	0.4	0.5	4.2	0.3	0.4	0.1
E	0.2	0.2	0.0	0.2	0.9	0.1	0.8	0.1	0.3	0.3	1.2	0.3
F	0.2	0.2	4.1	0.1	0.3	3.9	0.2	0.2	0.3	0.8	0.2	0.2
G	0.1	0.4	0.2	0.6	0.3	0.3	0.3	0.8	0.3	0.1	0.2	0.2
H	0.3	0.2	0.2	0.1	0.4	2.8	0.7	0.2	0.2	0.3	0.4	0.2

A245

Conversion NRS301-36												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.2	0.5	0.1	0.4	0.2	0.2	0.1	0.4	0.4	0.1	0.2	4.3
B	0.2	0.1	0.3	0.2	0.2	0.3	0.5	0.2	0.3	0.0	1.0	2.7
C	4.0	0.2	4.0	5.2	0.8	2.2	0.4	0.4	0.3	0.0	0.8	3.6
D	0.3	0.2	0.0	0.5	0.2	0.2	0.5	0.2	0.2	0.1	0.4	0.2
E	0.3	0.4	0.5	0.0	0.1	0.1	0.8	0.1	0.1	0.2	0.2	0.2
F	0.3	0.1	0.3	0.3	0.3	0.0	0.3	4.7	0.0	3.3	0.3	0.2
G	0.5	3.4	0.3	0.2	0.2	0.3	0.2	0.2	0.1	0.3	0.3	0.1
H	0.1	3.8	0.2	0.3	0.3	0.6	0.3	1.5	0.4	0.0	0.3	0.0

I244

NRS301-43												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0	16.7	12.5	29.1	1.3	1.0	4.3	0.0	26.6	7.4	21.3	23.3
B	13.2	7.3	5.9	0.0	0.0	26.9	4.4	15.5	0.0	0.0	0.6	7.5
C	0.6	8.6	1.2	20.6	6.8	22.4	16.7	12.6	3.6	26.0	0.0	23.8
D	0.0	9.4	2.6	0.0	8.5	25.9	0.0	0.6	0.0	24.2	5.1	0.0
E	1.2	19.7	13.9	2.0	0.0	0.3	0.2	8.7	0.0	3.5	30.9	0.0
F	0.0	0.0	1.5	10.4	28.9	13.3	26.0	3.4	0.0	31.2	0.4	0.0
G	0.0	0.0	16.3	9.1	9.8	6.2	4.8	17.1	12.1	20.7	0.0	0.0
H	0.0	8.0	0.0	1.5	7.0	6.7	13.4	2.1	2.6	0.0	4.1	0.0

Y89

NRS301-44												
	1	2	3	4	5	6	7	8	9	10	11	12
A	1.9	3.4	2.0	0.0	0.0	1.6	0.0	0.0	0.0	0.0	1.4	16.1
B	2.5	4.6	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.9
C	0.7	0.0	0.0	1.7	0.0	0.0	0.0	5.8	3.0	1.5	0.0	16.6
D	0.0	0.0	1.6	0.0	0.0	1.5	1.0	3.7	0.0	0.0	0.0	0.0
E	0.0	0.0	0.9	0.0	0.9	1.8	0.0	0.0	0.0	1.7	0.0	0.0
F	1.8	0.0	0.0	1.2	0.0	0.0	3.4	0.0	0.0	1.5	0.8	2.0
G	1.5	0.0	0.0	0.0	0.0	0.0	1.0	1.9	0.0	1.3	0.0	0.0
H	0.0	0.0	0.0	0.0	0.0	0.0	1.4	0.0	1.3	0.0	1.7	0.0

R34

NRS301-46												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0	16.6	0.0	0.0	21.4	18.4	0.0	19.1	0.0	17.2	23.3	20.2
B	13.8	19.2	19.4	24.0	19.5	24.9	19.9	29.0	9.9	21.7	16.0	23.9
C	0.0	17.8	0.0	0.0	0.0	22.6	0.0	0.0	67.6	0.0	17.0	14.4
D	17.8	0.0	0.0	0.0	24.8	14.9	20.3	18.4	17.3	26.5	15.1	0.0
E	29.7	16.9	13.8	18.0	0.0	0.0	0.0	23.4	0.0	6.6	18.8	0.0
F	18.2	17.9	0.0	21.1	0.0	20.8	22.6	27.5	0.0	19.4	16.5	0.0
G	30.6	0.0	0.0	22.5	19.5	0.0	20.5	30.5	17.9	0.0	10.1	2.6
H	0.0	15.0	22.5	0.0	2.5	41.3	0.0	28.9	0.0	0.0	22.4	2.9

T56

NRS301-47												
	1	2	3	4	5	6	7	8	9	10	11	12
A	12.2	8.4	0.0	10.3	0.0	5.5	5.0	10.1	14.6	3.4	0.0	10.4
B	15.9	6.1	6.9	0.0	4.1	11.8	0.0	0.0	0.0	25.2	11.8	22.3
C	0.0	6.9	4.7	12.1	0.0	5.5	6.3	8.1	0.0	0.0	14.7	21.2
D	0.0	9.7	7.6	0.0	10.7	10.2	0.0	0.0	15.6	6.0	0.0	0.0
E	6.8	4.3	8.7	3.4	8.6	4.9	4.0	18.4	4.9	14.3	12.5	0.0
F	10.8	0.0	9.2	0.0	11.7	0.0	0.0	8.7	0.0	16.2	18.0	0.0
G	7.7	4.7	5.1	3.9	0.0	13.0	7.9	3.7	10.4	0.0	8.9	0.0
H	12.3	0.0	0.0	7.7	0.0	17.5	0.0	7.1	6.3	16.3	16.1	3.1

N293

NRS301-48												
	1	2	3	4	5	6	7	8	9	10	11	12
A	3.8	0.0	0.0	3.4	5.9	0.0	5.9	0.0	4.7	0.0	0.0	31.6
B	0.0	0.0	0.0	0.0	0.0	5.0	0.0	0.0	0.0	0.0	0.0	27.0
C	0.0	2.4	0.0	4.2	0.0	4.5	0.0	0.0	8.7	0.0	0.0	19.1
D	0.0	0.0	0.0	4.3	3.3	0.0	0.0	0.0	0.0	4.2	0.0	0.0
E	13.4	0.0	9.6	0.0	1.6	3.0	4.0	0.0	0.0	0.0	0.0	0.0
F	0.0	0.0	0.0	2.3	0.0	0.0	0.0	8.0	0.0	5.4	1.8	0.0
G	0.0	0.0	0.0	3.4	4.0	5.8	0.0	0.0	3.0	0.0	2.0	3.5
H	0.0	0.0	3.7	3.9	0.0	6.3	0.0	0.0	0.0	4.7	4.6	0.0

E391

NRS301-51												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0	3.3	0.0	1.7	0.0	0.0	5.6	6.7	0.0	5.5	7.8	8.5
B	2.6	0.0	0.0	5.9	0.0	4.8	0.0	0.0	5.5	4.0	2.4	2.8
C	5.2	4.6	2.8	0.0	0.0	4.3	3.1	0.0	5.3	5.6	1.6	3.1
D	0.0	0.0	7.3	3.3	2.7	4.0	0.0	5.1	0.0	0.0	5.7	0.0
E	0.0	1.9	0.0	0.0	3.5	3.2	0.0	0.0	4.9	3.8	2.5	0.0
F	0.0	0.0	1.2	0.0	0.0	2.9	0.0	2.0	4.2	3.9	4.9	0.0
G	2.5	3.9	0.0	4.6	3.9	0.0	10.0	2.2	5.1	4.8	5.1	0.0
H	0.0	2.3	2.4	5.1	4.6	0.0	0.0	3.0	6.3	5.9	0.0	0.0

V63

NRS301-52												
	1	2	3	4	5	6	7	8	9	10	11	12
A	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5
B	0.0	0.0	0.0	0.0	0.0	2.1	0.0	0.0	0.0	0.0	0.0	2.9
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.8
D	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
E	0.0	0.0	0.0	0.0	0.0	0.0	9.1	0.0	0.0	0.0	0.0	0.0
F	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.9	0.0	0.0
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
H	0.0	0.0	0.0	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

K84

NRS301-53													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	0.0	0.0	2.0	5.0	3.0	7.6	5.1	5.0	5.3	1.7	5.5	5.4	12
B	0.0	3.9	1.8	2.9	1.0	0.0	5.9	4.0	0.0	6.9	3.8	8.0	24
C	4.2	2.0	3.7	5.3	0.0	0.0	4.3	0.0	0.0	4.8	0.0	3.1	36
D	0.0	0.0	5.2	4.9	0.0	0.0	0.0	3.8	0.0	0.0	0.0	0.0	48
E	3.9	0.0	0.0	0.0	0.0	0.0	2.1	5.4	3.1	4.4	5.7	0.0	60
F	0.0	0.0	0.0	2.6	6.3	7.5	0.0	0.0	3.3	7.1	4.0	0.0	72
G	0.0	0.0	0.0	3.3	0.0	0.0	3.1	3.8	0.0	0.0	2.3	0.0	84
H	0.0	0.0	1.9	0.0	4.5	4.8	5.0	5.5	0.0	2.1	0.0	0.0	96

L241

NRS301-54												
	1	2	3	4	5	6	7	8	9	10	11	12
A	3.8	3.0	0.0	0.0	0.0	0.0	0.0	6.9	1.1	0.0	0.0	8.9
B	0.0	3.8	2.1	2.8	0.0	3.2	0.0	0.0	0.0	0.0	0.0	5.1
C	0.0	1.9	0.0	0.0	0.0	0.0	0.0	6.6	0.0	1.7	0.0	0.0
D	0.0	0.0	0.0	0.0	0.0	0.0	1.8	0.0	0.0	0.0	0.0	0.0
E	4.1	0.0	4.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0
F	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
G	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.9	0.0	0.0
H	0.0	3.6	0.0	3.8	1.5	0.0	0.0	0.0	0.0	0.6	0.0	0.0

Y175

Conversion NRS301-37													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	0.4	0.4	3.9	0.3	0.3	2.6	1.1	0.4	0.2	0.4	0.8	4.3	12
B	0.0	0.3	0.4	1.0	5.1	0.6	0.2	0.6	0.1	0.0	0.2	4.0	24
C	0.1	0.2	0.2	0.7	0.5	0.1	0.6	1.2	0.2	0.9	0.0	4.0	36
D	3.1	2.9	0.2	0.4	0.4	0.2	1.0	0.5	2.9	3.5	6.1	0.1	48
E	0.2	1.5	1.1	0.2	0.2	0.3	2.6	1.5	0.2	0.2	0.3	0.1	60
F	0.7	0.1	0.3	1.1	0.2	3.5	0.7	1.0	2.5	0.2	0.3	0.1	72
G	0.4	0.7	0.5	0.3	0.5	0.6	0.5	0.8	3.9	2.6	0.2	0.1	84
H	0.1	0.4	0.4	0.2	0.5	0.5	0.1	0.0	0.0	0.1	0.0	0.2	96

G178

Conversion NRS301-39													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	0.3	0.6	0.3	3.7	0.3	4.4	0.4	0.5	0.4	3.6	0.2	3.9	12
B	0.8	2.9	2.9	0.8	0.3	0.4	0.3	0.4	0.6	0.5	0.6	4.3	24
C	3.6	0.2	0.2	0.6	3.2	0.5	6.0	3.5	4.3	0.4	0.7	3.7	36
D	0.6	0.3	0.8	0.4	0.3	0.1	0.3	0.3	0.2	0.6	0.6	0.2	48
E	0.8	0.4	0.4	0.4	0.3	2.9	2.1	0.6	2.2	0.8	0.4	0.4	60
F	1.1	0.5	0.8	3.0	0.4	0.3	3.4	0.7	8.0	0.5	0.8	0.2	72
G	0.4	0.7	0.4	0.5	0.4	0.5	2.5	3.1	0.5	2.0	2.3	0.2	84
H	0.3	0.2	0.4	0.7	1.3	1.3	3.4	0.5	0.5	3.5	2.6	0.1	96

P179

Conversion NRS301-40													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	2.6	2.2	0.2	0.0	2.5	0.9	0.2	1.5	0.0	1.5	0.6	5.6	12
B	0.2	0.3	0.5	0.5	1.4	0.6	0.5	2.7	0.1	1.0	0.2	4.5	24
C	0.2	0.7	0.9	1.4	0.2	0.5	1.3	0.4	1.3	0.1	0.6	0.4	36
D	0.5	0.3	0.5	1.8	0.1	0.2	1.6	0.2	0.5	0.1	0.7	0.2	48
E	2.8	0.4	0.5	3.7	1.3	0.4	1.3	0.1	0.6	0.2	0.3	0.5	60
F	0.4	1.9	0.6	0.4	1.5	0.5	1.4	1.0	0.5	0.1	0.2	0.2	72
G	0.3	0.6	0.2	0.2	0.2	0.3	0.1	0.4	0.4	1.6	1.3	0.1	84
H	0.3	2.0	0.2	0.4	0.4	0.7	1.3	1.8	2.0	0.8	1.8	0.1	96

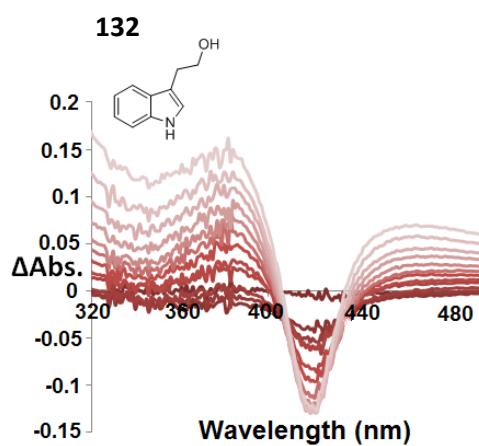
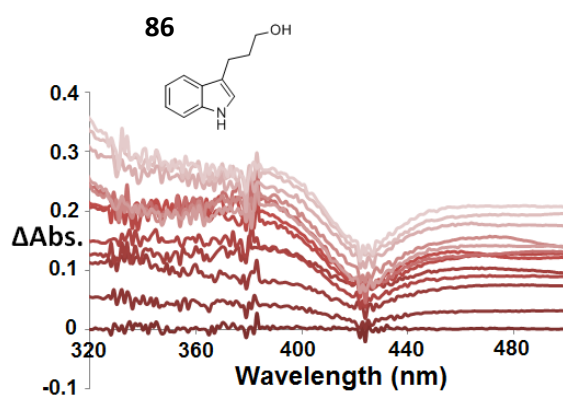
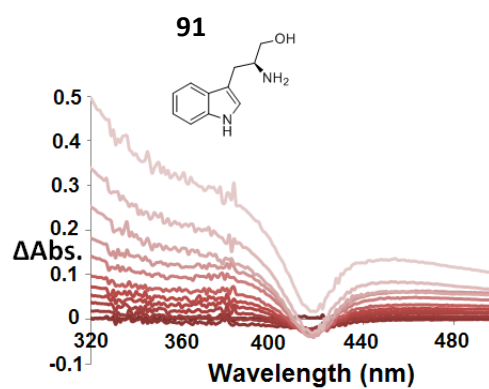
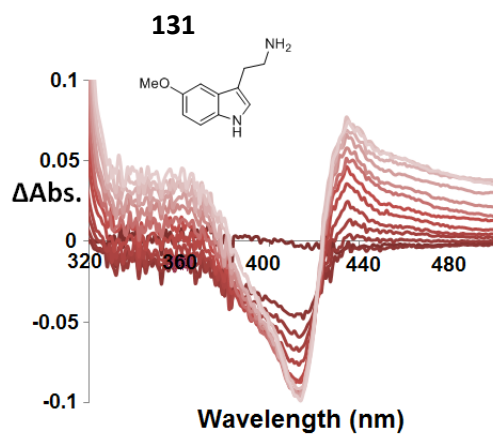
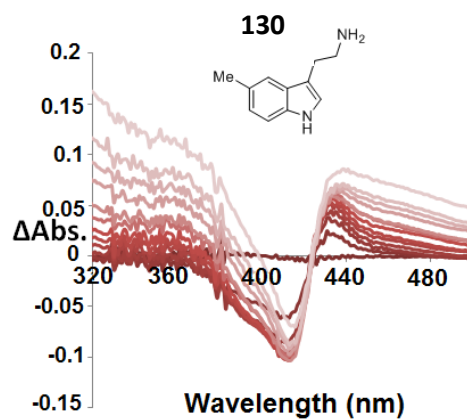
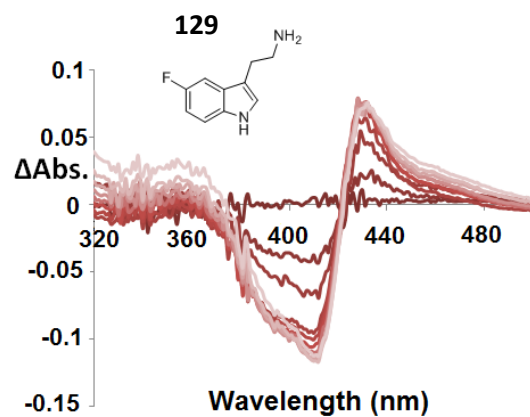
K180

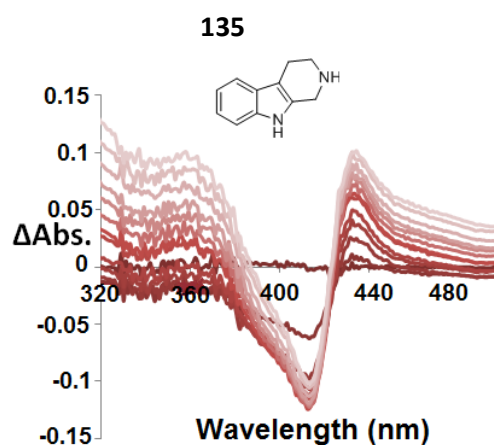
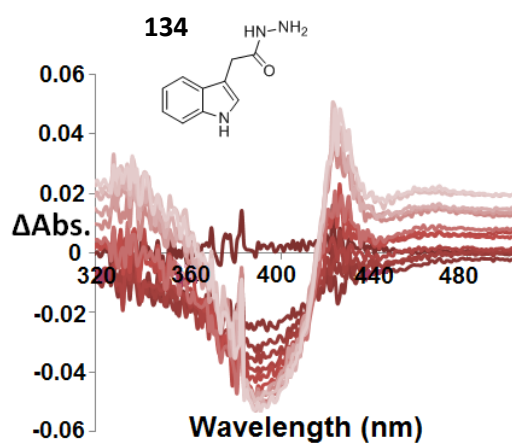
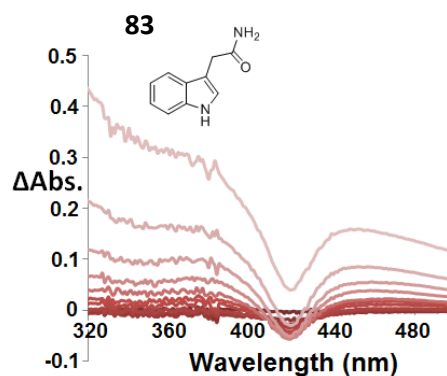
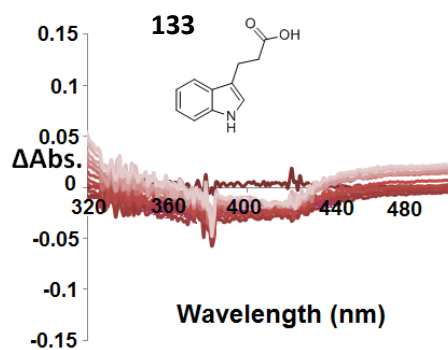
Conversion NRS301-41													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	4.8	0.2	4.9	0.3	0.3	0.9	4.6	2.9	2.8	4.8	3.9	5.6	12
B	4.4	0.4	4.0	0.1	3.8	0.2	2.3	3.4	0.8	0.2	0.3	5.0	24
C	0.2	0.3	0.4	0.2	0.5	0.3	1.3	3.9	0.3	0.6	3.4	4.5	36
D	5.0	3.0	1.0	2.7	0.7	0.8	4.0	1.6	0.2	0.1	8.3	0.4	48
E	0.2	4.5	0.9	1.1	2.9	0.2	3.8	2.8	1.3	4.2	2.3	0.2	60
F	4.1	0.6	0.2	0.2	2.3	0.7	2.4	4.6	0.3	2.9	0.7	0.2	72
G	3.6	5.8	0.3	0.2	4.4	0.3	0.6	2.1	0.3	0.7	0.3	0.2	84
H	0.3	0.2	4.8	0.3	5.5	0.4	0.2	3.7	4.9	4.3	0.3	0.1	96

D181

Conversion NRS301-42													
	1	2	3	4	5	6	7	8	9	10	11	12	
A	4.2	3.0	2.9	0.2	0.3	4.2	4.2	4.0	4.4	2.8	0.3	5.0	12
B	1.5	1.9	0.2	3.9	4.5	2.4	0.3	0.2	0.4	3.3	4.3	1.1	24
C	0.6	4.7	0.3	3.6	2.4	2.5	0.2	3.7	3.5	2.7	3.8	4.4	36
D	3.1	3.1	4.9	3.1	4.2	0.4	3.2	0.4	0.2	2.6	11.0	0.3	48
E	2.9	0.6	4.5	3.1	3.6	3.3	2.0	3.2	0.0	3.8	0.2	0.2	60
F	2.0	2.7	4.4	2.8	0.2	0.9	0.2	3.2	0.3	0.2	0.2	0.3	72
G	3.3	3.8	3.3	0.0	3.9	2.3	3.4	0.3	2.8	0.4	3.5	0.1	84
H	2.0	3.1	2.3	1.2	2.8	4.2	3.4	12.6	3.7	1.8	1.0	0.2	96

II: UV-Vis analysis of substrate binding to TxtE-BM3R(R59C)





III: HRMS of products from Chapter 5

Product	Expected $[\text{M}+\text{H}]^+$	Found $[\text{M}+\text{H}]^+$
129b	224.0830	224.0831
130b	220.1081	220.1077
131b	236.1030	236.1023
91b	236.1030	236.1027
135b	218.0924	218.0926, 218.0926