**Investigating the evolvability of
Escherichia coli nfsA via simultaneous
site-directed mutagenesis**

By

Kelsi Rose Hall

A thesis submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Biotechnology

Victoria University of Wellington
(2019)

Abstract

Bacterial nitroreductases are flavoenzymes able to catalyse the reduction of nitroaromatic compounds. The research presented in this thesis focused on NfsA_Ec, a nitroreductase from *E. coli*. NfsA_Ec is a promiscuous enzyme that can reduce a wide range of nitroaromatic antibiotics and prodrugs. This research sought to use NfsA_Ec as a model to improve our understanding of directed evolution, and also to identify NfsA_Ec variants exhibiting improved activation with a range of nil-bystander prodrugs for use in a targeted cell ablation system in zebrafish.

There is a substantial gap between the levels of enzyme activity that nature can achieve and those that scientists can evolve in the lab. This suggests that conventional directed evolution techniques involving incremental improvements in enzyme activity may frequently fail to ascend even local fitness maxima. We sought to contrast such approaches with simultaneous site-directed mutagenesis, employing a library of 252 million unique *nfsA* variants. To determine whether two superior NfsA_Ec variants recovered from this library could have been identified using a conventional stepwise approach we generated all possible intermediates of these two enzyme variants and recreated the most logical evolutionary trajectory for each enzyme variant. This revealed that a stepwise mutagenesis approach could indeed have yielded both of these variants, but also that very few evolutionary trajectories were accessible due to complex epistatic interactions between substitutions in these enzymes. Moreover, many conventional stepwise mutagenesis approaches such as iterative saturation mutagenesis would have failed to identify key substitutions in these variants. We also investigated the “black-box” effect of directed evolution, using NfsA_Ec and a panel of nitroaromatic compounds to model the off-target effects an evolved enzyme can have within an existing metabolic network. We found that selection for improved niclosamide and chloramphenicol detoxification also improved activity with some structurally distinct prodrugs, but not others. Using a dual positive-negative selection, we recovered NfsA_Ec variants that were more specialised for their primary activities, however this came at a cost in terms of overall activity levels.

The simultaneous site-directed *nfsA_Ec* mutagenesis library also had practical applications, enabling recovery of NfsA_Ec variants for targeted cell ablation in zebrafish models. These models involve the selective ablation of nitroreductase expressing cells without harming adjacent cells, to mimic a degenerative disease. Several NfsA_Ec variants were identified which were highly active with the nil-bystander prodrugs metronidazole, tinidazole, RB6145 and misonidazole when expressed in *E. coli*. However, these NfsA_Ec variants had inconsistent activities in our eukaryotic cell model (HEK-293). To expand the utility of the core ablation system, we sought to identify pairs of nitroreductases with non-overlapping prodrug specificities, suitable for use in a multiplex cell ablation system. Using a dual positive-negative selection, we recovered several NfsA_Ec variants that exhibited preferential nitrofurazone activation over metronidazole. Our lead variants for both applications are currently being trialed in zebrafish for their utility in generating degenerative disease models.

Acknowledgements

Firstly, I would like to thank the Royal Society of New Zealand Marsden fund and the Victoria University of Wellington for providing me with funding which enabled me to complete my studies. I would also like to thank other funding bodies; Research for Life, VUW Postgraduate Student Association, Royal Society of New Zealand, Claude McCarthy, VUW Faculty Strategic Research Grants, Maurice Wilkins Centre, New Zealand Microbiological Society, Centre for Biodiscovery and the Graduate Women Wellington, for providing me with travel funding to attend several conferences and participate in an international research collaboration during my PhD studies.

I would like to thank my supervisor Dave, for your continued support and guidance which enabled me to complete my PhD. I will be forever grateful for you allowing me to be a part of your research group which is filled with the most amazing, fun and supportive group of people, all of which is down to you being a great supervisor. I am also thankful for all the opportunities you provided me with, enabling me to attend conferences and participate in research collaborations. I would like to thank Jeff Mumm and Meera Saxena for providing me with the opportunity to spend time in their research lab at Johns Hopkins University in Baltimore. You guys welcomed us so openly into your research group and enabled me to learn an entire new set of research skills. I would also like to thank the technicians and support staff who keep the department running which enabled me to complete my research. Thank you to Dr Gary Evans and Dr Ralf Schwoerer for synthesising RB6145 for use in this study. I would also like to thank my secondary supervisor and Honours supervisor Wayne Patrick, for encouraging me to continue studying and for introducing me to Dave. A special thank you to Katherine for mentoring me when I first joined the lab and for providing advice throughout my PhD. I would also like to thank all other past and present members of the Ackerley and Owen lab groups for their guidance and who have made working in the lab so enjoyable over the past few years. I have countless fond memories at the many Castlepoint lab retreats and at our conference in Toulouse where I learnt many life lessons such as never to interact with strangers on a foreign bus.

Thank you to all my friends and family for your continued support throughout my many years of study, in particular to my Mum and Dad for trying your hardest to understand what I do and for sending me care packages to keep me going in my final few weeks of writing. Finally, I would like to thank my other half Rhys for keeping me fed in the last few months, keeping me sane and for painstakingly reading every page of this thesis. I am incredibly grateful for your support that helped me get through the last few years of study.

Table of Contents

Abstract.....	i
Acknowledgements	ii
Table of Contents	iii
List of Figures.....	ix
List of Tables	xii
Abbreviations and Definitions	xv
Chapter 1 Introduction.....	1
1.1 Protein Engineering	1
1.1.1 Rational protein design.....	1
1.1.2 Directed evolution	1
1.1.3 Promiscuous enzymes and directed evolution.....	4
1.2 Nitroreductase enzymes.....	5
1.2.1 Catalytic mechanisms.....	5
1.2.2 Structure	6
1.3 Biotechnological applications of nitroreductases	7
1.3.1 Nitroreductase mediated biocatalysis	7
1.3.2 Nitroreductase mediated bioremediation.....	7
1.3.3 Nitroreductase mediated prodrug activation	8
1.3.4 Targeted cell ablation in zebrafish	10
1.4 NfsA_Ec versus NfsB_Ec.....	15
1.4.1 Structure of NfsA	15
1.4.2 Native function of NfsA	16
1.4.3 Promiscuous activities of NfsA.....	16
1.5 Aims.....	20
Chapter 2 Methods.....	21
2.1 Chemicals reagents, enzymes and antibodies	21
2.2 Synthetic genes and oligonucleotide primers	21
2.3 Bacterial strains, mammalian cell lines and plasmids used in this study.....	22
2.3.1 Bacterial strains	22

2.3.2	Mammalian cell lines.....	22
2.3.3	Plasmids.....	22
2.4	Bacterial growth and maintenance.....	23
2.4.1	Growth media	23
2.4.2	Bacterial antibiotics and induction agents	23
2.4.3	Growth and storage.....	24
2.5	Standard molecular biology techniques	24
2.5.1	Isolation of plasmid DNA.....	24
2.5.2	Polymerase chain reaction (PCR) protocol.....	24
2.5.3	Restriction enzyme digest.....	25
2.5.4	DNA product purification.....	26
2.5.5	Agarose gel electrophoresis	26
2.5.6	Ligation.....	26
2.5.7	DNA quantification	27
2.5.8	DNA sequencing and analysis	27
2.5.9	Transformation of <i>E. coli</i>	27
2.6	Library Creation.....	29
2.6.1	Small-scale library creation (<5,000,000 clones)	29
2.6.2	Large-scale library creation (>5,000,000 clones)	30
2.6.3	Selection of nitroreductase libraries using positive selection compounds.....	30
2.7	Nitroreductase candidate library	31
2.8	Characterisation of nitroreductases	32
2.8.1	Growth Inhibition Assays	32
2.8.2	IC ₅₀ Assays.....	33
2.8.3	Whole cell fluorescence assay	33
2.8.4	SDS-PAGE of intracellular over-expressed nitroreductases	34
2.8.5	Protein purification	35
2.8.6	<i>In vitro</i> His ₆ -tagged protein kinetics	37
2.9	Generation and maintenance of HEK-293 cell lines.....	39
2.9.1	Growth of HEK-293 cell lines.....	39
2.9.2	Storage and recovery of frozen stocks.....	39
2.9.3	Generation of F279-V5:NTR constructs using Gateway® technology	40
2.9.4	Transfection of HEK-293 cells.....	41
2.9.5	Characterisation of HEK-293 cells expressing nitroreductases.....	41

2.10	Expression of nitroreductases in zebrafish	44
2.10.1	Ethics statement	44
2.10.2	Zebrafish maintenance	45
2.10.3	Dechoriation.....	45
2.10.4	Generation of transgenic zebrafish lines	45
2.10.5	Assessment of prodrug-induced lethality in non-transgenic zebrafish	47
2.10.6	Cell ablation assays.....	47
2.10.7	Confocal Microscopy.....	48
2.11	Biological and technical replicates	48
Chapter 3	Simultaneous site-directed mutagenesis of <i>nfsA</i>	49
3.1	Introduction	49
3.1.1	Simultaneous site-directed mutagenesis.....	50
3.1.2	The 7SM <i>nfsA</i> Library.....	50
3.1.3	Using NfsA as a model to understand enzyme evolution.....	52
3.2	Aims.....	52
3.3	Results	52
3.3.1	The 8SM <i>nfsA</i> Library.....	52
3.3.2	Final selection of eight-targeted residues for a simultaneous site-directed mutagenesis library.....	53
3.3.3	Generation of a large simultaneous site-directed mutagenesis <i>nfsA</i> library	59
3.3.4	Using a positive selection to recover superior NfsA variants	64
3.3.5	Characterisation of the top chloramphenicol reducing NfsA variants	66
3.3.6	Mutations in the plasmid backbone of selected clones	73
3.4	Discussion.....	76
Chapter 4	Using NfsA to model evolutionary trajectories	79
4.1	Introduction	79
4.1.1	Superior chloramphenicol detoxifying NfsA variants.....	79
4.2	Aims.....	80
4.3	Results	81
4.3.1	Creation and characterisation of intermediates of the top two chloramphenicol detoxifying NfsA variants	81
4.3.2	Recreating the evolutionary trajectories of NfsA 36_37 and 20_39	81
4.3.3	Epistatic interactions between amino acids in NfsA 36_37 and 20_39	87

4.3.4	Consequence of evolving NfsA variants towards chloramphenicol detoxification..	92
4.4	Discussion	93
Chapter 5 Using NfsA to investigate the off-target effects of directed evolution .		97
5.1	Introduction.....	97
5.1.1	Examples of metabolic interference	98
5.1.2	NfsA as a model for metabolic interference	98
5.2	Aims.....	99
5.3	Results.....	99
5.3.1	Effects of subjecting the Marsden library to positive selection pressures	99
5.3.2	Effects of subjecting the Marsden library to a dual positive-negative selection pressure	105
5.3.3	Summary of the five selection conditions.....	112
5.3.4	Kinetic analysis of the promiscuous activities of His ₆ -tagged NfsA	114
5.4	Discussion	118
Chapter 6 Identifying improved NfsA variants as tools for targeted cell ablation		121
6.1	Introduction.....	121
6.1.1	NfsB_Ec mediated cell ablation	121
6.1.2	NfsB_Ec triple mutant mediated cell ablation.....	122
6.1.3	NfsB_Vv F70A/F108Y.....	122
6.1.4	Evolving NfsA variants for use in targeted cell ablation.....	123
6.2	Aims.....	124
6.3	Results.....	124
6.3.1	Creation of stable zebrafish lines.....	124
6.3.2	Characterisation of NfsB_Vv F70A/F108Y in zebrafish.....	126
6.3.3	RB6145 toxicity in non-transgenic zebrafish larvae.....	131
	Recovery and characterisation of NfsA variants exhibiting improved activation of 5-nitroimidazole prodrugs in <i>E. coli</i>	132
6.3.4	.....	132
6.3.5	Expression and characterisation of <i>nfsA</i> variants with improved activation of metronidazole and tinidazole in HEK-293 cells.	135
6.3.6	Recovery and characterisation of NfsA variants exhibiting improved activation of 2-nitroimidazole prodrugs in <i>E. coli</i>	143

6.3.7	Expression and characterisation of <i>nfsA</i> variants with improved activation of RB6145 and misonidazole in HEK-293 cells.....	147
6.3.8	Development of a transient cell ablation assay in zebrafish.....	152
6.4	Discussion.....	153
Chapter 7	Directed evolution of NfsA for Multiplex Targeted Cell Ablation	157
7.1	Introduction	157
7.1.1	Multiplex Cell Ablation.....	157
	Previous multiplex cell ablation successes in the Ackerley lab group.....	159
7.1.2	.....	159
7.2	Aims.....	160
7.3	Results	160
7.3.1	Improving the tinidazole selectivity of NfsA in <i>E. coli</i>	160
7.3.2	Expression and characterisation of <i>nfsA</i> variants exhibiting tinidazole selectivity in HEK-293 cells.	165
7.3.3	Investigating alternative multiplex cell ablation prodrug combinations in <i>E. coli</i> .	171
7.3.4	Expression and characterisation of <i>nfsA</i> variants exhibiting nitrofurazone selectivity in HEK-293 cells.	177
7.4	Discussion.....	181
Chapter 8	Key findings, conclusions and future directions	185
8.1	Research motivation	185
8.2	Summary of key findings.....	186
8.2.1	Identifying NfsA variants exhibiting superior chloramphenicol detoxification.....	186
8.2.2	Using NfsA 36_37 and 20_39 to improve our understanding of directed evolution....	186
8.2.3	Evolution of NfsA variants for improved targeted cell ablation in zebrafish.....	187
8.2.4	Evolution of nitroreductases to enable a multiplex targeted cell ablation system in zebrafish.	188
8.2.5	Insights into the contributions of the eight-targeted residues in the Marsden library	189
8.3	Critical evaluation of the methods used in this research.....	189
8.3.1	Selection of the eight residues to target in NfsA.....	189
8.3.2	The use of the pUCX plasmid, library construction and IPTG induction	190
8.3.3	<i>In vivo</i> (<i>E. coli</i> 7NT) selection of NfsA variants.....	190

8.3.4	Dual positive-negative selection.....	191
8.4	Future directions	191
8.4.1	Additional evolutionary trajectories of top chloramphenicol detoxifying NfsA variants.....	191
8.4.2	Targeted cell ablation in zebrafish.....	192
8.4.3	Alternative selection to identify NfsA variants for multiplex cell ablation.....	192
8.5	Concluding remarks	193
References		195
Appendices.....		215
	Appendix A: Top chloramphenicol detoxifying NfsA variants.....	215
	Appendix B: IC ₅₀ values of all NfsA 36_37 and 20_39 intermediates	217
	Appendix C: Growth inhibition assay and sequence data of NfsA variants recovered from various selection conditions.....	219
	Appendix D: Michalis-Menten plots of wild-type NfsA with nitroaromatic compounds. ..	224
	Appendix E: Development of a transient cell ablation assay in zebrafish.....	225

List of Figures

Figure 1.1: Schematic of type I (oxygen-insensitive) and type II (oxygen-sensitive) nitroreductase mediated reduction of a nitroaromatic compound.	6
Figure 1.2: Chemical structure of the prodrug CB1954 (5-(1-Aziridiny1)-2,4-dinitrobenzamine).	9
Figure 1.3: Schematic of nitroreductase mediated targeted cell ablation in zebrafish.	13
Figure 1.4: Chemical structures of the nil bystander prodrugs metronidazole, tinidazole, azomycin, RB6145 and misonidazole.	14
Figure 1.5: Structure of NfsA.	15
Figure 1.6: Chemical structures of the prodrugs nitrofurazone, nitrofurantoin and 2,4-DNT.	17
Figure 1.7: Chemical structures of the nitroaromatic compounds niclosamide and chloramphenicol.	18
Figure 2.1: Schematic showing the generation of transgenic zebrafish lines stably expressing a nitroreductase.	46
Figure 3.1: Targeted positions in the 7SM and 8SM <i>nfsA</i> libraries.	53
Figure 3.2: Ratio of <i>E. coli</i> 7NT cells expressing <i>pUCX_empty</i> to <i>pUCX_nfsA</i> at different time points in a liquid culture.	63
Figure 3.3: SDS-PAGE showing the relative protein expression levels of ten <i>E. coli</i> 7NT strains expressing <i>pUCX_nfsA</i> variants in the absence of IPTG.	66
Figure 3.4: Enzyme model of NfsA depicting the most common substitutions present in the top 30 chloramphenicol detoxifying variants.	69
Figure 3.5: Chloramphenicol IC ₅₀ assays of <i>E. coli</i> 7NT cells expressing <i>nfsA</i> variants with or without mutations in the <i>lacI</i> region.	75
Figure 3.6: SDS-PAGE gel showing relative nitroreductase enzyme levels of <i>E. coli</i> 7NT strains expressing <i>pUCX_nfsA</i> constructs with or without mutations in the <i>lacI</i> region.	75
Figure 4.1: Evolutionary trajectory of NfsA 36_37 and 20_39.	82
Figure 4.2: Positive evolutionary trajectories of NfsA 36_37 and 20_39.	85
Figure 4.3: Positive evolutionary trajectories of the NfsA variants 36_37* and 20_39*.	86
Figure 4.4: The fitness effect of the single amino acid substitutions present in NfsA 36_37 and 20_39.	88
Figure 4.5: Epistatic interactions of S41Y, T219Y and R225V/D.	89
Figure 4.6: Epistatic interactions of the H215C and H215N substitutions.	90
Figure 4.7: Evolutionary progression of NfsA 20_39 with and without the addition of H215N.	91

Figure 4.8: Activity of enzyme intermediates of NfsA 36_37 and 20_39 with a range of nitroaromatic prodrugs.	92
Figure 5.1: Growth inhibition assay data of <i>E. coli</i> 7NT strains expressing <i>nfsA</i> variants selected on positive selection conditions.	102
Figure 5.2: Schematic of a positive selection and dual positive-negative selection.	105
Figure 5.3: Growth inhibition assay data of <i>E. coli</i> 7NT expressing <i>nfsA</i> variants recovered from dual positive-negative selection conditions.	109
Figure 5.4: Summary graph of the 57 NfsA variants recovered under each of the six selection conditions.	113
Figure 5.5: Linear regression of the log k_{cat}/K_M of nitroaromatic compounds in a NfsA catalysed reaction and their single-electron reduction potential (E^I_7).	117
Figure 6.1: Schematic of the Gal4/UAS trans expression system.	125
Figure 6.2: Metronidazole induced ablation in transgenic zebrafish larvae expressing <i>nfsB_Vv F70A/F108Y</i> across various challenge times.	127
Figure 6.3: Metronidazole induced cell ablation in transgenic zebrafish larvae expressing <i>nfsB_Vv F70A/F108Y</i> at decreased metronidazole concentrations challenged for a 24- or 6-hour period.	128
Figure 6.4: Confocal imaging of transgenic zebrafish.	129
Figure 6.5: Chemical structures of the prodrugs RB6145 and misonidazole.	130
Figure 6.6: RB6145 and misonidazole induced ablation in transgenic zebrafish larvae expressing <i>nfsB_Vv F70A/F108Y</i>	130
Figure 6.7: RB6145 toxicity of non-transgenic zebrafish larvae.	132
Figure 6.8: Chemical structures of 5-nitroimidazole prodrugs metronidazole and tinidazole.	132
Figure 6.9: Activation of FSL61 by <i>E. coli</i> 7NT strains expressing the top metronidazole and tinidazole reducing <i>nfsA</i> variants.	137
Figure 6.10: HEK-293 cell lines stably expressing <i>nfsA</i> variants visualised by the activated FSL61 fluorophore.	138
Figure 6.11: Flow cytometry of mixed populations of wild-type HEK-293 cells and HEK-293 cells expressing wild-type <i>nfsA</i> incubated with FSL61.	139
Figure 6.12: Flow cytometry of HEK-293 cells stably expressing the top metronidazole and tinidazole activating <i>nfsA</i> variants.	140
Figure 6.13: Western blot depicting the expression levels of <i>nfsA</i> variants in HEK-293 cells for variants exhibiting improved activation with metronidazole and tinidazole.	141
Figure 6.14: Activation of FSL61 by <i>E. coli</i> 7NT strains expressing the top RB6145 and misonidazole activating <i>nfsA</i> variants.	148
Figure 6.15: HEK-293 cell lines stably expressing <i>nfsA</i> variants visualised using the nitro-quenched fluorophore FSL61.	149

Figure 6.16: Flow cytometry of HEK-293 cells stably expressing the top RB6145 and misonidazole activating <i>nfsA</i> variants.....	150
Figure 6.17: Western blot depicting the expression levels of <i>nfsA</i> variants when expressed in HEK-293 cells.	150
Figure 7.1: Nitroreductase mediated multiplex cell ablation.....	158
Figure 7.2: Chemical structures of nil-bystander nitroaromatic prodrugs.....	159
Figure 7.3: Amino acid substitutions in NfsA 5_B9 and NfsA 5_B9 derived variants...	163
Figure 7.4: Activation of FSL61 by <i>E. coli</i> 7NT strains expressing the top <i>nfsA</i> variants exhibiting tinidazole selectivity.	166
Figure 7.5: Visualisation of HEK-293 cell lines stably expressing <i>nfsA</i> variants using FSL61.....	168
Figure 7.6: Flow cytometry of HEK-293 cells stably expressing the top tinidazole selective <i>nfsA</i> variants.....	169
Figure 7.7: Western blot depicting <i>nfsA</i> variant expression in HEK-293 cells.	170
Figure 7.8: Activation of FSL61 by <i>E. coli</i> 7NT strains expressing the top <i>nfsA</i> variants exhibiting nitrofurazone selectivity.	178
Figure 7.9: Visualisation of HEK-293 cell lines stably expressing <i>nfsA</i> variants using the nitro-quenched fluorophore FSL61.	179
Figure 7.10: Flow cytometry of HEK-293 cells stably expressing the top nitrofurazone selective <i>nfsA</i> variants.....	180
Figure 7.11: Western blot depicting <i>nfsA</i> variant expression in HEK-293 cells.	180

List of Tables

Table 2.1: Oligonucleotide primers used in this study.....	22
Table 2.2: <i>E. coli</i> strains used in this study.....	22
Table 2.3: Plasmids used in this study	23
Table 2.4 Liquid growth media recipes.	23
Table 2.5: Buffer recipes used in the isolation of plasmid DNA.	24
Table 2.6: Biomix Red ^{TM} PCR components and parameters.....	25
Table 2.7: Q5 ® High Fidelity Master Mix PCR components and parameters.	25
Table 2.8: GeneMorph ® II PCR components and parameters.	25
Table 2.9: Ligation components.....	26
Table 2.10: Oxidoreductase enzymes in the Ackerley group nitroreductase candidate library.	32
Table 2.11: SDS-PAGE solution recipes.	35
Table 2.12: Buffer recipes for His ₆ -tag purification using Ni-NTA His-Bind ® Resin.	36
Table 2.13: Extinction coefficients of compounds used in this study.....	37
Table 2.14: Western blot buffers.....	43
Table 2.15: Antibodies and dilutions used in western blots.	44
Table 2.16: Zebrafish maintenance media recipes.	45
Table 3.1: Percentage growth inhibition and sequencing results for 7SM <i>nfsA</i> variants selected on niclosamide or chloramphenicol.	56
Table 3.2: Percentage growth inhibition and sequencing results for 8SM <i>nfsA</i> variants selected on niclosamide and chloramphenicol.	57
Table 3.3: Percentage variation at the 15 positions targeted in the 7SM and 8SM <i>nfsA</i> libraries.....	58
Table 3.4: Degenerate codons utilised in the Marsden library.....	59
Table 3.5: Number of transformants with no anomalies.....	61
Table 3.6: The percentage of the Marsden library able to grow on various selective concentrations of chloramphenicol or niclosamide.....	65
Table 3.7: Percentage of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of the top 114 <i>NfsA</i> variants selected on 45 and 47.5 μ M chloramphenicol.....	66
Table 3.8: Chloramphenicol IC ₅₀ values and amino acid substitutions for the <i>E. coli</i> 7NT cells expressing top 30 <i>nfsA</i> variants.	68

Table 3.9: Kinetic parameters for the reduction of chloramphenicol by the top five NfsA variants.....	71
Table 3.10: Kinetic parameters for the reduction of p-benzoquinone by the top five NfsA variants.....	73
Table 4.1: Amino acids at the eight-targeted positions in the top two chloramphenicol detoxifying NfsA variants.....	81
Table 4.2: Structure of the amino acids present at positions 225 and 227 in NfsA 36_37, NfsA 20_39 and wild-type NfsA.	84
Table 4.3: Summary of the effect each amino acid substitution in NfsA 36_37 or NfsA 20_39 had on overall enzyme fitness.....	87
Table 5.1: Three positive selection conditions and percentage of the Marsden library able to grow under these conditions.	100
Table 5.2: Percentage occurrence of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library for NfsA variants recovered from the positive selection conditions.....	104
Table 5.3: The five positive and dual positive-negative selections used with the resulting percentage selection.	107
Table 5.4: Median percentage growth of NfsA variants recovered from various selection conditions challenged with niclosamide.	111
Table 5.5: Percentage occurrence of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of NfsA variants selected on dual positive-negative selection conditions.	111
Table 5.6: Kinetic parameters for the reduction of CB1954 by His ₆ -tagged NfsA with and without the NADPH regeneration system.	116
Table 5.7: Kinetic parameters for the reduction of CB1954, nitrofurazone, 2,4-DNT, RB6145 and metronidazole by His ₆ -tagged NfsA.	116
Table 6.1: Metronidazole IC ₅₀ values and sequences of the top 10 nfsA variants expressed in <i>E. coli</i> 7NT.....	134
Table 6.2: Tinidazole IC ₅₀ values and sequences of the top 10 nfsA variants expressed in <i>E. coli</i> 7NT cells.	135
Table 6.3: Metronidazole and tinidazole IC ₅₀ values of HEK-293 cell lines stably expressing <i>nfsA</i> variants.....	142
Table 6.4: RB6145 IC ₅₀ values and sequences of the top ten nfsA variants expressed in <i>E. coli</i> 7NT.....	145
Table 6.5: Misonidazole IC ₅₀ values and sequences of the top ten nfsA variants expressed in <i>E. coli</i> 7NT.....	146
Table 6.6: RB6145 and misonidazole IC ₅₀ values of HEK-293 cell lines stably expressing <i>nfsA</i> variants.....	151
Table 6.7: Summary of the <i>nfsA</i> variants cloned into the zebrafish pTOL2 vector.	152

Table 7.1: Tinidazole and metronidazole IC ₅₀ values and sequences of the top four <i>nfsA</i> 5_B9 derived <i>nfsA</i> variants expressed in <i>E. coli</i> 7NT.	162
Table 7.2: Tinidazole and metronidazole IC ₅₀ values and sequences of the top ten Marsden <i>nfsA</i> variants expressed in <i>E. coli</i> 7NT.	164
Table 7.3: Percentage of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of the top 10 NfsA variants.	165
Table 7.4: Tinidazole and metronidazole IC ₅₀ values of HEK-293 cell lines stably expressing <i>nfsA</i> variants.	171
Table 7.5: Nitrofurazone, nitrofurantoin and metronidazole IC ₅₀ values of <i>E. coli</i> 7NT strains expressing <i>nfsA_Ls</i> or <i>nfsA_Ec</i>	172
Table 7.6: Nitrofurazone, nitrofurantoin and metronidazole IC ₅₀ values and sequences of the top <i>nfsA</i> variants expressed in <i>E. coli</i> 7NT.	174
Table 7.7: Metronidazole and nitrofurazone IC ₅₀ values and sequences of the top nitrofurazone selective <i>nfsA</i> variants expressed in <i>E. coli</i> 7NT.	176
Table 7.8: Percentage of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of the top ten nitrofurazone selective NfsA variants....	176
Table 7.9: Nitrofurazone and metronidazole IC ₅₀ values of HEK-293 cell lines stably expressing <i>nfsA</i> variants.	181

Abbreviations and Definitions

× g	× g-force (also known as relative centrifugal force, RCF)
2,4-DNT	2,4-dinitrotoluene
7SM	<i>nfsA</i> mutagenesis library with seven positions mutated (precursor to the Marsden library)
8SM	<i>nfsA</i> mutagenesis library with eight positions mutated (precursor to the Marsden library)
Amp	Ampicillin
BCA	Bicinchoninic Acid
BDEPT	Bacterial directed enzyme prodrug therapy
BSA	Bovine serum albumin
CM	Chloramphenicol
ddH ₂ O	Double distilled H ₂ O
DEPT	Directed enzyme prodrug therapy
DMSO	Dimethyl sulfoxide
dpf	Days post fertilisation
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
FACS	Fluorescence-activated cell sorting
FCS	Fetal calf serum
FMN	Flavin mononucleotide
Gal4	Yeast transcription factor
GDEPT	Gene directed enzyme prodrug therapy
GFP	Green fluorescent protein
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His ₆	Polyhistidine tag containing six histidine residues
IC ₅₀	Concentration of drug required to inhibit growth of an organism by 50%
IPTG	Isopropyl-β-D-1-thiogalactopyranoside
k_{cat}	Kinetic constant describing the theoretical maximum rate of catalysis for a specific enzyme with a specific substrate under the conditions tested
K_M	Kinetic constant describing the concentration of substrate at which the rate of catalysis of a specific enzyme is half the k_{cat} under the conditions tested
LB	Lysogeny broth
Marsden library	Large-scale <i>nfsA</i> mutagenesis library with eight position mutated
MTZ	Metronidazole
NADH	Nicotinamide adenine dinucleotide (reduced)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NCS	Niclosamide
NFN	Nitrofurantoin
NfsA	NfsA_Ec (nitroreductase from <i>E. coli</i>)
NFZ	Nitrofurazone
NTR	Nitroreductase
OD ₆₀₀	Optical density at 600 nm wavelength (also known as cell turbidity)
PBS	Phosphate buffered saline

PCR	Polymerase chain reaction
PDB	Protein data bank
Pen-Strep	Penicillin-Streptomycin
PET	Positron emission tomography
PTU	1-phenyl-2-thiourea
RFU	Relative fluorescence units
rpm	Revolutions per minute
SD	Standard deviation
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SOC	Super optimal broth with catabolite repression
TE	Tris-EDTA
TNP	2,4,6-trinitrophenol
TNT	2,4,6-trinitrotoluene
TNZ	Tinidazole
Tris	Tris(hydroxymethyl)aminomethane
UAS	Upstream activating sequence
v/v	Volume per volume
VDEPT	Viral directed enzyme prodrug therapy
w/v	Weight per volume
YFP	Yellow fluorescent protein

Chapter 1 Introduction

1.1 Protein Engineering

Protein engineering is the process of modifying proteins towards a user-defined goal. This goal may include adapting the protein to achieve a change in substrate specificity, catalytic efficiency or increased stability in different temperatures, organic solvents and/or pH. Two general strategies are used in protein engineering; (1) rational protein design and (2) directed evolution.

1.1.1 Rational protein design

Rational protein design is one of the earliest protein engineering approaches and relies heavily on computational methods and structural biology to create or modify a protein. This approach involves the development of a molecular model using algorithms to predict the possible effects of different substitutions, which are subsequently generated in the lab (Hellinga, 1997; Kapoor *et al.*, 2017). A comprehensive understanding of the structure and function of the protein is often required to achieve success with this method. Although this approach has achieved numerous successes (Pinto *et al.*, 1997; Selzer *et al.*, 2000; Gocke *et al.*, 2008), it usually fails to address the role distant substitutions can have on protein structure and function, and the unexpected effects which can occur from combinations of amino acid substitutions (Chen *et al.*, 1999). In addition, rational protein design often focuses on single substitutions, limiting the amount of sequence space that is explored.

1.1.2 Directed evolution

Directed evolution is a predominately laboratory-based method, which mimics the process of natural selection to evolve proteins towards a user-defined goal, condensing the time scale of what may take nature millions of years to achieve into months or weeks (Arnold, 1998). This technique acknowledges that we often have a limited capability to predict the effects that individual amino acid substitutions can have on protein structure and function, and does not mandate prior extensive structural information. The general method involves the creation of large sets of diverse protein sequences known as a library. This library is then subjected to screening or a selection to recover proteins with the desired function. Directed evolution has gained a lot of media attention since Frances Arnold won the 2018 Nobel Prize in Chemistry for her efforts pioneering work in this field. This technique has

been used in a variety of applications and is helping to solve many industrial, medical and environmental problems our world is currently facing. Examples include the directed evolution of enzymes for bioremediation (Kumamaru *et al.*, 1998; Barriault *et al.*, 2002), insect-specific pest control (Deist *et al.*, 2014; Shao *et al.*, 2016) and cancer gene therapy (Guise *et al.*, 2007; Williams *et al.*, 2015). While the overall flow of a directed evolution experiment is the same, the type of mutagenesis used to create the library and the screening or selection method can vary.

1.1.2.1 Types of mutagenesis used in this study

1.1.2.1.1 Error-prone PCR

Error-prone PCR involves the use of a polymerase which exhibits a high error rate to randomly introduce mutations throughout a gene. Early error-prone PCR methods enhanced the intrinsic error rate of *Taq* polymerase, by modifying buffer components (e.g., $MnCl_2$) and/or the ratio of dNTPs present to reduce base pairing specificity (Cadwell and Joyce, 1992; Copp *et al.*, 2014a). However, this method has been shown to cause mutational bias, resulting in an overrepresentation of AT to GC changes (Cirino *et al.*, 2003; Copp *et al.*, 2014a). Alternative polymerases have been developed which provide a more balanced mutational spectrum (Rasila *et al.*, 2009), and are frequently sold as proprietary enzymes in commercially available kits (e.g., GeneMorph II sold by Agilent). The stochastic nature of error-prone PCR can introduce mutations which may not have been predicted to yield a fitness gain using a rational approach, such as mutations distant from the active site that nevertheless have profound effects on enzyme fitness (Rajagopalan *et al.*, 2002; Reetz *et al.*, 2004). While error-prone PCR can explore large amounts of sequence space, it requires large libraries and subsequently high-throughput screening to offset the large number of inevitable neutral or deleterious mutations. Key residues are also likely to be missed, as it is not possible to create a library large enough to mutate every residue to all 20 amino acids. Even a small protein of 100 amino acids would require 20^{100} possible combinations to sample all 20 amino acids at each position, more than the number of atoms present in the universe (Romero and Arnold, 2009).

1.1.2.1.2 Site-saturation mutagenesis

Site-saturation mutagenesis is an example of a semi-rational design approach, and one that frequently draws upon both computationally based rational design and random mutagenesis. Structural or functional information is used to rationally target specific

residues in an enzyme, normally around the active site. This is combined with a non-rational approach, using degenerate codons to encode a broad choice of amino acids at a given position (Reetz *et al.*, 2008). This approach allows for larger regions of sequence space to be explored while still taking a targeted approach, thus creating smarter libraries (Chica *et al.*, 2005). Mutations focused around the active site have been shown to be more effective than distant mutations when evolving enzymes to modify enantioselectivity, substrate selectivity and alternative catalytic activities (Morley and Kazlauskas, 2005). Site-directed mutagenesis can be used in conjunction with a high-throughput screening technique to accelerate directed evolution.

1.1.2.1.3 Epistatic interactions in evolved enzymes

Epistasis describes interactions between two or more mutations, where the effect of one mutation depends on the presence of another (Romero and Arnold, 2009). Simultaneous site-directed mutagenesis increases the chances of positive epistatic or synergistic interactions, as opposed to mutating positions individually or at entirely random positions (Chica *et al.*, 2005). Epistatic interactions are crucial to the evolution of an enzyme because of their potential to amplify individually small increases in enzyme fitness (Mildvan, 2004; Weinreich *et al.*, 2006). However, epistatic interactions can constrain an evolutionary network, by limiting the number of evolutionary trajectories that are selectively accessible (Poelwijk *et al.*, 2011). This is because epistasis can create rugged fitness landscapes, where combinations of mutations are required to reach a fitness optimum, as opposed to a smooth fitness landscape which can be scaled by the accumulation of single beneficial mutations (Romero and Arnold, 2009; Kvitek and Sherlock, 2011; Poelwijk *et al.*, 2011). It is important to consider epistatic interactions when determining the best evolutionary strategy, as not all types of mutagenesis allow for synergistic or other epistatic interactions (Voigt *et al.*, 2000).

1.1.2.2 Recovering improved variants from an enzyme library

When working with a large enzyme library it is critical to have an efficient screen or selection to isolate variants exhibiting the desired activity. A screen refers to the direct examination of every variant in the library looking for a desired property, while a selection eliminates undesirable variants (Leemhuis *et al.*, 2009). Screens often employ the use of assays, which can be conducted either *in vivo* or *in vitro* to identify enzyme variants with a desired function. These assays use a variety of techniques to measure enzyme activity,

including monitoring cell growth, colorimetric read outs, substrate consumption or product formation and ligand binding assays (Aharoni *et al.*, 2005b; Engqvist and Rabe, 2019). These approaches are time consuming and often rely heavily on the use of plate-based assays and robots (Xiao *et al.*, 2015). A selection has the advantage of being more high-throughput than a screen, and links the function of the engineered enzyme directly to the survival of an organism (Engqvist and Rabe, 2019). This method is limited to the use of enzymes exhibiting detoxifying abilities or enzymes which synthesise essential nutrients, however provides an efficient method of isolating enzyme variants exhibiting the desired activity from large libraries (Leemhuis *et al.*, 2009).

1.1.3 Promiscuous enzymes and directed evolution

To ensure the metabolic network of a cell can operate efficiently, enzymes must exhibit substrate specificity, excluding structurally similar molecules. However, there is also pressure on a cell to be flexible enough to respond and adapt to changing environmental pressures. This flexibility often comes in the form of promiscuous enzyme functions, which a cell can draw upon in a changing environment. Promiscuous enzymes are defined as enzymes that are able to catalyse secondary reactions that are physiologically irrelevant to the cell (Copley, 2017). Enzymes can exhibit either substrate promiscuity or catalytic promiscuity. Catalytic promiscuity is when an enzyme is able to catalyse two or more mechanistically distinct reactions (O'Brien and Herschlag, 1999). In contrast, an enzyme exhibiting substrate promiscuity involves the same mechanistic reaction carried out on one or more non-canonical substrates (Copley, 2017). Many scientists have reported substantial changes in promiscuous activity as a result of only a few amino acid changes (Harel *et al.*, 2004; Aharoni *et al.*, 2005a). Promiscuous enzymes have therefore become a logical starting point in many directed evolution experiments (Toscano *et al.*, 2007; Bloom and Arnold, 2009).

The work conducted in this study focused on the directed evolution of a nitroreductase, an example of a promiscuous enzyme. In addition to their highly promiscuous nature, nitroreductases have a protein scaffold with an unusually plastic active site, making them highly amenable to mutations in that region. This makes nitroreductases excellent candidates for directed evolution.

1.2 Nitroreductase enzymes

Nitroreductases are flavoenzymes, capable of NAD(P)H-dependant reduction of nitroaromatic compounds. The nitroreductase superfamily is comprised of greater than 24,000 known sequences from all domains of life which share structural folds, select active site traits and mechanistic features (Akiva *et al.*, 2017). Despite the “nitroreductase” naming convention, members of this family are not all capable of nitro-reduction, but collectively are known to catalyse at least ten distinct enzymatic reactions. In addition to nitro-reduction, these reactions include; dehydrogenation (Gondry *et al.*, 2001; Cortial *et al.*, 2012), flavin fragmentation (Taga *et al.*, 2007; Yu *et al.*, 2012) and dehalogenation (Thomas *et al.*, 2009; Mukherjee and Rokita, 2015). These activities can act on a broad range of substrates including nitroaromatic (Roldan *et al.*, 2008), enone (Hou *et al.*, 2015), quinone (Liochev *et al.*, 1999), flavin mononucleotide (FMN) (Chung and Tu, 2012) and metal ion compounds (Ackerley *et al.*, 2004).

1.2.1 Catalytic mechanisms

Bacterial nitroreductases can be classified into two functional classes, based upon their oxygen sensitivity. Type II nitroreductases are oxygen-sensitive and catalyse a single-electron transfer from NAD(P)H to the nitro group, producing a nitro anion radical (Figure 1.1). This radical ion intermediate can be rapidly oxidised back to the parent compound in the presence of oxygen, generating a superoxide anion in the process (Roldan *et al.*, 2008). In contrast, type I nitroreductases are oxygen insensitive and can generate end products in the presence of oxygen. These nitroreductases catalyse concerted and consecutive two-electron transfers from NAD(P)H to the nitro group, generating nitroso, hydroxylamine and amino derivatives (Peterson *et al.*, 1979; Roldan *et al.*, 2008). These two-electron transfers prevent the formation of radical ion intermediates and are therefore not influenced by the oxygen status of the environment (Bryant *et al.*, 1981).

Figure 1.1: Schematic of type I (oxygen-insensitive) and type II (oxygen-sensitive) nitroreductase mediated reduction of a nitroaromatic compound. Type I nitroreductases catalyse the reduction of the nitro-group via concerted and consecutive two-electron transfers. Type II nitroreductases catalyse consecutive one-electron transfers, producing a nitro anion radical which can be rapidly oxidised back to the parental compound in the presence of oxygen.

Nitroreductase catalysis proceeds by a ping-pong bi-bi kinetic mechanism involving two redox half reactions. In the oxidative half of the reaction, NAD(P)H enters the active site and electrons are transferred to the FMN, the oxidised NAD(P)⁺ then exits the active site. In the reductive half of the reaction a terminal electron acceptor such as a nitroaromatic compound, enters the active site and subsequently accepts the pair of electrons from the FMN, the reduced substrate then dissociates from the active site. During this time, the FMN cofactor cycles between oxidised and reduced states (Roldan *et al.*, 2008). The two-electron chemistry of type I nitroreductases makes the one-electron reduced semiquinone state of FMN thermodynamically unfavourable, explaining the oxygen-insensitivity of type I nitroreductases (Koder *et al.*, 2002). Generation of the amine end-product is rarely achieved, as further reduction from the hydroxylamine has a much lower midpoint potential (E°) and is considerably less favourable (Miller *et al.*, 2018).

1.2.2 Structure

Nitroreductases are typically homodimers, consisting of two monomeric subunits that combine to form two FMN-binding active sites, with each monomer contributing to both active sites. The FMN moiety is non-covalently bound in the active site and is central to catalysis (Macheroux *et al.*, 2011). In 2017, crystal structures had been determined for 54 nitroreductase family members comprising 16 out of the 22 superfamily subgroups (Akiva *et al.*, 2017). In a 2017 study, these structures were used to determine the nitroreductase superfamily minimal scaffold. The scaffold created consisted of an $\alpha + \beta$ homodimeric

fold, harbouring key FMN interacting residues. These interactions with the FMN isoalloxazine ring were found to modulate the redox potential and influence catalysis (Akiva *et al.*, 2017).

1.3 Biotechnological applications of nitroreductases

The promiscuous activities of nitroreductases have been exploited for a range of biotechnological applications including biocatalysis (Hoogenraad *et al.*, 2004), bioremediation (Heiss and Knackmuss, 2002) and cell ablation (Grove *et al.*, 2003; Curado *et al.*, 2007; Vass *et al.*, 2009; Prosser *et al.*, 2010; Swe *et al.*, 2012). Nitroreductase mediated cell ablation can be used to target both cancerous cells via gene therapies and specific cell populations in zebrafish. The latter application is one of the primary foci of the research described in this thesis.

1.3.1 Nitroreductase mediated biocatalysis

Biocatalysis refers to the use of enzymes to speed up synthetic chemical reactions, generating useful compounds. Aromatic amines generated by nitroreductases have potential to be used as building blocks to generate high-value pharmaceutical compounds from readily available nitroaromatic compounds (Hoogenraad *et al.*, 2004). Although most well characterised nitroreductases seldom reduce nitroaromatic compounds through to the amine product, a few nitroreductases have been identified which do produce the amine end-product (Koder and Miller, 1998; Race *et al.*, 2005). Nitroreductases from *Clostridium* (Kutty and Bennett, 2005), *Klebsiella* (Kim and Song, 2005; Shin and Song, 2009), *Bacillus* (Chaignon *et al.*, 2006; Gwenin *et al.*, 2015), *Mycobacterium* (Manina *et al.*, 2010) and *Gluconobacter* (Yang *et al.*, 2016) have all demonstrated the ability to produce amine end-products from a range of nitroaromatic compounds (Miller *et al.*, 2018).

1.3.2 Nitroreductase mediated bioremediation

Widespread industrialisation has led to increasing amounts of environmental pollutants. Xenobiotic compounds constitute a large number of environment pollutants and includes nitroaromatic compounds, which are rarely found in nature (Ramos *et al.*, 2005; Ju and Parales, 2010). 2,4,6-trinitrotoluene (TNT) and 2,4,6-trinitrophenol (TNP) are examples of such nitroaromatic xenobiotic compounds and are serious environmental pollutants due to their large-scale production for explosives. These compounds are toxic to humans and are resistant to degradation (Hawari *et al.*, 2000). Nitroreductases have been used in both bioslurries and in combination with phytoremediation for the bioremediation of these

compounds (Hannink *et al.*, 2001; Heiss and Knackmuss, 2002; Kurumata *et al.*, 2005; Ramos *et al.*, 2005; Van Dillewijn *et al.*, 2008).

It has also been reported that some nitroreductases exhibit the ability to reduce chromate. Chromate [Cr(VI)] is a serious environmental pollutant and is produced as a by-product of many industrial processes. Chromate is readily taken up by bacterial and mammalian cells, and through a series of cellular processes generates a highly unstable Cr(V) radical and large quantities of reactive oxygen species, which contribute to its high toxicity (Venitt and Levy, 1974). Chromate is water soluble and toxic, however its reduced form, Cr(III), is insoluble in most environmental conditions, impermeable to most cells and less toxic, thus is the desired form (Imai and Gloyna, 1990). Nitroreductases from *Vibrio harveyi* (Kwak *et al.*, 2003) and *E. coli* (Ackerley *et al.*, 2004; Robins *et al.*, 2013) have been identified as chromate reductases and have shown promise for the bioremediation of chromate.

1.3.3 Nitroreductase mediated prodrug activation

Nitroreductase mediated prodrug activation is another leading biotechnological application of nitroreductases. The reduction of a nitro-group to a hydroxylamine or amine product is one of the largest electronic shifts achieved in biology (Denny, 2002). This has been exploited in prodrug design as addition of a strongly electron-withdrawing nitro-group on an aromatic ring can pull the electron density away from the highly reactive ring, rendering it inactive (Siim *et al.*, 1997). Reduction of the nitro-group by a nitroreductase subsequently pushes the electrons back onto the aromatic ring, generating a cytotoxic compound. This allows for tighter control over reactivity within a cell. Nitroreductase mediated cancer cell ablation and targeted cell ablation in zebrafish will be discussed here.

1.3.3.1 Nitroreductase mediated cancer cell ablation

One of the current limitations of conventional chemotherapy is the off-target damage associated with administering a cytotoxic compound that targets all rapidly dividing cells (Brown and Giaccia, 1998). Treatments such as directed enzyme prodrug therapy (DEPT) are providing promise for more targeted cancer therapies. In DEPT, an exogenous prodrug-activating enzyme is selectively delivered to tumour cells, sensitising these cells to a prodrug. One of the biggest limitations of this therapy is the difficulty of achieving effective delivery of the enzyme to the tumour. Gene directed enzyme prodrug therapy (GDEPT) is one approach, which involves delivery of a gene encoding the therapeutic enzyme to the

tumour. This is generally mediated by either a tumour tropic bacterial vector (BDEPT: bacterial DEPT) or viral vector (VDEPT: viral DEPT) (Palmer *et al.*, 2004; Lehouritis *et al.*, 2013). In the case of BDEPT, the therapeutic enzyme is expressed by the bacterial vector, while in VDEPT it is transcribed by the cancerous cells. The therapeutic enzyme is then able to convert a systemically administered non-toxic prodrug into its cytotoxic form, killing cancerous cells. The localisation of prodrug activation increases the therapeutic index and decreases systemic side effects of this cancer therapy. Prodrugs are chosen which result in a “high-bystander” effect. This is when an activated prodrug is able to diffuse out of the cell it was activated in and kill neighbouring cells (Freeman *et al.*, 1993; Chan-Hyams *et al.*, 2018). This increases the efficacy of this therapy as it allows the activated prodrug to kill more than one cell, without having to deliver the therapeutic enzyme to every cancer cell. This is especially critical for bacterial vectors, which are killed by the activated prodrug (Portsmouth *et al.*, 2007).

Nitroreductases have been used for DEPT strategies. The first nitroreductase-prodrug pairing to be used in cancer therapy was NfsB from *E. coli* in combination with the prodrug CB1954 (5-(1-Aziridiny)-2,4-dinitrobenzamine) (Figure 1.2).

Figure 1.2: Chemical structure of the prodrug CB1954 (5-(1-Aziridiny)-2,4-dinitrobenzamine).

CB1954 can be reduced at either the 2- or 4-nitro group, creating 2- or 4-hydroxylamine reaction products, respectively. The 4-hydroxylamine reacts with a thioester such as acetyl-CoA forming a bi-functional alkylating agent capable of crosslinking DNA (Knox *et al.*, 1993). The 2-hydroxylamine undergoes disproportionation to form the 2-nitroso and 2-amine forms. The 2-amine form is a toxic end-product that can form DNA mono-adducts and exhibits a higher bystander effect than the 4-hydroxylamine (Helsby *et al.*, 2004). CB1954 was first synthesised in Chester Beatty laboratories in the late 1960s to test for

activity against transplantable rat 256 carcinoma (Khan and Ross, 1969). This compound led to complete cures in rats, however activity did not translate in humans (Cobb *et al.*, 1969). The high activity observed in rats was later found to be a consequence of the rat oxidoreductase NQO1 (but not the human equivalent) reducing CB1954 to a 4-hydroxylamine derivative (Knox *et al.*, 1991). Despite these initial disappointing findings, CB1954 has dominated the field for use in cancer gene therapy (Grove *et al.*, 2003; Vass *et al.*, 2009; Prosser *et al.*, 2010; Swe *et al.*, 2012). However, therapy has been limited by the low achievable serum concentrations of CB1954 and the high K_M of NfsB_Ec (i.e., low prodrug affinity). Alternative CB1954-activating nitroreductases have been investigated for GDEPT, with a review by Williams *et al.*, (2015) listing over 30 nitroreductases capable of reducing CB1954, most at greater levels than NfsB_Ec. Existing nitroreductases have also been engineered to improve their ability to activate CB1954 (Grove *et al.*, 2003; Barak *et al.*, 2006; Guise *et al.*, 2007; Swe *et al.*, 2012) or next-generation nitroaromatic prodrugs such as dinitrobenzamide mustards (DNBM)s which appear far more effective (Prosser *et al.*, 2013; Copp *et al.*, 2017).

1.3.4 Targeted cell ablation in zebrafish

Targeted cell ablation in zebrafish can be used to model degenerative diseases and subsequently study regenerative medicine. Zebrafish (*Danio rerio*) are a powerful model organism for studying vertebrate biology owing to their high fecundity, short generation time, translucence during embryogenesis and ability to undergo external development (Dooley and Zon, 2000). They can be easily genetically manipulated and used in high-throughput screens (Driever *et al.*, 1996; Haffter *et al.*, 1996; Walker *et al.*, 2012). Other vertebrate models such as the mouse undergo early development *in utero*, require a lot of space and are prohibitively expensive. Zebrafish can be used to study vertebrate specific features such as the kidneys, multichambered heart, multi-lineage haematopoiesis, notochord and neural crest cells (Dooley and Zon, 2000). To date, a range of genes associated with different degenerative diseases have been identified in zebrafish, including haematopoietic (Weinstein *et al.*, 1996), neurodegenerative (Xi *et al.*, 2011), cardiovascular (Fishman and Olson, 1997) and kidney diseases (Drummond *et al.*, 1998). Several cell-specific ablation systems have been developed in an attempt to mimic degenerative diseases in zebrafish. These models are then used to study cell regeneration to develop treatments for human degenerative diseases.

1.3.4.1 Light/laser ablation

Light ablation has been used to ablate the zebrafish retina, which is particularly vulnerable to intense light. In this approach, prolonged intense light treatments are used to ablate rod and cone cells of the eye (Vihtelic and Hyde, 2000). An argon laser has also been used to achieve selective ablation of photoreceptors (Wu *et al.*, 2001). While this approach has shown targeted specificity, it is labour intensive, which limits sample sizes. This approach is also limited to cell ablation within the eye.

1.3.4.2 Chemical ablation

A variety of cell toxins have been used to elicit cell-specific ablation, generating a broad range of degenerative disease models. Some examples include MoTP [(2-morpholinobutyl)-4-thiophenol] to ablate melanocytes (Yang and Johnson, 2006), Triton X-100 to ablate olfactory epithelial cells (Iqbal and Byrd-Jacobs, 2010), streptozotocin to ablate pancreatic β -cells (Moss *et al.*, 2009), aminoglycosides to ablate mechanoreceptor hair cells in the neuromasts (Harris *et al.*, 2003) and dexamethasone to ablate T-cells (Langenau *et al.*, 2004). The biggest limitation with chemically induced cell ablation is the small range of cell types that can be targeted with each chemical. Some chemicals such as MoTP and streptozotocin also require manual injection into individual larvae, which is technically challenging and time consuming.

1.3.4.3 Transgene mediated targeted cell ablation

Several systems have been developed that involve the expression of transgenes to induce targeted cell ablation. This either involves the direct expression of a transgene which can elicit cell death (e.g., diphtheria toxic or caspases) or the expression of enzymes which convert non-toxic compounds such as a prodrug into cytotoxins (e.g., thymidine kinase or nitroreductases). The latter approach mimics several techniques used in cancer gene therapy. The direct expression of a cytotoxic transgene requires inducible transgene expression to allow temporal control over cell ablation, however, enzyme induced cell ablation can be controlled by the timing of prodrug administering. Transgene mediated targeted cell ablation has several advantages over light and chemical cell ablation methods: (1) Specificity, nearly any cell type can be targeted based on the localisation of the transgene; (2) Temporal control as noted above; (3) High-throughput, cell ablation can be simultaneously achieved in thousands of fish; and (4) Ease of use, once the transgenic zebrafish line has been generated, targeted cell ablation can be easily achieved by adding a

prodrug to the media (White and Mumm, 2013). Some examples of transgene mediated cell ablation are outlined below, highlighting key benefits and limitations of each system.

1.3.4.3.1 *Diphtheria toxin mediated cell ablation*

The diphtheria toxin is an endotoxin secreted by *Corynebacterium diphtheriae* and encoded by the *tox* gene (Holmes, 2000). Transgenic expression of the diphtheria toxin has been achieved in the exocrine of the zebrafish pancreas under the control of the *elastaseA* promoter (Wan *et al.*, 2006). The endocrine cells of the pancreas and other endodermal derivatives such as the liver and intestine were unaffected by the diphtheria toxin, allowing scientists to investigate exocrine pancreatic development. The diphtheria toxin has also been expressed in cardiomyocytes using the Cre-lox system, allowing for study of cardiomyocyte regeneration in zebrafish (Wang *et al.*, 2011). One limitation of this system is the extreme potency of the diphtheria toxin, which can result in off-target ablation, particularly when not under the control of a strict promoter.

1.3.4.3.2 *Caspase mediated cell ablation*

Caspases are a family of proteases that play an essential role in programmed cell death and inflammation. A modified caspase-3 containing chemically induced dimerization binding sites was used to induce hepatocyte ablation (Mallet *et al.*, 2002). Administering the chemical inducer resulted in dose-dependent hepatocyte ablation. Caspase-8 has also been used to mediate cell ablation. This was achieved by fusing caspase-8 to the estrogen receptor ligand-binding domain which induces *caspase-8* expression upon exposure to tamoxifen (Chu *et al.*, 2008). While both these systems have been reported in mouse models, they have yet to be shown effective in zebrafish models.

1.3.4.3.3 *Thymidine kinase/ganciclovir mediated cell ablation*

Thymidine kinase is an enzyme expressed in most living cells. The predominant role of thymidine kinase is to phosphorylate thymidine for the synthesis of DNA. The Herpes virus form of this enzyme is uniquely inhibited by the nucleotide analogue prodrug ganciclovir. The phosphorylated and activated form of ganciclovir inhibits the incorporation of deoxyguanosine triphosphates (dGTP) into DNA, leading to cell cycle arrest and death. Due to this mechanism of action, ganciclovir only proves cytotoxic towards actively dividing cells, limiting the application of this method. The high-bystander effect of activated ganciclovir would also limit the use of this methodology (Elshami *et al.*, 1996; Hamel *et al.*, 1996). This system has been used in gene therapy trials for targeting cancerous

cells (Oldfield *et al.*, 1993; Rainov and Grp, 2000), however like the caspase system, this has not been adapted into zebrafish (White and Mumm, 2013).

1.3.4.3.4 Nitroreductase prodrug mediated targeted cell ablation

Nitroreductase mediated cell ablation is one of the leading methods for achieving targeted cell ablation. This system involves localised expression of a nitroreductase, selectively sensitising cells to a nil-bystander prodrug, which upon nitro-reduction produces a cytotoxic compound that results in specific cell ablation (Curado *et al.*, 2007) (Figure 1.3). Nitroreductase expression is controlled by a cell-type specific promoter, allowing ablation to occur in a broad range of cell types. Another benefit of this system is that ablation can be monitored by co-expressing or fusing a fluorescent reporter protein such as YFP to the nitroreductase (Medico *et al.*, 2001).

Figure 1.3: Schematic of nitroreductase mediated targeted cell ablation in zebrafish. A nitroreductase and a fluorescent reporter gene are co-expressed under the control of a cell-type specific promoter, selectively sensitising cells to a prodrug such as metronidazole. Nitro-reduction of the prodrug produces a cytotoxic compound resulting in targeted cell ablation.

In contrast with GDEPT where a high-bystander effect is desirable, the use of a nil-bystander prodrug is essential to targeted cell ablation. The nil-bystander prodrug metronidazole is the most commonly used prodrug in nitroreductase mediated cell ablation. Metronidazole is a 5-nitroimidazole prodrug and is a common antibiotic used to treat various protozoans and anaerobic bacteria (Freeman *et al.*, 1997) (Figure 1.4). It has been shown that loss of nitroreductase activity is a common method of resistance against metronidazole (Goodwin *et al.*, 1998; Paulish-Miller *et al.*, 2014). The mechanism of action

of the reduced form of metronidazole is not well understood, however it is believed to be involved in DNA binding and adduct formation, resulting in strand breakage (Edwards, 1980; Goldman *et al.*, 1986; Tocher and Edwards, 1994). Metronidazole is also cell cycle independent, unlike ganciclovir, and is cost effective allowing for high-throughput screens (White and Mumm, 2013).

Metronidazole is most commonly used in combination with the nitroreductase NfsB_Ec. Like in cancer gene therapy, this system is limited by the relatively poor metronidazole activation by NfsB_Ec (Mathias *et al.*, 2014). Alternative nitroreductase prodrug pairings with improved prodrug activation could vastly improve this system. One of the aims of the work described in this thesis was to identify alternative nitroreductase-prodrug pairings, including the alternative nil-bystander prodrugs tinidazole, misonidazole and RB6145 (Figure 1.4).

1.3.4.3.4.1 Nil-bystander prodrugs

Figure 1.4: Chemical structures of the nil bystander prodrugs metronidazole, tinidazole, azomycin, RB6145 and misonidazole.

Tinidazole is a 5-nitroimidazole compound and is an analogue of metronidazole that was originally used in the treatment of protozoans (Fung and Doan, 2005) (Figure 1.4). The activated form of tinidazole is believed to have a similar mechanism of action to metronidazole (Edwards, 1980; Goldman *et al.*, 1986; Tocher and Edwards, 1994). RB6145, azomycin and misonidazole are all 2-nitroimidazole compounds (Figure 1.4). RB6145 is a pre-prodrug form of the chemotherapeutic and hypoxic cell radio-sensitising agent RSU1069 (Chaplin *et al.*, 1986; Bremner, 1993). Misonidazole has also been used as a hypoxic cell radio-sensitising agent (Dische *et al.*, 1979). Azomycin is an antibiotic produced by a strain of *Nocardia mesenterica* that is used to treat infections caused by a

variety of bacteria (Nakamura, 1955). The mechanisms of action of the activated forms of RB6145, azomycin and misonidazole are not well understood. It has been proposed that misonidazole may have a similar mechanism of action to metronidazole, destabilising the DNA helix via extensive strand breakage (Knight *et al.*, 1979).

1.4 NfsA_Ec versus NfsB_Ec

NfsB_Ec is the canonical nitroreductase that has dominated the nitroreductase field for applications in bioremediation, cancer gene therapy and targeted cell ablation in zebrafish. The widespread use of NfsB_Ec has been such that in several studies it is simply referred to as NTR or NR as a general abbreviation for nitroreductase (Friedlos *et al.*, 1997; Hannink *et al.*, 2001; Grove *et al.*, 2003). However, the large size of the nitroreductase superfamily means there is scope for improving all the applications mentioned above through the use of alternative nitroreductases (Akiva *et al.*, 2017). The work described in this thesis has predominately focused on an alternative nitroreductase from *E. coli*, NfsA_Ec, which from hereon will simply be referred to as NfsA. NfsA has shown to be superior in purified enzyme kinetics with the prodrug CB1954 in multiple independent studies (Vass *et al.*, 2009; Prosser *et al.*, 2010). Additional studies in the Ackerley lab have shown NfsA is more active than NfsB_Ec with a range of nitroaromatic prodrugs including metronidazole, S33, RB6145 and PR-104A (Williams, 2013; Rich, 2017; Sharrock, 2018).

1.4.1 Structure of NfsA

Figure 1.5: Structure of NfsA. (A) Structure of one of the active sites of NfsA. The active site is formed at the dimeric interface with residues contributing from both monomers. (B) Top-view of NfsA homodimer showing the two active sites each with FMN bound. Each monomer is shaded either blue or green with the FMN molecule shown in yellow. PDB accession number 1f5v.

The crystal structure of NfsA was published in 2001 and exhibits an $\alpha + \beta$ homodimeric fold, including the nitroreductase superfamily minimal scaffold (Kobori *et al.*, 2001; Akiva *et al.*, 2017). The monomeric subunits of NfsA interact in a head-to-head/tail-to-tail conformation forming two active sites per dimer at the dimeric interface, with residues contributing from both monomers (Figure 1.5). In contrast to NfsB family nitroreductases, NfsA exhibits a strong preference for NADPH over NADH (Zenno *et al.*, 1996). The structures of NfsA and other nitroreductases are believed to remain rigid during catalysis. Residues in the active site are not believed to play a direct role in substrate reduction, rather involved in stabilisation of the active site pocket (Williams, 2013; Pitsawong *et al.*, 2014). In order for catalysis to proceed, the 2'-phosphate group of NADPH needs to interact with the active site residue R203, and the nicotinamide ring of NADPH must come into close proximity with the isoalloxazine ring of FMN to allow for efficient hydride transfer (Kobori *et al.*, 2001).

1.4.2 Native function of NfsA

Very few nitroaromatic compounds exist in nature. Known examples include chloramphenicol, nitropyoluteorin, oxypyrronitritin and phidolopin (Roldan *et al.*, 2008). Most nitroaromatic compounds are artificially synthesised either for medical purposes (i.e., CB1954 and metronidazole), or for other applications, such as use in explosives (i.e., 2,4,6-trinitrotoluene and 2,4,6-trinitrophenol). The rarity of nitroaromatic compounds in nature suggests that they are unlikely the primary substrate for these enzymes. A biological role for NfsA and other nitroreductases as quinone reductases has been suggested (Liochev *et al.*, 1999; Wang and Maier, 2004; Green *et al.*, 2014; Copp *et al.*, 2017). Two-electron reduced quinones (quinols) are able to scavenge and detoxify reactive oxygen species, thereby maintaining a pool of reduced quinones, contributing to the antioxidant defence of a cell (Soballe and Poole, 1999). However, the highly promiscuous nature of NfsA and other nitroreductases makes it hard to conclusively define the native function of these enzymes.

1.4.3 Promiscuous activities of NfsA

A truly promiscuous activity is defined as a secondary reaction of an enzyme which is physiologically irrelevant to the organism (Copley, 2017). Although we can only predict the primary function of NfsA, we can confidently state it is a truly promiscuous enzyme due to the artificial and synthetic nature of the prodrugs it can reduce. Moreover, most

nitroaromatic compounds become more toxic following nitro-reduction in turn killing the cell, making it hard to understand why this would have evolved as a native or primary function. In addition to the promiscuous activities already noted (i.e., reduction of CB1954, metronidazole, tinidazole, RB6145, azoymcin and misonidazole), some additional promiscuous activities of direct relevance to this study are introduced in Section 1.4.3.1 below.

1.4.3.1 Nitrofurazone, nitrofurantoin and 2,4-dinitrotoluene (2,4-DNT)

Figure 1.6: Chemical structures of the prodrugs nitrofurazone, nitrofurantoin and 2,4-DNT.

Nitrofurazone and nitrofurantoin are broad-spectrum antibiotics with similar chemical structures (Figure 1.6). As early as the 1950s it was shown that the sensitivity of bacterial strains to nitrofurazone was the result of a flavoprotein catalysing the reduction of the drug to a cytotoxic compound (Asnis, 1957). A large number of nitroreductases from many different bacterial species have been reported to catalyse the reduction of nitrofurazone (Bryant *et al.*, 1981; Bryant and DeLuca, 1991; Zenno *et al.*, 1998; Kwak *et al.*, 2003; Olekhovich *et al.*, 2009). This is likely due to the relatively low one-electron reduction potential ($E^{\prime}_7$) of nitrofurazone. The name NfsA is derived from its ability to catalyse the reduction of nitrofurans, standing for Nitrofuran sensitive gene A (McCalla *et al.*, 1978). Both nitrofurazone and nitrofurantoin are nil-bystander prodrugs and therefore could also have potential as agents in targeted cell ablation (Matin *et al.*, 2010)

2,4-DNT is a polynitrated compound with a relatively simple structure consisting of a benzene ring with two nitro-groups and one methyl-group (Figure 1.6). This compound is the precursor to trinitrotoluene (TNT), however is mainly used in the production of toluene diisocyanate, for generation of polyurethane foams (Dent *et al.*, 1981). Residual TNT and 2,4-DNT have been found as environmental pollutants near manufacturing plants and land mines (Gong *et al.*, 1999; Jenkins *et al.*, 2001). Although 2,4-DNT is not considered a

prodrug, previous members of the Ackerley lab group have shown it can be treated like one in IC₅₀ assays to quantify the ability of a nitroreductase to reduce this compound (Rich, 2017).

1.4.3.2 Niclosamide and chloramphenicol

Figure 1.7: Chemical structures of the nitroaromatic compounds niclosamide and chloramphenicol.

The nitroaromatic compound *N*-(2'-chloro-4'-nitrophenyl)-5-chlorosalicylamide, more commonly known as niclosamide (Figure 1.7), was first used clinically as an anthelmintic to kill parasitic worms through the uncoupling of oxidative phosphorylation (Pearson and Hewlett, 1985; Frayha *et al.*, 1997). Niclosamide is effective against Gram-positive bacteria such as *Staphylococcus* and *Clostridium* species (Rajamuthiah *et al.*, 2015; Gooyit and Janda, 2016). Niclosamide has also recently been shown to be effective against the Gram-negative bacterium *Helicobacter pylori*, by changing the transmembrane pH and disrupting the proton motive force (Tharmalingam *et al.*, 2018). Previous work in the Ackerley lab group has also shown niclosamide is selectively toxic to *E. coli* cells lacking the drug efflux pump TolC (Rich, 2017).

Chloramphenicol is a broad-spectrum antibiotic (Figure 1.7), first isolated from *Streptomyces venezuelae* (Akagawa *et al.*, 1975). This compound is effective against both Gram-positive and Gram-negative bacteria and works by binding to the 50S ribosomal subunit, inhibiting the peptidyltransferase step during protein synthesis (Rossolini, 2017). It was first reported in the 1980s that certain bacterial species might be able to evade the toxicity of chloramphenicol through nitro-reduction (Yunis, 1988). More recently it has been proven that a nitroreductase from *Haemophilus influenzae* is capable of conferring

chloramphenicol resistance through reduction of the nitro-group (Smith *et al.*, 2007; Crofts *et al.*, 2019; Green, 2019).

Unlike the other nitroaromatic compounds presented earlier, niclosamide and chloramphenicol are toxic prior to nitro-reduction. Previous work in the Ackerley lab has shown that both niclosamide and chloramphenicol can be used as positive selection tools to recover active nitroreductase variants. Overexpression of nitroreductases from a range of families was shown to provide a protective effect against these compounds (Little, 2015; Rich, 2017). A clear correlation has been found between the enhanced detoxification of these positive selection compounds and improvement in the nitroreductase activity with a number of nitroaromatic prodrugs. Plating *nfsA* variant libraries on low concentrations of niclosamide or chloramphenicol has proved an efficient selection, enabling recovery of active nitroreductase variants while removing inactive variants (Little, 2015; Rich, 2017).

1.4.3.3 Promiscuous activities and metabolic interference

Although promiscuous activities can provide a starting point for directed evolution experiments, they can also be a hindrance due to their off-target effects (Pitera *et al.*, 2007; Bantscheff *et al.*, 2009; Kim and Copley, 2012). This phenomenon is known as metabolic interference and occurs when metabolites or enzymes interfere with metabolic pathways they are not normally involved in (Kim and Copley, 2012). This can become a problem when an essential metabolite or pathway is inhibited, having potentially lethal consequences on a cell. Promiscuous enzymes are often a source of these adventitious interactions and can cause problems when exposed to molecules not previously encountered. The number of promiscuous activities a single enzyme can catalyse is vast. One study from 2015 looked at the activities of 217 members of the haloacid dehalogenase family with 169 different phosphorylated compounds. Fifty of the enzymes were found to be broadly promiscuous, acting on more than 41 substrates, with one enzyme able to act on 143 different substrates (Huang *et al.*, 2015). This study emphasises how widespread promiscuous activities are, and the numerous opportunities for metabolic interference. Directed evolution of promiscuous enzymes to improve these promiscuous activities only exacerbates this problem and this is believed to be key reason why many evolved enzymes do not function as anticipated when introduced into an existing metabolic network.

In addition to the biotechnological applications of the promiscuous activities of NfsA, these activities can also be of interest from an evolutionary standpoint. We want to use NfsA to gain a better understanding of how promiscuous activities can be evolved towards a user-defined goal, and to improve our understanding of the off-target effects of directed evolution.

1.5 Aims

The research presented in this study sought to use simultaneous site-directed mutagenesis of *nfsA* to identify NfsA variants exhibiting superior nitro-reduction of a range of nitroaromatic prodrugs. These NfsA variants were evaluated in both fundamental and applied studies.

The primary aims of this study were:

- Create a simultaneous site-directed mutagenesis library of *nfsA* to identify variants exhibiting superior detoxification of chloramphenicol or niclosamide.
- Evaluate different NfsA variants to improve our understanding of evolutionary trajectories and how promiscuous activities can interfere with directed evolution.
- Develop improved methods for targeted cell ablation in zebrafish through:
  - a) Identifying NfsA variants exhibiting improved activation with metronidazole, tinidazole, RB6145 and misonidazole.
  - b) Creation of a multiplex system to enable independent ablation of two different cell populations within a zebrafish.

Chapter 2 Methods

2.1 Chemicals reagents, enzymes and antibodies

RB6145 (Jenkins *et al.*, 1990) was synthesised by Dr Gary Evans and Dr Ralf Schwoerer for use in this study (*VUW Ferrier Institute*). The fluoroprobe FSL61 (Su *et al.*, 2013) was synthesised by chemists at the Auckland Cancer Society Research Centre and provided for use in this study by Dr Jeff Smaill and Dr Adam Patterson (*University of Auckland*). CB1954 was purchased from MedKoo Biosciences Inc. (*Morrisville, NC, USA*). Additional nitroaromatic compounds and quinones (metronidazole, tinidazole, nitrofurazone, nitrofurantoin, misonidazole, 2,4-dinitrotoluene, niclosamide, chloramphenicol and *p*-benzoquinone) were purchased from Sigma-Aldrich (*St Louis, MO, USA*). All drug stocks were made up in 100% anhydrous dimethyl sulfoxide (DMSO) and stored at -80 °C. All other chemicals and reagents were purchased from Sigma Aldrich (*St Louis, MO, USA*), Thermo Fisher Scientific (*Waltham, MA, USA*) or Total Lab Systems (*Auckland, NZ*) unless otherwise stated. Restriction enzymes were purchased from Thermo Fisher Scientific (*Waltham, MA, USA*). BioMix^{TM} Red Polymerase Mastermix was purchased from Bioline (*London, UK*). Agilent Technologies GeneMorph® II Random Mutagenesis Kit was purchased from Integrated Sciences (*Sydney, NSW, Australia*). Q5® High Fidelity 2 × Master Mix and T4 DNA ligase were purchased from New England Biolabs (*Ipswich, MA, USA*). Bugbuster® HT Protein Extraction Reagent was purchased from Merck (*Kenilworth, NJ, USA*). The sheep anti-NfsA antibody was donated by Dr Peter Searle for use in this study (*University of Birmingham; UK*). All other antibodies were purchased from Abcam (*Cambridge, UK*).

2.2 Synthetic genes and oligonucleotide primers

All synthetic genes used in this study were ordered from Twist Bioscience (*San Francisco, CA, USA*) and provided as lyophilised powders. Oligonucleotide primers used in this study were supplied as lyophilised powders by Integrated DNA Technologies (*IDT; Coralville, IA, USA*). Primers were resuspended to 100 µM in TE buffer (10 mM Tris.HCl pH 8.0, 0.1 mM EDTA) to generate stock solutions and 10 µM working solutions were diluted in sterile ddH₂O. Stock and working solutions were stored at -20 °C. The sequences of all primers used in this study are listed in Table 2.1.

Primer Name	Sequence 5' → 3' (underlined text indicates restriction sites used for cloning)	Function of Primer
NfsA_Ec_fwd	GGCATATGACGCCAACCATTGAAC	Amplifying NfsA_Ec
NfsA_Ec_rev	GGGTCGACTTAGCGCGTCGCCCAACCCTG	Amplifying NfsA_Ec
NfsA_Ec_GW_fwd	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGA AGGAGATAGAACCATGGGCACGCCAACCATTGAACT	Construction of Gateway constructs
NfsA_Ec_GW_rev	GGGGACCACTTTGTACAAGAAAGCTGGGTCCTAG CGCGTCGCCCAACCCTG	Construction of Gateway constructs
pTOL2_fwd	AGCGGAGCTACTAACTTC	Amplifying inserts from the pTOL2 vector
F279V5-fwd	CCAGTTCAATTACAGCTC	Amplifying inserts from the F279_V5 vector
M13_fwd	GTAAAACGACGGCCA	Amplifying inserts from the pDONR221 vector
NfsA_Ec_Xmal	CCCCCGGGCCTATGACGCCAACC	Introduce Xmal restriction site for pTOL2 cloning
pUCX_fwd	GACATCATAACGGTTCTG	Amplifying inserts from the pUCX vector
pUCX_rev	GTTTCACTTCTGAGTTCTG	Amplifying inserts from the pUCX vector
pET28a+_fwd	AGATCTCGATCCCGCGA	Amplifying inserts from the pET28a+ vector

Table 2.1: Oligonucleotide primers used in this study.

2.3 Bacterial strains, mammalian cell lines and plasmids used in this study

2.3.1 Bacterial strains

All *E. coli* expression and screening strains used in this study are listed in Table 2.2.

Strains	Description	Source
7NT	W3110 Δ nfsA Δ nfsB Δ yeiF Δ ycaK Δ mdaB Δ azoR Δ nemA Δ tolC	Copp <i>et al.</i> , (2014b)
SOS-R2	ADA-510 Δ nfsA Δ nfsB Δ azoR Δ nemA Δ tolC	Prosser <i>et al.</i> , (2013)
BL21(DE3)	F ⁻ ompT gal dcm lon hsdSB(r _B ⁻ m _B ⁻) λ (DE3)	Novagen (Kenilworth, NJ, USA)
EC100™	F ⁻ mcrA Δ (mrr-hsdRMS-mcrBC) Φ 80dlacZ Δ M15 Δ lacX74 recA1 endA1 araD139 Δ (ara, leu)7697 galU galK λ - rpsL (Str ^R) nupG	Lucigen (Middleton, WI, USA)

Table 2.2: *E. coli* strains used in this study.

2.3.2 Mammalian cell lines

HEK-293, an adherent human embryonic kidney cell line was purchased from ATCC (American Type Culture Collection; Manassas, VA, USA).

2.3.3 Plasmids

All plasmids used in this study are listed in Table 2.3.

Plasmid	Resistance	Description	Source
pUCX	Ampicillin	Nitroreductase expression vector used in library creation.	Prosser <i>et al.</i> , (2010)
pET28a+	Kanamycin	Expression vector for His ₆ -tagged enzyme purification	Novagen (Kenilworth, NJ, USA)
pDONR221	Kanamycin	Entry vector for Gateway recombination cloning	Invitrogen (Carlsbad, CA, USA)
F279_V5	Ampicillin/ Puromycin	Destination vector for Gateway recombination cloning. Mammalian expression vector.	Prosser <i>et al.</i> , (2013)
pTOL2	Ampicillin	Zebrafish expression vector	Dr Jeff Mumm (Wilmer Eye Institute, Johns Hopkins University)

Table 2.3: Plasmids used in this study

2.4 Bacterial growth and maintenance

2.4.1 Growth media

Unless otherwise stated, all growth media was prepared in ddH₂O, autoclaved at 121 °C for 15 minutes and stored at room temperature. All liquid media recipes are listed in Table 2.4. Solid media was prepared by the addition of 1.5% agar to lysogeny broth (LB) medium prior to autoclaving. Prior to pouring plates, media was cooled to < 50 °C and supplemented with the appropriated antibiotic.

Media	Components
LB Low Salt ^a	10 g.L ⁻¹ Tryptone 5 g.L ⁻¹ Yeast extract 5 g.L ⁻¹ NaCl
M9 Minimal Media	1 x M9 salts ^a 2 mM MgSO ₄ ^b 100 µM CaCl ₂ ^b 0.4% Glucose ^b
Super Optimal Media with Catabolite Repression (SOC)	20 g.L ⁻¹ Tryptone 5 g.L ⁻¹ Yeast extract 10 mM NaCl 2.5 mM KCl 10 mM MgCl ₂ ^b 10 mM MgSO ₄ ^b 0.2% Glucose ^b

^a = LB and M9 salts were purchased as a premixed powder

^b = Filter-sterilised and added post-autoclave.

Table 2.4 Liquid growth media recipes.

2.4.2 Bacterial antibiotics and induction agents

Stock solutions of antibiotics were prepared in ddH₂O at 1000 × the final concentration and filter sterilised using a 0.22 µm filter. Aliquots were stored at -20 °C. Ampicillin (Amp)

and kanamycin were the two antibiotics used in this study and were used at a final concentration of 100 $\mu\text{g.mL}^{-1}$. Isopropyl β -D-1-thiolactopyranoside (IPTG) was used for the induction of genes from the *lac* operator. IPTG was made up in ddH₂O to a concentration of 420 mM (10% w/v), filter sterilised using a 0.22 μM filter and stored in aliquots at -20 °C.

2.4.3 Growth and storage

Unless otherwise stated, bacteria inoculated in liquid media were grown at 37 °C at 200 rpm and bacteria inoculated solid media were grown at 37 °C in a static incubator. For long-term storage, liquid cultures were mixed 1:1 with autoclaved 80% (w/v) glycerol and stored indefinitely at -80 °C.

2.5 Standard molecular biology techniques

2.5.1 Isolation of plasmid DNA

Isolation of plasmid DNA was carried out using a standard protocol. DNA spin columns were purchased from Epoch Life Sciences (*Missouri City, Texas, USA*). Buffers used in the isolated of plasmid DNA are listed in Table 2.5.

Buffer	Components
P1 (Resuspension Buffer)	50 mM Tris.HCl pH 8.0 10 mM EDTA 100 $\mu\text{g.mL}^{-1}$ RNase A
P2 (Cell Lysis Buffer)	200 mM NaCl 1% SDS
N3 (Neutralisation Buffer)	4.2 M guanidine hydrochloride 0.9 M potassium acetate Add glacial acetic acid to pH 4.8
PB Wash Buffer	5 M guanidine hydrochloride 30% isopropanol
PE Wash Buffer	10 mM Tris.HCl pH 7.5 80% ethanol
Elution Buffer	10 mM Tris.HCl pH 8.5

Table 2.5: Buffer recipes used in the isolation of plasmid DNA.

2.5.2 Polymerase chain reaction (PCR) protocol

Biomix Red^{TM} was used for colony screening, Q5® High-Fidelity 2 × Master Mix was used for amplifying DNA for cloning purposes and the GeneMorph® II Random Mutagenesis Kit was used for the creation of error-prone gene libraries. To prevent mutational bias occurring in the error-prone PCR (when a mutation introduced early in the reaction can become overrepresented in the final library), the reactions were split into multiple tubes

prior to starting the reaction. The reaction components and the PCR cycle parameters are outlined in Table 2.6, Table 2.7 and Table 2.8.

Component	Final Concentration	Temperature	Time
DNA template	10-50 ng ^a	95 °C	5 min
Forward primer	0.5 µM	95 °C	15 sec
Reverse primer	0.5 µM	T _A ^b °C	30 sec/kb
Biomix Red™	1 ×	72 °C	1 min/kb
Nuclease free H ₂ O	To final volume	72 °C	5 min

^a A sterile pipette tip was used to inoculate the PCR master mix with a colony sample

^b T_A calculated as the lowest primer T_M

Table 2.6: Biomix Red™ PCR components and parameters.

Component	Final Concentration	Temperature	Time
DNA template	1-25 ng	98 °C	30 sec
Forward primer	0.5 µM	98 °C	10 sec
Reverse primer	0.5 µM	T _A ^a °C	30 sec/kb
Q5® High Fidelity Master Mix	1 ×	72 °C	30 sec/kb
Nuclease free H ₂ O	To final volume	72 °C	2 min

^a T_A calculated as the lowest primer T_M

Table 2.7: Q5® High Fidelity Master Mix PCR components and parameters.

Component	Final Concentration	Temperature	Time
DNA template	5 ng ^a	95 °C	2 min
Forward primer	0.25 µM	95 °C	30 sec
Reverse primer	0.25 µM	51 °C	30 sec
Mutazyme II Polymerase	2.5 U/50 µL	72 °C	1 min
10 x Mutazyme II Buffer	1 ×	72 °C	10 min
dNTPs	800 µM		
Nuclease Free H ₂ O	To final volume		

^a DNA template for error-prone PCR was amplified using the standard Q5® PCR protocol. The amount of DNA template was used to control the error rate.

Table 2.8: GeneMorph® II PCR components and parameters.

2.5.3 Restriction enzyme digest

Restriction digests were carried out as per manufacturer's instructions. For vector preparation, where the digested fragment was < 100 bp, the digested vector was purified using a DNA Clean and Concentrator-5 Kit (Section 2.5.4.1). If the digested fragment was larger, the digested vector was gel extracted (Section 2.5.4.2). Digested vector was stored in aliquots at -20 °C to minimise freeze/thaw cycles.

2.5.4 DNA product purification

2.5.4.1 DNA cleaning and concentrating

PCR and restriction digest products were purified using a DNA Clean and Concentrator-5 Kit (*Zymo Research; Irvine, CA, USA*) as per manufacturer's instructions.

2.5.4.2 Gel extraction

Vector restriction digest preparations containing a digested fragment > 100 bp were run on an agarose gel containing 1 × SYBR^{TM} Safe (*Thermo Fisher Scientific; Waltham, MA, USA*) alongside a HyperLadder^{TM} 1 kb molecular weight marker (*Bioline; London, UK*). DNA bands were visualised under blue light and DNA bands of the correct size were excised using a clean scalpel, minimising the amount of agarose gel not containing any DNA. DNA was then purified using the QIAquick Gel Extraction Kit (*Qiagen; Hilden, Germany*) as per manufacturer's instructions.

2.5.5 Agarose gel electrophoresis

For the analysis of DNA samples, aliquots of DNA were mixed with DNA loading dye and loaded into 1% agarose gels (1% (w/v) agarose in 1 × TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.3), 1 µg.mL⁻¹ ethidium bromide) and run in 1 × TAE buffer at 135 V for 30 minutes. DNA bands were visualised under ultra violet (UV) light and size was determined by comparison to a HyperLadder^{TM} 1 kb molecular weight marker (*Bioline; London, UK*).

2.5.6 Ligation

T4 DNA ligase (*New England Biolabs; Ipswich, MA, USA*) was used to ligate DNA fragments as outlined in Table 2.9. Ligation reactions were incubated at 16 °C overnight.

Component	Final Concentration
Digested vector DNA (50 ng)	1-10 µg.mL ⁻¹
Digested insert DNA (6:1 ratio over vector)	
5 x ligase buffer	1 ×
Ligase	1 U per 10 µL
ddH ₂ O	Up to 10 µL

Table 2.9: Ligation components

2.5.6.1 tRNA assisted DNA precipitation

When a ligation was used for library creation it was precipitated with yeast tRNA using a protocol modified from Zhu and Dean, (1999). The ligation was added to 1 mg.mL⁻¹ yeast tRNA, 0.5 M sodium acetate pH 5.2 and 1 volume of isopropanol and centrifuged at

21,000 × g for 20 minutes at 4 °C. The supernatant was removed and the pellet was washed with 1 mL ice-cold 70% ethanol, then centrifuged at 21,000 × g for 10 minutes at 4 °C. The supernatant was removed and the pellet air-dried for 5 minutes. The air-dried pellet was resuspended in ddH₂O in half the starting ligation volume.

2.5.7 DNA quantification

The purity and concentration of DNA products was determined using a NanoPhotometer® spectrophotometer (*Implen; München, Germany*), according to the manufacturer's instructions.

2.5.8 DNA sequencing and analysis

DNA sequencing was carried out by Macrogen Inc. (*Seoul, South Korea*). Purified plasmids were sent as per recommendations by Macrogen to determine the sequence of newly cloned variants and any mutants of interest. All sequences were analysed using SnapGene® 4.1.9 (*Chicago, IL, USA*).

2.5.9 Transformation of *E. coli*

2.5.9.1 Preparation of chemically competent cells

An overnight culture of the *E. coli* strain to be made competent was grown in 2 mL of LB containing any relevant antibiotics at 37 °C, 200 rpm overnight. A 100-fold dilution of the overnight culture was added to a 100 mL flask of LB containing any relevant antibiotics and grown at 37 °C, 200 rpm until an OD₆₀₀ of 0.4-0.6 was reached, after which the culture was chilled on ice for 10 minutes. Half of the chilled culture was transferred into a pre-chilled 50 mL centrifuge tubes and cells were pelleted by centrifugation at 2500 × g for 10 minutes at 4 °C. The supernatant was discarded and this process was repeated with the remaining culture. Once all the culture had been pelleted, the cell pellet was resuspended in 50 mL ice-cold 100 mM MgCl₂ then cells were pelleted by centrifugation at 2500 × g for 10 minutes at 4 °C. The supernatant was discarded and the cell pellet was resuspended in 50 mL ice-cold 100 mM CaCl₂, then cells were pelleted by centrifugation at 2500 × g for 10 minutes. The supernatant was removed and the cell pellet resuspended in 2 mL of ice-cold 85 mM CaCl₂, 15% glycerol. Cells were aliquoted into single use (10 µL) or multiple use (100 µL) aliquots. Aliquots of cells were snap-frozen and stored at -80 °C until required.

2.5.9.2 Transformation of chemically competent cells

Chemically competent cells were transformed with plasmid DNA or ligation for high-throughput cloning. In the case of high-throughput cloning, this protocol used PCR strip tubes or a 96-well PCR plate and a PCR machine. Frozen aliquots of cells were thawed on ice. Plasmid or ligation ($5 \text{ ng} \cdot \mu\text{L}^{-1}$ in a $1 \mu\text{L}$ volume) was added to individual aliquots of cells and incubated on ice for 10 minutes. Cells were heat shocked by incubating at 42°C for 90 seconds, followed by incubation on ice for 60 seconds. Following heat shock, cells were recovered in nine times the volume of LB medium at 37°C for 1 hour, then spread on solid medium containing the appropriate antibiotics. As a guideline, when using $10 \mu\text{L}$ aliquots of cells, when cells were transformed using plasmid, 20% of the resulting recovery culture was plated and the entire recovery culture was plated when cells were transformed using ligation.

2.5.9.3 Preparation of electrocompetent cells

An overnight culture of the *E. coli* strain to be made competent was grown in 5 mL of LB containing any relevant antibiotics at 37°C , 200 rpm overnight. A 100-fold dilution of the overnight culture was added to two 400 mL flasks of LB containing any relevant antibiotics and grown at 37°C , 200 rpm until an OD_{600} of 0.4-0.6 was reached, at which point the culture was chilled on ice for 20 minutes. Half of the chilled culture was transferred into eight pre-chilled 50 mL centrifuge tubes (400 mL total) and cells were pelleted by centrifugation at $2500 \times g$ for 10 minutes at 4°C . The supernatant was decanted and this process was repeated with the remaining culture. Once all the culture had been pelleted, the resulting cell pellets were gently resuspended in 25 mL of ice-cold sterile ddH_2O and pooled into four tubes. The resuspended cells were pelleted by centrifugation at $2500 \times g$ for 10 minutes at 4°C . The supernatant was decanted and the cell pellets were gently resuspended in 25 mL ice-cold 10% (v/v) glycerol and pooled into two tubes. The resuspended cells were pelleted by centrifugation at $2500 \times g$ for 15 minutes at 4°C . The supernatant was decanted and the cell pellets were gently resuspended in 25 mL ice-cold 10% (v/v) glycerol and pooled into one tube. The resuspended cells were pelleted by centrifugation $2500 \times g$ for 20 minutes at 4°C . The supernatant was decanted and the cell pellet was gently resuspended in residual 10% (v/v) glycerol plus an additional $\sim 400 \mu\text{L}$ of ice-cold 10% (v/v) glycerol, resulting in a final volume of $\sim 1200 \mu\text{L}$. Single ($40 \mu\text{L}$) or double ($80 \mu\text{L}$) aliquots of cells were snap-frozen using a metal tube block cooled to -80°C , and stored at -80°C .

2.5.9.4 Transformation of electrocompetent cells

Aliquots of frozen cells were thawed on ice. DNA (0.5-1 μ L of ligation or diluted plasmid) was added to 5 μ L of ddH₂O and gently mixed with 40 μ L of cells to ensure the cells were evenly mixed with the DNA. The cell/DNA mixture was moved into 2 mm gap electroporation cuvettes and electroporated using a Bio-Rad Gene Pulser Xcell^{TM} (Hercules, CA, USA), pulsed at 2.5 kV, 25 μ F, 200 Ω . Following electroporation, 1 mL of SOC media was immediately added to the cells and the cells were transferred into a 15 mL tube. An additional 1 mL of SOC media was added to the electroporation cuvette and transferred into the 15 mL tube to ensure as many cells as possible were removed from the cuvette. Cells were recovered at 37 °C, 200 rpm for 1 hour, after which dilutions were plated on agar plates supplemented with the appropriate antibiotics. The first time a batch of electrocompetent cells was tested, a negative control containing no DNA was included to ensure the cells were not contaminated.

2.6 Library Creation

2.6.1 Small-scale library creation (<5,000,000 clones)

Cells were transformed as per Section 2.5.9.4. Following recovery of cells in SOC media, all transformations were pooled and 0.01 μ L, 0.1 μ L and 1 μ L samples were plated in triplicate on agar plates supplemented with the appropriate antibiotic and incubated at 37 °C overnight. These plates were used to estimate the total number of clones in a library. The remaining culture was plated over multiple large agar plates (150 mm) supplemented with the appropriate antibiotic. A volume was plated which allowed for individual colonies to be formed (a single transformation was normally tested the prior day to determine this volume). Plates were incubated at 37 °C overnight. The following morning cells were pooled together by pipetting 2 mL of LB onto each plate and scrapping the colonies off using a glass plate spreader. The resulting culture was pelleted at 2500 $\times$ g for 30 minutes at 4 °C to create a concentrated library pool. The supernatant was discarded and the resulting pellet was gently resuspended in a small volume of LB. The library culture was mixed 1:1 with 80% (v/v) glycerol, aliquoted and stored at -80 °C indefinitely. The OD₆₀₀ of a 1 in 100 dilution of the library was measured and was used to determine the density of cells in the final library (an OD₆₀₀ of 1.0 was estimated to contain 1×10^9 cells).

2.6.2 Large-scale library creation (>5,000,000 clones)

Cells were transformed as per Section 2.5.9.4. Following recovery of cells in SOC media, all transformations were pooled and 0.01 μL , 0.1 μL and 1 μL samples were plated in triplicate on agar plates supplemented with the appropriate antibiotic and incubated at 37 °C overnight. These plates were used to estimate the total number of clones in a library. The remaining culture was transferred to a suitable sized flask containing a 30-fold larger volume of LB medium than the recovery culture being added (i.e., 10 mL of recovery culture into 300 mL of LB medium) and grown at 37 °C, 200 rpm for 6 hours to propagate the library. Following this, 10% of the resulting culture was pelleted by centrifugation at $2500 \times g$ for 30 minutes at 4 °C. The supernatant was decanted and the cell pellet was gently resuspended in the residual LB medium. The resulting culture was mixed 1:1 with 80% (v/v) glycerol and stored indefinitely at -80 °C. The OD_{600} of a 1 in 100 dilution of the library was measured (an OD_{600} of 1.0 was estimated to contain 1×10^9 cells) and was used to determine the density of cells in the final library.

2.6.3 Selection of nitroreductase libraries using positive selection compounds

A sample of a frozen library stock (20 μL) was resuspended in 1 mL of LB medium. Using this cell suspension, six 10-fold serial dilutions were prepared, transferring 100 μL into 900 μL of LB medium sequentially. The concentrated resuspension or the 1 in 10 dilution was plated on agar plates supplemented with 50 μM IPTG, appropriate antibiotic and the desired concentration of niclosamide and/or chloramphenicol. In some cases, IPTG was excluded from the media when the IPTG appeared toxic to the growth of the cells. Plates containing niclosamide were incubated at 37 °C overnight (~16 hours) and plates containing chloramphenicol were incubated for 40 hours. To estimate the number of cells plated on the selection plate, 100 μL from dilution 5 and 6 (1 in 10,000 and 1000,000 dilutions) were plated on agar containing the appropriate antibiotics and incubated at 37 °C overnight. Colonies which grew on selective media were picked into the inner 60-wells of 96-well plates containing LB supplemented with the appropriate antibiotic and 20% glucose (w/v) and incubated at 30 °C overnight. These plates were supplemented 1:1 with 80% glycerol (v/v) and stored at -80 °C indefinitely. These plates were subjected to further screening with specific prodrugs using growth inhibition and IC_{50} assays.

2.7 Nitroreductase candidate library

The Ackerley lab nitroreductase library comprises 58 oxidoreductases from a variety of bacterial species (Table 2.10). Prior to this project, each oxidoreductase gene had been PCR amplified from gDNA, cloned into the pUCX expression vector and used to transform the *E. coli* 7NT strain (Prosser *et al.*, 2013; Williams, 2013; Rich, 2017). This nitroreductase library was used as a source of alternative nitroreductases in this study. Throughout this thesis, each oxidoreductase is referred to by the family name followed by a two- or three-letter abbreviation from the genus and species of the bacteria that the enzyme was derived from.

Gene	Organism	Family	Super family	GenBank Accession #
NfsA_Ec	<i>Escherichia coli</i>	NfsA	Nitro FMN reductase	APC51123.1
NfsA_St	<i>Salmonella enterica</i> serovar Typhimurium	NfsA	Nitro FMN reductase	WP_000075306.1
NfsA_Ck	<i>Citrobacter koseri</i>	NfsA	Nitro FMN reductase	WP_012133106.1
NfsA_Kp	<i>Klebsiella pneumoniae</i>	NfsA	Nitro FMN reductase	WP_004179133.1
NfsA_Cs	<i>Cronobacter sakazakii</i>	NfsA	Nitro FMN reductase	WP_007848889.1
NfsA_Eca	<i>Pectobacterium atrosepticum</i> (previously <i>Erwinia carotovora</i>)	NfsA	Nitro FMN reductase	WP_039295639.1
NfsA_Vf	<i>Aliivibrio fischeri</i> (previously <i>Vibrio fischeri</i>)	NfsA	Nitro FMN reductase	WP_054775176.1
NfsA_Vv	<i>Vibrio vulnificus</i>	NfsA	Nitro FMN reductase	WP_017421853.1
Frp_Vh	<i>Vibrio harveyi</i>	NfsA	Nitro FMN reductase	WP_038898257.1
CO-Frp_Vh	<i>Vibrio harveyi</i>	NfsA	Nitro FMN reductase	Q56691.1
NfrA_Bs	<i>Bacillus subtilis</i>	NfsA	Nitro FMN reductase	WP_003222161.1
NfsA_Lw	<i>Listeria welshimeri</i>	NfsA	Nitro FMN reductase	WP_011701746.1
NfsA_Li	<i>Listeria innocua</i>	NfsA	Nitro FMN reductase	WP_003761397.1
NfsA_Bc	<i>Bacillus coagulans</i>	NfsA	Nitro FMN reductase	WP_029141841.1
NfsA_Np	<i>Nostoc punctiforme</i>	NfsA	Nitro FMN reductase	ACC79374.1
NfsA_Bt	<i>Bacillus thuringiensis</i>	NfsA	Nitro FMN reductase	WP_070179995.1
NfsA_Ls	<i>Lactobacillus sakei</i>	NfsA	Nitro FMN reductase	WP_035146643.1
EcD_Pp	<i>Pseudomonas putida</i>	NfsA	Nitro FMN reductase	WP_010953425.1
YcnD_Bs	<i>Bacillus subtilis</i>	NfsA	Nitro FMN reductase	WP_003225013.1
NfsA_Ms	<i>Mycobacterium smegmatis</i>	NfsA	Nitro FMN reductase	WP_011728689.1
YcnD_Bli	<i>Bacillus licheniformis</i>	NfsA	Nitro FMN reductase	WP_003178951.1
NfrA_Bli	<i>Bacillus licheniformis</i>	NfsA	Nitro FMN reductase	WP_003186165.1
NfsB_Ec	<i>Escherichia coli</i>	NfsB	Nitro FMN reductase	CQR80176.1
NfsB_Ck	<i>Citrobacter koseri</i>	NfsB	Nitro FMN reductase	WP_047458272.1
NfsB_St	<i>Salmonella enterica</i> serovar Typhimurium	NfsB	Nitro FMN reductase	WP_000355874.1
NfsB_Kp	<i>Klebsiella pneumoniae</i>	NfsB	Nitro FMN reductase	WP_004178896.1
NfsB_Vv	<i>Vibrio vulnificus</i>	NfsB	Nitro FMN reductase	WP_017421465.1
NfsB_Pp	<i>Pseudomonas putida</i>	NfsB	Nitro FMN reductase	WP_010953384.1
NfsB_Cs	<i>Cronobacter sakazakii</i>	NfsB	Nitro FMN reductase	WP_004384956.1
Frasel_Vf	<i>Aliivibrio fischeri</i>	NfsB	Nitro FMN reductase	P46072.3

Gene	Organism	Family	Super family	GenBank Accession #
NfsB_Vh	<i>Vibrio harveyi</i>	NfsB	Nitro FMN reductase	WP_005435698.1
YfkO_Bs	<i>Bacillus subtilis</i>	NfsB	Nitro FMN reductase	ADE73858.1
YfkO_Bli	<i>Bacillus licheniformis</i>	NfsB	Nitro FMN reductase	WP_003179779.1
NfsB_Nme	<i>Neisseria meningitidis</i>	NfsB	Nitro FMN reductase	WP_002225416.1
Ydgl_Bs	<i>Bacillus subtilis</i>	NTR 1	Nitro FMN reductase	WP_003225379.1
NfsB_Pa	<i>Pseudomonas aeruginosa</i>	NTR 1	Nitro FMN reductase	WP_003100063.1
AzoR_Ec	<i>Escherichia coli</i>	AzoR	FMN reductase	WP_000048950.1
AzoR_St	<i>Salmonella enterica</i>	AzoR	FMN reductase	WP_000048924.1
AzoR_Vv	<i>Vibrio vulnificus</i>	AzoR	FMN reductase	WP_017421098.1
AzoR_Pp	<i>Pseudomonas putida</i>	AzoR	FMN reductase	ABQ77513.1
NemA_Ec	<i>Escherichia coli</i>	OYE-like	TIM phosphate binding	WP_000093589.1
NemA_St	<i>Salmonella enterica</i>	OYE-like	TIM phosphate binding	WP_000092935.1
NemA_Kp	<i>Klebsiella pneumoniae</i>	OYE-like	TIM phosphate binding	WP_004148597.1
NemA_Vv	<i>Vibrio vulnificus</i>	OYE-like	TIM phosphate binding	WP_017421369.1
YwrO_Bs	<i>Bacillus subtilis</i>	MdaB	FMN reductase	WP_003219573.1
YwrO_Li	<i>Listeria innocua</i>	MdaB	FMN reductase	WP_003772432.1
YwrO_Bam	<i>Bacillus amyloliquefaciens</i>	MdaB	FMN reductase	WP_014471115.1
YieF_Ec	<i>Escherichia coli</i>	SsuE	FMN reductase	WP_001291268.1
YieF_Pa	<i>Pseudomonas aeruginosa</i>	SsuE	FMN reductase	WP_003082483.1
MdaB_Ec	<i>Escherichia coli</i>	MdaB	FMN reductase	WP_050864130.1
MdaB_Ps	<i>Pseudomonas syringae</i>	MdaB	FMN reductase	WP_011167644.1
WrbA_Ec	<i>Escherichia coli</i>	WrbA	FMN reductase	WP_001151437.1
WrbA_Ps	<i>Pseudomonas syringae</i>	WrbA	FMN reductase	WP_004657185.1
YdjA_Ec	<i>Escherichia coli</i>	Arsenite oxidase	Nitro FMN reductase	EGI21339.1
YcdI_Ec	<i>Escherichia coli</i>	Nitroreductase 5	Nitro FMN reductase	WP_001001176.1
YcdI_Kp	<i>Klebsiella pneumoniae</i>	Nitroreductase 5	Nitro FMN reductase	WP_004141442.1
YcaK_Ec	<i>Escherichia coli</i>	MdaB	FMN reductase	WP_001190363.1
YcaK_Pa	<i>Pseudomonas aeruginosa</i>	MdaB	FMN reductase	WP_003114232.1

Table 2.10: Oxidoreductase enzymes in the Ackerley group nitroreductase candidate library.

2.8 Characterisation of nitroreductases

2.8.1 Growth Inhibition Assays

A 96-well plate containing 200 μ L LB medium supplemented with 0.2% glucose (w/v) and Amp was inoculated with *E. coli* 7NT pUCX:nitroreductase strains and incubated at 30 °C, 200 rpm for 16 hours. A 15 μ L sample of the overnight culture was used to inoculate 200 μ L of induction media (LB supplemented with Amp, 0.2% (w/v) glucose and 50 μ M IPTG) and was incubated at 30 °C, 200 rpm for 2.5 hours. Cultures were split into 30 μ L aliquots and used to inoculate wells of a 384-well plate, two wells supplemented with 30 μ L of challenge media (LB supplemented with Amp, 0.2% (w/v) glucose and prodrug) and two

wells supplements with 30 μ L of control media (LB supplemented with Amp, 0.2% (w/v) glucose and an equivalent volume of DMSO). OD₆₀₀ readings were recorded using an EnSpire™ Multilabel Reader (*PerkinElmer; Waltham, MA, USA*). Cultures were incubated for a further 4 hours, after which the OD₆₀₀ was recorded once more. The percentage growth inhibition was determined by calculating the relative increase in OD₆₀₀ for challenged versus control wells.

2.8.2 IC₅₀ Assays

A 96-well plate containing 200 μ L LB media supplemented with 0.2% glucose (w/v) and Amp was inoculated with *E. coli* 7NT pUCX:nitroreductase strains and incubated at 30 °C, 200 rpm for 16 hours. A 100 μ L sample of the overnight culture was used to inoculate 2 mL of induction media (LB supplemented with Amp, 0.2% (w/v) glucose and 50 μ M IPTG) in a 15 mL tube and incubated at 30 °C, 200 rpm for 2.5 hours. A 30 μ L sample of each culture was added to wells plates of a 384-well plate containing 30 μ L of induction media with supplemented with two times the final prodrug concentration. Each culture was exposed in duplicate to 7-15 prodrug concentrations containing 1.5-fold increasing titrations of prodrug and one induction media only control. The induction media was supplemented with DMSO as appropriate, ensuring that the final concentration remained < 4%. OD₆₀₀ readings were recorded using an EnSpire™ Multilabel Reader (*PerkinElmer; Waltham, MA, USA*) at $t = 0$ (pre-challenge) and at $t = 4$ (post 4-hour challenge). The percentage growth inhibition was determined at each prodrug concentration by calculating the relative increase in OD₆₀₀ for challenged versus control wells. The IC₅₀ value (the concentration of prodrug at which 50% growth inhibition occurs) of technical replicates was calculated using a dose-response inhibition four-parameter variable slope equation in GraphPad Prism 8.0 (*GraphPad Software Inc; La Jolla, USA*). The IC₅₀ values of biological replicates were averaged to give the final IC₅₀ value.

2.8.3 Whole cell fluorescence assay

A 96-well plate containing 200 μ L LB supplemented with 0.2% (w/v) glucose and Amp was inoculated with *E. coli* 7NT pUCX:nitroreductase strains and incubated at 30 °C, 200 rpm overnight. A 15 μ L sample of the overnight culture was used to inoculate duplicate wells of a 96-well plate containing 200 μ L induction minimal media (minimal media supplemented with Amp, 0.2% (v/v) glucose and 50 μ M IPTG). Cells were incubated at 37 °C, 200 rpm for 2.5 hours. Each duplicate well was split into 2 $\times$ 100 μ L aliquots and

used inoculate induction minimal media containing 20 μ M of fluoroprobe and another well containing induction minimal media containing the equivalent concentration of DMSO (serving as a control for the detection of strain-dependent auto-fluorescence). Cultures were incubated for a further hour at 37 °C, 200 rpm after which the fluorescence and OD₆₀₀ was determined using an EnSpire™ Multilabel Reader (*PerkinElmer; Waltham, MA, USA*). Excitation and emission wavelengths of 350/455 nm, 355/460 nm and 405/585 nm were used for the detection of FSL41, FSL61 and FSL76 respectively. The fluorescence of each culture was normalised by cell turbidity (OD₆₀₀) and relative fluorescence units (RFU) were calculated from the fold difference between fluorescence values of nitroreductase cultures and the pUCX_empty control.

2.8.4 SDS-PAGE of intracellular over-expressed nitroreductases

E. coli 7NT pUCX:nitroreductase strains were used to inoculate 200 μ L LB media supplemented with 0.2% (w/v) glucose and Amp. Cultures were incubated overnight at 30 °C, 200 rpm. Following incubation 100 μ L of the overnight culture was used to inoculate 2 mL of LB induction medium (LB supplemented with 0.2% (w/v) glucose, Amp and 50 μ M IPTG). Day cultures were grown at 30 °C, 200 rpm for 6.5 hours, after which the cultures were pelleted by centrifugation at 2500 $\times$ g for 5 minutes. The supernatant was decanted and the cell pellets were resuspended in ~100 μ L of LB medium and the OD₆₀₀ of a 1 in 100 dilution was measured. Cell cultures were normalised by dilution with additional LB medium so that a 1 in 100 dilution would give an OD₆₀₀ reading of 0.1. A 20 μ L sample of each culture was mixed with SDS loading buffer (Table 2.11), heated at 95 °C for 5 minutes and 15 μ L was loaded onto a 4-12% ExpressPlus™ PAGE gel (*GenScript®; NJ, USA*) as per Section 2.8.4.1.

2.8.4.1 SDS-PAGE

For visualisation of proteins, 4-12% ExpressPlus™ PAGE gels (*GenScript, NJ, USA*) were used as per the manufacturer's instructions. Protein bands were stained by gentle shaking in Coomassie Blue stain solution for 30 minutes. Destaining was carried out by rinsing the gel of all Coomassie Blue stain in running tap water, following by gently shaking in destain solution for as long as necessary to achieve the desired contrast. Protein size was determined by comparison by to PageRuler™ Prestained Protein Ladder 10-180 kDa (*Thermo Fisher Scientific; Waltham, MA, USA*) of which 5 μ L was run alongside protein samples. Solution recipes are listed in Table 2.11.

Solution	Components
5 x SDS-PAGE loading buffer	5% β -mercaptoethanol 0.02% (w/v) bromophenol blue 30% (v/v) glycerol 10% (w/v) SDS 250 mM Tris.HCl pH 6.8
Coomassie stain solution	1% (w/v) Coomassie Brilliant Blue 40% ethanol 10% glacial acetic acid
Destain solution	40% ethanol 10% glacial acetic acid

Table 2.11: SDS-PAGE solution recipes.

2.8.5 Protein purification

2.8.5.1 Growth of protein over-expression cultures

E. coli BL21 (DE3) cells expressing a pET28a+:nitroreductase construct were used to inoculate a 3 mL culture of LB medium supplemented with kanamycin and grown overnight at 37 °C, 200 rpm. The overnight culture was used to inoculate a flask containing 50-100 mL of LB medium supplemented with kanamycin at a 1 in 100 dilution and incubated at 37 °C, 200 rpm until an OD₆₀₀ of 0.6-0.8 was reached. The culture was chilled on ice for 20 minutes after which IPTG was added to a final concentration of 0.5 mM to induce nitroreductase expression. The culture was incubated for a further 24 hours at 18 °C, 200 rpm. After incubation, the cells were collected by centrifugation at 2500 × g for 20 minutes at 4 °C. The cell pellet was processed immediately or stored at -20 °C.

2.8.5.2 Cell lysis

Cell pellets were resuspended in 2.5 mL of BugBuster® (*Merck Millipore; MA, USA*) and incubated at room temperature on a shaking platform for 30 minutes. If a protein was insoluble or gave a low yield using BugBuster® lysis, a French press was used to lyse cells. Cell pellets were resuspended in 20 mL of bind buffer (Table 2.12) passed through the French press chamber twice at 40,000 psi. Lysed cell cultures were centrifuged at 21,000 × g for 20 minutes to separate the insoluble and soluble fractions. Soluble fractions were decanted and stored on ice until purification.

2.8.5.3 Immobilised metal affinity chromatography purification

Purification of His₆-tagged proteins from the soluble fraction was carried out using Ni-NTA His.Bind® Resin (*Merck Millipore; MA, USA*) and a peristaltic pump. Recipes for the protein purification buffers used are listed in Table 2.12. A 2 mL sample of Ni-NTA

His.Bind® Resin was packed into a 10 mL Pierce disposable column (*Thermo Fisher Scientific; Waltham, MA, USA*) and connected to a peristaltic pump. The resin was washed with 4 mL of ddH₂O and charged by running through 8 mL of charge buffer followed by 4 mL of bind buffer. The soluble fraction of the lysed cells was loaded onto the column, mixed with the resin and left until the resin had settled. The soluble lysate was pumped through the column, followed by 10 mL of bind buffer and 8 mL of wash buffer. The pump tubing was drained, after which 4 mL of elution buffer was pumped through the column and 500 μ L fractions were collected. The three most yellow fractions (indicative of FMN-bound nitroreductase) were pooled. The resin was washed with 4 mL of strip buffer and stored in 20% ethanol at 4 °C. The buffer levels were maintained at a higher level than the resin level to ensure air was not introduced.

Buffer	Components
Charge buffer	50 mM NiSO ₄
Bind Buffer	500 mM NaCl 20 mM Tris.HCl pH 7.9 5 mM imidazole
Wash Buffer	500 mM NaCl 20 mM Tris.HCl pH 7.9 60 mM imidazole
Elution Buffer	500 mM NaCl 20 mM Tris.HCl pH 7.9 1 M imidazole
Strip Buffer	500 mM NaCl 20 mM Tris.HCl pH 7.9 100 mM EDTA

Table 2.12: Buffer recipes for His₆-tag purification using Ni-NTA His-Bind® Resin.

2.8.5.4 FMN supplementing and desalting

Pooled protein fractions were incubated with excess FMN cofactor (1 mM) on ice for 1 hour. Buffer exchange into 40 mM Tris.HCl pH 7.0 was carried out using a 5 mL HiTrap™ desalting column (*GE Healthcare; Chicago, IL, USA*) as per the manufacturer's instructions. Once the column was equilibrated and loaded with the protein fraction, the first 1.5 mL of flow-through was collected (flow-through beyond this was likely to contain contaminating free FMN). His₆-tags were not removed at any point during protein purification as the presence of this tag had not been previously seen to affect nitroreductase catalytic activity (Prosser *et al.*, 2013). Purified protein was mixed 1:1 with 80% (w/v) glycerol and stored in single use aliquots at -80 °C.

2.8.5.5 Protein quantification

The absorbance of protein samples was measured at 280 nm using a NanoPhotometer® spectrophotometer (*Implen; München, Germany*), according to the manufacturer's instructions. The Beer-Lambert equation ($A = \epsilon \cdot c \cdot l$) was used to determine the final protein concentration. The online tool ProtParam (*Gasteiger et al., 2005; <https://web.expasy.org/protparam/>*) was used to determine the theoretical molar extinction coefficients for proteins.

2.8.6 *In vitro* His₆-tagged protein kinetics

2.8.6.1 Determination of extinction coefficients of compounds

The molar extinction coefficient of metronidazole at 340 nm was determined by incubating 100 μ M of metronidazole with 300 μ M NADPH and excess purified NfsB_Vv for 30 minutes. NfsB_Vv was used as it possess low-level NADPH oxidase activity, such that it slowly oxidises all remaining NADPH in the cuvette following full reduction of metronidazole. The molar extinction coefficients of RB6145 at 340 nm (*Williams, 2013*) and CB1954 at 420 nm (*Race et al., 2007*) had been previously determined (Table 2.13). The reduction of nitrofurazone, 2,4-DNT, chloramphenicol and *p*-benzoquinone did not interfere with absorbance at 340 nm when reduced. As such, kinetics performed with nitrofurazone, 2,4-DNT and chloramphenicol utilised the extinction coefficient of $2 \times$ the molar extinction coefficient of NADPH ($2 \times \epsilon = 12,400 \text{ M}^{-1}\text{cm}^{-1}$), representing the assumed oxidation of two moles of NADPH per mole of substrate reduced to the hydroxylamine. Kinetics performed with *p*-benzoquinone utilised the extinction coefficient of $1 \times$ the molar extinction coefficient of NADPH ($6,220 \text{ M}^{-1}\text{cm}^{-1}$) representing the assumed oxidation of one mole of NADPH per mole of *p*-benzoquinone. A summary of the extinction coefficients used in this study are listed in Table 2.13.

Compound	Extinction Coefficient ($\text{M}^{-1}\text{cm}^{-1}$)
Metronidazole	1980
RB6145	19500
CB1954	1200
Nitrofurazone	12400
2,4-DNT	12400
Chloramphenicol	12400
<i>p</i> -benzoquinone	6220

Table 2.13: Extinction coefficients of compounds used in this study.

2.8.6.2 Standard purified protein kinetics

Apparent steady state enzyme kinetics for nitroreductases with nitroaromatic and quinone substrates (with the exception of metronidazole) were assessed by monitoring the decrease in absorbance at 340 nm (RB6145, nitrofurazone, 2,4-DNT, chloramphenicol and *p*-benzoquinone) or the increase in absorbance at 420 nm (CB1954) using the extinction coefficients outlined in Table 2.13. Reactions were prepared in a 60 μ L total volume in UVettes^{TM} (*Eppendorf; Hamburg, Germany*) with a 2 mm path length. Reactions contained 10 mM Tris.HCl pH 7.0, varying concentrations of substrate (made up at $25 \times$ final concentration in 100% DMSO), 250 μ M NADPH and an appropriate concentration of purified enzyme (ranging from 0.1-10 μ M). Reactions were initiated by the addition of NADPH and purified enzyme ($\sim$ 10 μ M of enzyme for slow reactions and $\sim$ 0.1-1 μ M enzyme for faster reactions). The linear decrease or increase in absorbance was measured following initiation and used to determine the initial rate of reaction. Absorbance at the desired wavelength was measured at room temperature using a SpectraSuite® USB4000 UV/Vis spectrophotometer (*Ocean Optics; Largo, FL, USA*). Technical replicates were plotted using GraphPad 8.0 software and non-linear regression analysis and Michaelis-Menten curve fitting was performed (*GraphPad Software Inc; La Jolla, CA, USA*).

2.8.6.3 Purified enzyme kinetics using NADPH regeneration system

Apparent steady state enzyme kinetics for NfsA and metronidazole were measured using a NADPH regeneration system so that the intrinsic NADPH oxidase activity of NfsA would not interfere with measurements. This meant that the reduction of metronidazole could be measured at 340 nm, independent of NADPH oxidation. Reactions were prepared in quartz cuvettes with a 1 mm path length. The NADPH regeneration system was modified from Robins *et al.*, (2013), and consisted of 100 μ L reactions containing 50 mM sodium phosphate pH 7.0, 5 mM glucose, 100 μ M NADPH, 0.55 μ M glucose dehydrogenase from *Bacillus subtilis*, 0-1500 μ M metronidazole and 10 μ M purified NfsA. Reactions were initiated by the addition of purified NfsA. The linear decrease in absorbance was measured and used to determine the initial rate of reaction. Absorbance at 340 nm was measured at room temperature using a SpectraSuite® USB4000 UV/Vis spectrophotometer (*Ocean Optics; Largo, FL, USA*). Technical replicates were plotted using GraphPad 8.0 software and non-linear regression analysis and Michaelis-Menten curve fitting was performed (*GraphPad Software Inc; La Jolla, CA, USA*).

2.9 Generation and maintenance of HEK-293 cell lines

2.9.1 Growth of HEK-293 cell lines

HEK-293 cell lines were maintained in RPMI Medium 1640, no glutamine, supplemented with glutaMAX^{TM} (*Thermo Fisher Scientific; Waltham, MA, USA*) and 10% fetal calf serum (FCS). Wild-type HEK-293 cells were grown in medium additionally supplemented with 1% Penicillin Streptomycin (Pen-Strep), while HEK-293 cell lines carrying a F279-V5 plasmid were maintained in medium supplemented with puromycin (1-3 μ M). All media were stored at 4 °C and warmed to 37 °C before use.

Cell lines were grown in a monolayer up to 80% confluency in a humidified incubator (37 °C, 5% CO₂) and sub-cultured at least every 4 days. Cells were harvested from growth flasks and microplates using a trypsin/EDTA protocol. Growth medium was aspirated, and cells were washed briefly with pre-warmed phosphate buffered saline (PBS). A volume of trypsin/EDTA sufficient to cover the surface of the cell monolayer was added into the flask and incubated for 3 minutes in a humidified incubator. The flask was agitated to detach adherent cells, and trypsin was inactivated by addition of three volumes of FCS-containing medium. Cells were removed from the flask or microplate, transferred to a 15 mL centrifuge tube and centrifuged at 300 $\times$ g for 5 minutes to remove trypsin-containing medium. The cell pellet was resuspended in RPMI medium and the desired dilution was seeded into a new flask. If exact numbers of cells were required for seeding, 20 μ L of the cell suspension were mixed with 20 μ L of trypan blue and counted using a haemocytometer (*Improved Neubauer; Hausser, Horsham, PA*) and seeded into fresh medium at the desired density.

2.9.2 Storage and recovery of frozen stocks

For long term storage, 1×10^6 cells resuspended in 500 μ L of RPMI medium were mixed with 500 μ L of freeze medium (RPMI medium, no glutamine supplemented with glutaMAX^{TM}, 40% (v/v) FCS and 20% (v/v) DMSO) in a cryotube. The cells were frozen at a controlled rate of -1 °C per minute using a Mr Frosty® Cryo 1 °C Freezing Container (*Thermo Fisher Scientific; Waltham, MA, USA*). When a temperature of -80 °C was reached, the cryotubes were placed in a liquid nitrogen dewar for long-term storage. Cell recovery was achieved by removing cell aliquots from the liquid nitrogen dewar and rapidly thawing in a 37 °C bead bath before diluting the entire aliquot in 9 mL of pre-warmed

RPMI medium. The diluted cell suspension was centrifuged at $300 \times g$ for 5 minutes to remove the DMSO-containing medium, after which the cell pellet was resuspended in the desired medium and the entire volume seeded into a T25 flask.

2.9.3 Generation of F279-V5:NTR constructs using Gateway® technology

2.9.3.1 Generation of Gateway® compatible PCR products

Nitroreductase genes were amplified in Q5® PCR reactions (Section 2.5.2) using Gateway compatible primers (Table 2.1) that introduced a mammalian Kozak sequence, a Shine-Dalgarno prokaryote consensus sequence, Invitrogen Gateway® BP recombination sites, an extra GGC codon immediately following the ATG start codon and a TAG stop codon. PCR products were purified as per Section 2.5.4.1 and used in a BP Clonase recombination reaction (Section 2.9.3.2).

2.9.3.2 BP Clonase recombination

pDONR221:nitroreductase gateway entry clone constructs were created through a BP recombination step using BP Clonase™ II enzyme mix (*Thermo Fisher Scientific; Waltham, MA, USA*). pDONR221 vector (150 ng) was mixed with 50 fmol of PCR product, 0.5 µL BP Clonase enzyme mix and TE buffer (10 mM Tris.HCl, 1 mM EDTA, pH 8.0) to a final volume of 8 µL. The reaction was incubated at room temperature overnight and stopped by addition of 1 µL of 2 µg.µL⁻¹ Proteinase K solution, incubated at 37 °C for 10 minutes. The reaction was stored at -20 °C or used immediately whereby 1 µL of the reaction was used to transform electrocompetent *E. coli* EC100 cells. Recovered cells were plated on agar plates supplemented with kanamycin. Colonies were analysed for the presence of the nitroreductase gene by conducting colony PCR reactions using both a nitroreductase-specific reverse primer and a pDONR221-specific forward primer and sequencing of the plasmid. Colonies which contained the correct plasmid were used to inoculate 3 mL overnight cultures of LB supplemented with kanamycin, from which the plasmid was isolated the following day as per Section 2.5.1.

2.9.3.3 LR clonase recombination

Recombination of the nitroreductase gene from pDONR221 into the Gateway destination vector F279-V5 was accomplished using LR Clonase™ II enzyme mix (*Thermo Fisher Scientific; Waltham, MA, USA*). pDONR221 vector (150 ng) was mixed with 150 ng of F279-V5 vector, 1 µL LR Clonase enzyme mix and TE buffer pH 8.0 to a final volume of

8 μL . The reaction was incubated at room temperature overnight and stopped by addition of 1 μL of 2 $\mu\text{g}\cdot\mu\text{L}^{-1}$ Proteinase K solution and incubated at 37 °C for 10 minutes. The reaction was stored at -20 °C or used immediately whereby 1 μL of the reaction was used to transform electrocompetent *E. coli* EC100 cells. Recovered cells were plated on agar plates supplemented with ampicillin. Colonies were analysed for the presence of the nitroreductase gene by conducting colony PCR reactions using both a nitroreductase-specific reverse primer and a F279-V5-specific forward primer and sequencing of the plasmid. Colonies containing the correct plasmid were used to inoculate 3 mL overnight culture of LB supplemented with ampicillin, from which plasmid was isolated the following day as per Section 2.5.1, eluting in sterile ddH₂O.

2.9.4 Transfection of HEK-293 cells

Wild-type HEK-293 cells were seeded into a 6-well plate and grown until 80-90% confluency. Four wells were transfected with each F279-V5:nitroreductase construct. Prior to transfection, cell growth medium was replaced with RPMI medium lacking antibiotic. Transfections were carried out using Lipofectamine 3000 (*Thermo Fisher Scientific; Waltham, MA, USA*) as per manufacturer's instructions with 2 μg of F279-V5 construct per well. Twenty-four hours post transfection, one well transfected with each construct was tested for transient nitroreductase expression as per Section 2.9.5.1. Forty-eight hours post transfection, medium in the remaining wells was replaced with medium containing 1 μM puromycin. Cells which had stably integrated the construct were selected for by exposure to increasing concentrations of puromycin (1-3 μM). Once cell blooms of resistant cells had appeared in the 6-well plates (around 2-3 weeks post transfection), cells were passaged into T25 flasks and then characterised or frozen down for long-term storage.

2.9.5 Characterisation of HEK-293 cells expressing nitroreductases

2.9.5.1 Fluorescence microscopy with fluoroprobes

Activation of the fluoroprobe FSL61 was used as an indicator of nitroreductase expression. Cells from each HEK-293 line to be assessed were seeded into wells of a 6-well plate and grown until ~80% confluency. Cell medium was replaced with RPMI phenol free medium supplemented with 50 μM FSL61. HEK-293 cells were incubated for 1 hour (37 °C, 5% CO₂). Following the incubation, the fluoroprobe medium was removed and the cells were washed with PBS and replaced with 1 mL phenol free RPMI medium to prevent auto-fluorescence. Cells were visualised using an Olympus IX51 inverted microscope (*Olympus;*

Hamburg, Germany) using the DAPI filter. Cells were analysed using CellSens software (Olympus; Hamburg, Germany) using a constant exposure setting between samples.

2.9.5.2 Flow cytometry

Cells from each line to be assessed were harvested, resuspended in RPMI phenol-free medium and counted. A total of 1×10^6 cells from each cell line were seeded into duplicate wells in a non-tissue culture treated 24-well plate in a 500 μ L volume. One mL of medium containing 10 μ M FSL61 or the equivalent volume of DMSO was added to each duplicate well. Cells were protected from light and incubated for 1 hour (37 °C, 5% CO₂) and transferred to flow cytometry tubes prior to analysis. Fluorescence of single cell events was monitored using a FACS Canto II flow cytometer (BD Biosciences; Franklin Lakes, NJ, USA). Single cells were gated using forward- and side-scatter set to 170 V and 420 V respectively. The V450 channel was used to detect FSL61 fluorescence set to 265 V. Data was analysed using FLOWJO software (Tree Star, Ashland, OR, USA). Where possible at least 20,000 events were collected per sample.

2.9.5.3 Western Blots

2.9.5.3.1 Sample preparation and quantification

A T25 flask containing the desired cell line at 70-80% confluency was passaged and transferred to a 15 mL centrifuge tube. Cells were centrifuged at $300 \times g$ for 5 minutes. The supernatant was discarded and cells were gently washed with 3 mL ice-cold PBS. Cells were centrifuged at $300 \times g$ for 5 minutes, the supernatant was removed and the cell pellet placed on ice. The cell pellets were resuspended in 350 μ L of radioimmunoprecipitation assay (RIPA) buffer (Table 2.14) supplemented with 10 μ L.mL⁻¹ P8340 protease inhibitor cocktail (Sigma-Aldrich; St Louis, MO, USA) and incubated on ice for 30 minutes. Samples were transferred to pre-chilled microcentrifuge tubes and centrifuged at $21,000 \times g$ for 20 minutes at 4 °C to collect cell debris. The clear lysate was removed to a new tube and stored at -80 °C until required. The protein concentration of each sample was determined using a Pierce^{TM} Bicinchoninic Acid (BCA) Protein Assay Kit (Thermo Fisher Scientific; Waltham, MA, USA) according to manufacturer's instructions.

2.9.5.3.2 SDS-PAGE separation and membrane transfer

A total of 15 μ g of each western blot sample was mixed with SDS loading buffer and incubated at 95 °C for 5 minutes. Samples were run on a 4-12% ExpressPlus^{TM} PAGE gel

(*GenScript®; Nanjing, China*) as per the manufacturer's instructions. Following electrophoresis, the gel cast was disassembled and proteins were transferred onto a polyvinylidene difluoride 0.45 μ M pore size membrane (*Millipore Sigma; MA, USA*) using a Bio-Rad Mini Trans-Blot® Electrophoretic Transfer Cell System filled with ice-cold transfer buffer (Table 2.14) and run at 25 V for 18 hours at 4 °C.

Buffer	Components
RIPA Buffer	50 mM Tris.HCl pH 8.0 150 mM NaCl 1 mM EDTA 0.5% (v/v) sodium deoxycholate 0.1% (w/v) SDS 1% (v/v) Triton X-100
Transfer Buffer	25 mM Tris base 192 mM glycine 20% (v/v) methanol
TBS	20 mM Tris base 150 mM NaCl pH to 7.6

Table 2.14: Western blot buffers.

2.9.5.3.3 Antibody detection of proteins

The membrane was removed from the transfer apparatus and washed with TBS-T (TBS supplemented with 0.1% (v/v) Tween-20) and blocked with blocking buffer (TBS-T supplemented with 5% (w/v) bovine serum albumin (BSA)) for 1 hour on an orbital shaker at room temperature. The blocked membrane was gently washed in TBS-T. The primary antibody was diluted in TBS-T supplemented with 1% (w/v) BSA, added to the membrane and incubated for 4 hours on an orbital shaker at 4 °C. Following incubation, the buffer was decanted, washed three times for 5 min in TBS-T and the appropriate secondary antibody was added diluted in TBS-T supplemented with 1% (w/v) BSA. Secondary antibody incubation was carried out at room temperature for 2 hours in a lightproof bag. Following incubation, the buffer was decanted and washed three times for 5 minutes in TBS-T. For the detection of chemiluminescent secondary antibodies (Donkey anti-sheep IgG-HRP), the membrane was incubated with SuperSignal® West Pico Chemiluminescent Substrate (*Thermo Fisher Scientific; Waltham, MA, USA*) for 3 minutes. Image acquisition was accomplished using a Fujifilm LAS-4000 (*Fujifilm; Tokyo, Japan*). For the detection of fluorescent secondary antibodies (Goat anti-rabbit IgG-Alexa488), the membrane was visualised on a Fujifilm FLA 5100 fluorescence scanner (*Fujifilm; Tokyo, Japan*). Densitometry analysis was performed using ImageJ (version 1.8.0 of the public domain Java image processing program) and nitroreductase expression levels were normalised

against α -tubulin expression levels. All antibodies used for western blots are described in Table 2.15.

Antibody	Dilution	Source
Sheep anti-NfsA_Ec	1:5000	Dr Peter Searle (University of Birmingham)
Rabbit anti- α -tubulin	1:5000	Abcam (ab18251)
Goat anti-rabbit IgG-Alexa488	1:5000	Thermo Fisher Scientific (A-11008)
Donkey anti-sheep IgG-HRP	1:5000	Abcam (ab97125)

Table 2.15: Antibodies and dilutions used in western blots.

2.9.5.4 Cell viability assays

HEK-293 cell lines were seeded at a density of 13,000 cells per well in a 100 μ L volume (RPMI medium, no glutamine, supplemented glutaMAX^{TM}, 10% FCS and 2 μ M puromycin) into 96-well plates (6.25 mm diameter with cell adhesion) and left to adhere in a humidified incubator at 37 °C, 5% CO₂ for 24 hours. Following incubations, cells were challenged with 100 μ L of RPMI medium, no glutamine, supplemented with glutaMAX^{TM} and two times the final prodrug concentration. Cells were incubated at 37 °C, 5% CO₂ for 48 hours. Following prodrug challenge, 15 μ L of CellTitre 96 Aqueous One Solution Cell Proliferation Assay reagent (*Promega; Madison, WI, USA*) were added to each well (as well as to a medium-only containing well to allow for baseline medium absorbance subtraction). Cells were incubated for a further 2.5 hours at 37 °C, 5% CO₂. The absorbance of wells was measured at 490 nm. The baseline-subtracted absorbance of challenged wells was compared to that of an unchallenged control well to determine percentage cell viability for each prodrug concentration. The IC₅₀ value (the concentration of prodrug at which 50% growth inhibition occurs) of technical replicates was calculated using a dose-response inhibition four-parameter variable slope equation in GraphPad Prism 8.0 (*GraphPad Software Inc; La Jolla, USA*). The IC₅₀ values of biological replicates were averaged to give the final IC₅₀ value.

2.10 Expression of nitroreductases in zebrafish

2.10.1 Ethics statement

The zebrafish work completed in this study was carried out in accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care and use Committee of Johns Hopkins School of Medicine (Approval Identification Number: FI17M19). This institution has an Animal Welfare Assurance on file in the Office of

Laboratory Animal Welfare (Assurance Number: A3272-01). All efforts were made to minimise discomfort and suffering during all experimental procedures.

2.10.2 Zebrafish maintenance

Zebrafish larvae were reared in E3 medium (Table 2.16) in petri dishes in an incubator at 28.5 °C until they hatched around 3-5 days post fertilisation (dpf). Zebrafish larvae to be raised were then transferred into larval tanks containing E3 medium until 14 dpf. Following this, larvae were transferred into larger fish tanks and maintained at 28.5 °C in fish water (Table 2.16) on a 14 hour/10 hour light/dark schedule in a recirculating water system (*Aquatic Systems; USA*) at John Hopkins School of Medicine. From 5 dpf, zebrafish were fed a combination of dried fish food and brine shrimp. When using zebrafish larvae in cell ablation assays or for confocal imaging, E3 medium was supplemented with 0.2 mM of 1-phenyl-2-thiourea (PTU) to inhibit melanisation. When larvae anaesthetisation was required, E3 medium was supplemented with 0.2 mg.mL⁻¹ tricaine methansulfonate (*Argent; Redmond, WA, USA*).

Media	Component
E3 Media	5 mM NaCl 0.15 mM KCl 0.33 mM CaCl ₂ 0.33 mM MgSO ₄
Fish water	250 mg.mL ⁻¹ sodium bicarbonate 100 mg.mL ⁻¹ instant ocean salts

Table 2.16: Zebrafish maintenance media recipes.

2.10.3 Dechoriation

Zebrafish larvae which had not hatched naturally by 4 dpf were manually dechorionated under a dissecting microscope using forceps to gently tear open the chorion releasing the zebrafish larva.

2.10.4 Generation of transgenic zebrafish lines

2.10.4.1 Transgene injection into embryos

Zebrafish embryos at the single cell stage were injected using a 1.0-1.2 mm diameter calibrated microinjection capillary. Injections consisted of a 0.38-0.90 µL volume containing 50 ng.µL⁻¹ pTOL2 plasmid, 5% (v/v) phenol red dye, Danieau's buffer (17.4 mM NaCl, 0.21 mM KCl, 0.12 mM MgSO₄, 0.18 mM Ca(NO₃)₂, 1.5 mM HEPES pH 7.2) made up to volume with ddH₂O. Tol2 transposase mRNA was included when generating stable cell lines to assist in integration of the transgene into the genomic DNA

of the zebrafish. Phenol red dye was used to ensure injections were into the single cell of the embryo and not the yolk.

2.10.4.2 Generation of stable transgenic zebrafish lines

Embryos were injected as per Section 2.10.4.1 and kept in E3 medium at 28.5 °C. Injected embryos were monitored under a fluorescent microscope for evidence of mosaic TagYFP fluorescence (indicative of nitroreductase expression). After 5 dpf, all larvae exhibiting TagYFP fluorescence, were raised as putative founders (P_0) until they reached sexual maturity (around 3 months). The putative founders were mated, and their offspring were checked for broad TagYFP fluorescence, indicative of germline transgenesis. Any offspring exhibiting broad TagYFP fluorescence were raised as the F_1 generation, indicating a stable transgenic line had been established. Embryos generated from matings of this (F_1) and subsequent generations (F_2 , F_3 etc.) were used in cell ablation assays (Section 2.10.6). A schematic of this process is outlined in Figure 2.1.

Figure 2.1: Schematic showing the generation of transgenic zebrafish lines stably expressing a nitroreductase.

2.10.5 Assessment of prodrug-induced lethality in non-transgenic zebrafish

Non-transgenic zebrafish larvae were transferred to 60 mm petri dishes containing E3 medium at 4 dpf and treated with the prodrug of interest or a DMSO control. Viability was assessed and quantified by examination through a dissecting microscope. Viability was defined by a beating heart and response to stimulus such as gentle plate shaking.

2.10.6 Cell ablation assays

Transgenic zebrafish larvae were mated, and larvae were collected the following morning and transferred to E3 medium. Larvae were checked for TagYFP fluorescence under the fluorescence microscope at 4 dpf, indicative of nitroreductase expression. Larvae containing comparable TagYFP fluorescence were sorted and used in cell ablation assays. Non-transgenic larvae, also examined at 4 dpf, were used as a control to determine the background larval auto-fluorescence (greater than the number of treatment larvae). Fluorescent larvae were split into two equal groups (a control and treatment group). Larvae were incubated in 60 mm petri dishes in E3 medium supplemented with PTU and diluted prodrug (or an equivalent volume of DMSO) for the prescribed challenge time after which larvae were washed in E3 medium supplemented with PTU. Larva fluorescence was examined 48 hours following commencement of challenge. Where the prodrug challenge time was less than 48 hours, larvae were transferred into E3 medium until 48-hours post challenge commencement. This ensured that the immune system of the zebrafish had time to clear any fluorescent debris following cell ablation (*personal communication Assoc. Prof. Jeff Mumm*). To measure fluorescence, larvae were anaesthetised and transferred into single wells of a black polypropylene 96 well microplate with U-shaped wells (*Greiner-Bio One; Kremsmünster, Austria*) in 350 μ L of E3 medium. Fluorescence was read on a TECAN Infinite M1000 plate reader (*Tecan; Männedorf, Switzerland*). TagYFP fluorescence was measured using an excitation wavelength of 568 nm and an emission wavelength of 610 nm. The fluorescence of non-transgenic fish was used to establish background auto-fluorescence, and a 'signal threshold' above which signal could be attributed to fluorescent transgene expression. The signal threshold was calculated as the average of all non-transgenic fish included in the experiment plus three standard deviations of the maximal regional fluorescence values from non-transgenic control larvae. Signal in transgenic fish was defined as any single regional fluorescence value above the signal threshold. The calculated fluorescence of each larva from each prodrug condition was averaged and compared to the fluorescence of the DMSO control larvae to determine the

degree of prodrug-induced cell ablation that had occurred. When the fluorescence signal dropped below detection limits due to robust ablation, the signal threshold value was used.

2.10.7 Confocal Microscopy

Larvae were anaesthetised and mounted in a petri dish within 30 μL of low melting point agarose supplemented with 0.2 mM PTU and 0.2 mg.mL^{-1} tricaine methansulfonate. Once mounted, the petri dish was flooded with E3 medium containing 0.2 mM PTU and 0.2 mg.mL^{-1} tricaine methansulfonate. Confocal imaging was performed on an Olympus FV1000 upright confocal microscope equipped with 405, 440, 488, 515, 559 and 625 nm lasers using Olympus Fluoview FV10-ASW imaging software (*Olympus; Hamburg, Germany*). For each fish z-stack images were acquired. A non-transgenic larva was used to define background fluorescence values.

2.11 Biological and technical replicates

For each experiment, multiple biological and technical replicates were carried out to evaluate the variation inherent within the system under investigation, and the variation due to measurement techniques, respectively. For the bacterial experiments described in this thesis, biological replicates refer to independent overnight and day cultures of a particular strain. For mammalian cell line experiments described in this thesis, biological replicates refer to independently defrosted/seeded cell lines at similar passage number. In *E. coli* and HEK-293 cells, technical replicates refer to assays performed on the same day using the same overnight and day cultures/cell line but in different experimental wells. In zebrafish experiments sample sizes refer to individual zebrafish embryos.

Chapter 3 Simultaneous site-directed mutagenesis of *nfsA*

3.1 Introduction

Directed evolution experiments that seek to optimise activities with non-natural substrates typically fail to achieve the level of efficiency observed with the natural substrate (Bornscheuer *et al.*, 2012). In contrast, nature can find ways to rapidly evolve enzyme activity towards a non-natural substrate. One study from 2011 showed *E. coli* could develop resistance to the antibiotic ciprofloxacin in only ten hours (Zhang *et al.*, 2011). This rapid response of nature to xenobiotic compounds suggests the goal of achieving native levels of activity with a non-natural substrate is attainable, but how can we harness this when evolving a single enzyme in the laboratory? Most conventional directed evolution techniques involve the successive accumulation of single amino acid changes that provide incremental improvements in enzyme fitness (Poelwijk *et al.*, 2007). While these approaches have shown some success in improving catalytic efficiency (Chen and Arnold, 1993; Gulick and Fahl, 1995; Kim and Lei, 2008; Li *et al.*, 2017), in the majority of cases, modifying substrate specificity or evolving entirely new enzyme functions has been rarely demonstrated (Georgiou and DeWitt, 1999; Chen, 2001; Dalby, 2011). Even multiple rounds of directed evolution rarely yields enzymes that are as active with the alternative substrate(s) as they originally were with the natural substrate(s) (Bornscheuer *et al.*, 2012). Conventional stepwise approaches also fail to address the importance that individually neutral amino acid substitutions can have on the overall fitness of an enzyme. These substitutions may not appear advantageous, however can enhance the stability of an enzyme and increase its tolerance for acquiring additional beneficial-but-destabilising mutations (Bloom and Arnold, 2009). These neutral substitutions can also become the starting point for the evolution of new functions. However, such functionally neutral substitutions are unlikely to be selected for in a conventional stepwise approach due to their lack of immediate contribution to fitness. To overcome these limitations, a different approach could be taken such as simultaneous site-directed mutagenesis. Simultaneous site-directed mutagenesis has the potential to quickly evolve an enzyme that exhibits “peak” activity, bypassing any intermediate variants.

3.1.1 Simultaneous site-directed mutagenesis

Conventional directed evolution approaches often subject a gene to iterative rounds of mutagenesis, followed by a selection or screen to identify enzyme variants with a desired function. Mutations are often introduced in an untargeted fashion, e.g., using error-prone PCR. Error-prone PCR can explore large areas of sequence space, however the untargeted nature of these mutations requires a large library to overcome the inevitable number of neutral or deleterious mutations (Copp *et al.*, 2014a). In contrast a targeted approach such as iterative site-saturation mutagenesis can be used to generate “smarter” libraries and improve the likelihood of identifying improved variants (Reetz and Carballera, 2007). For targeted approaches, structural, functional and sequence information of an enzyme can be used to pre-select residues for mutagenesis, creating smaller enzyme libraries (Lutz and Patrick, 2004; Chica *et al.*, 2005). An example of this approach was reported by Sandstrom *et al.*, (2012), in which nine codons in the active site of a lipase were simultaneously modified. Despite creating a small library of only 1024 variants, enzyme variants were identified with substantial changes in activity and enantioselectivity, demonstrating the strength of a simultaneous site-directed mutagenesis approach. We considered that simultaneously targeting multiple important residues in combination with the creation of a large enzyme library would vastly increase the number of active variants in a library, allowing superior enzyme variants to be rapidly identified. In contrast to iterative mutagenesis, a simultaneous approach would also allow individually neutral mutations to be retained, which may provide a fitness benefit in combination with other mutations.

When examining a large enzyme library it is critical to have an efficient method of isolating variants with improved activities. An effective selection provides the most efficient way of interrogating a large enzyme library. Previous work in the Ackerley group has shown that both niclosamide and chloramphenicol can be used as positive selection tools to identify active nitroreductase variants from a large enzyme library (Section 1.4.3.2).

3.1.2 The 7SM *nfsA* Library

The 7SM library is a *nfsA* simultaneous site-directed mutagenesis library previously created in the Ackerley lab. This library has been used as a resource for identifying improved NfsA variants for a range of applications. This library was designed by Drs Janine Copp and Elsie Williams who used both empirical and computational data to choose the seven codon positions for randomisation. Five of these sites (encoding residues S41, F42, F83, K222

and R225) were selected based on previous directed evolution studies of NfsA, where amino acid substitutions at these positions had improved activity with the prodrug PR-104A (Copp *et al.*, 2017) and the positron-emission tomography (PET) imaging probe EF5 (Williams, 2013). The remaining two sites (S224 and F227) were chosen using the online software HotSpot wizard 1.7 (<http://loschmidt.chemi.muni.cz/hotspotwizard/>) used sequence, structural and evolutionary information to computationally determine residues which may be of catalytic importance (Pavelka *et al.*, 2009). The degenerate codon NDT (N = adenine, cytosine, guanine, thymine; D = adenine, guanine, thymine; T = thymine) was utilised at six of the seven positions (S41, F42, F83, S224, R225 and F227). This codon allows for more efficient directed evolution as it encodes a representative sample of 12 different amino acids, including the wild-type amino acid without having to use the conventional NNK codon, which uses 32 codon possibilities to encode 20 amino acids (Reetz *et al.*, 2008). As the NDT codon does not encode lysine, the NNK codon (K = guanine, thymine) was used at position K222 to ensure the wild-type amino acid, lysine, could be incorporated. The 7SM library had a total of approximately 60 million unique NfsA variants of which a library of 10 million variants was created. Data obtained from screening this library played a pivotal role in finalising the choice of target residues for the simultaneous site-directed mutagenesis library created in this study.

3.1.2.1 *E. coli* 7NT screening strain and pUCX expression vector

The 7SM library and other *nfsA* libraries described in this chapter were cloned into the *E. coli* screening strain 7NT. This *E. coli* strain contains knockouts of seven endogenous oxidoreductases, predicted to be nitroreductase candidates at the time of strain creation (*nfsA*, *nfsB*, *yieF*, *ycaK*, *mdaB*, *azoR*, *nema*) and an efflux pump (*tolC*) (Copp *et al.*, 2014b). The plasmid pUCX was used for nitroreductase expression in the 7SM library and all other *nfsA* libraries described in this thesis. This plasmid was derived from pUC19 by the addition of the *lacI* gene, *tac* promoter, *lac* operator, RBS region and *rrnB* terminator sequence from pMMB67EH (Gonzalez *et al.*, 2005; Prosser *et al.*, 2010). Two variations of this plasmid exist in the Ackerley lab, pUCX_low and pUCX_high. The 7SM library was created using pUCX_low, which unknown at the time, contained mutations that decreased *nitroreductase* expression (Rich, 2017). The other libraries described in this thesis were created using pUCX_high to allow for higher nitroreductase expression. *E. coli* 7NT cells expressing pUCX without a nitroreductase (pUCX_empty) were used as a control throughout this study.

3.1.3 Using NfsA as a model to understand enzyme evolution

The goal of this study was to identify NfsA variants exhibiting superior detoxification of niclosamide or chloramphenicol. We believed using simultaneous site-directed mutagenesis would provide us with the most efficient method of identifying improved NfsA variants. Due to the direct toxic effects niclosamide and chloramphenicol impose on *E. coli* 7NT, we believed focussing on one of these compounds would enable us to rapidly identify NfsA variants exhibiting substantial detoxification of these compounds over the parental form, wild-type NfsA. The top NfsA variants would then be used in future experiments as a model to improve our understanding of how new enzyme functions evolve *in vivo*.

3.2 Aims

- Use simultaneous site-directed mutagenesis of *nfsA* to create an enzyme variant library containing $>10^8$ enzyme variants.
- Use chloramphenicol and niclosamide as positive selection compounds to identify the most active NfsA variants within the library.
- Isolate and characterise lead NfsA variants exhibiting superior detoxification of chloramphenicol or niclosamide.

3.3 Results

3.3.1 The 8SM *nfsA* Library

To extend from the previous 7SM library, an 8SM library was designed by Drs David Ackerley, Michelle Rich and Katherine Robins and was subsequently created as a part of this project together with Dr Robins. This library was created with the purpose of investigating an additional eight target positions in NfsA (non-overlapping with the seven residues in the 7SM library) to finalise an optimal choice of target residues. Empirical data was used to choose six sites (L43, C45, E99, L103, N134 and H215) based on amino acid substitutions at these sites having previously improved activity with the prodrug PR-104A (Copp *et al.*, 2017) or the PET imaging probe EF5 (Williams, 2013). In addition to these, positions R133 and T219 were selected due to their close proximity to the active site (Figure 3.1). The degenerate codon NDT was used at six of the eight positions (L43, C45, L103, R133, N134 and H215). The degenerate codons, SNG (S = cytosine, guanine) and NHT

(H = adenine, cytosine, thymine) were used at positions E99 and T219 respectively, to ensure that the wild-type amino acid could be encoded at each position, while still retaining a balanced diversity of alternative amino acid possibilities. In total, the 8SM library encoded a total of approximately 287 million unique NfsA variants. A summary of the 15-targeted residues in the 7SM and 8SM libraries is shown below (Figure 3.1). To generate preliminary data to inform the final residue choices for a simultaneous site-directed mutagenesis library, a sub-library of 2 million transformants was generated in the *E. coli* strain 7NT. Transformants were subjected to niclosamide or chloramphenicol selection and were counter-screened in growth inhibition assays to assess activity.

Figure 3.1: Targeted positions in the 7SM and 8SM *nfsA* libraries. (A) Structure of NfsA with the 15 residues targeted in the 7SM and 8SM libraries shown as sticks. Orange residues (●) were targeted in the 7SM library and blue residues (●) were targeted in the 8SM library. Only one of the active sites present in the NfsA homodimer is shown, with the FMN molecule shown in yellow. Each monomer is shaded either blue or green. Protein Data Bank (PDB) accession number 1f5v. (B) Summary of the 7SM and 8SM libraries.

3.3.2 Final selection of eight-targeted residues for a simultaneous site-directed mutagenesis library

Both libraries were plated on LB agar containing either the positive selection compound niclosamide (5 μ M) or chloramphenicol (3 μ M) to recover a smaller population of NfsA variants, removing the majority of inactive variants. Fifty-seven colonies which grew under these selection conditions were picked into the inner 60 wells of 96-well plates alongside one pUCX_empty and two wild-type NfsA controls. Strains overexpressing *nfsA* variants were challenged with a panel of prodrugs (metronidazole, RB6145 or azomycin, CB1954,

nitrofurazone and 2,4-DNT). Prodrug concentrations were chosen which resulted in ~50% growth inhibition for *E. coli* 7NT cells expressing wild-type *nfsA*. This meant growth inhibition greater than 50% would indicate a more active NfsA variant and growth inhibition less than 50% a less active variant. Different concentrations were used between the 7SM and 8SM *nfsA* libraries due to differences in the expression vector (pUCX_low versus pUCX_high respectively). The 57 *nfsA* variants were also sequenced to determine the mutations introduced and to identify any duplicates. All experiments using the 8SM library were completed together with Dr Robins. All experiments using the 7SM library were previously completed by Dr Rich. Data from the 7SM library has been replicated below with permission from Dr Rich as it provided essential information alongside the 8SM library data to inform the final eight positions to target in a simultaneous site-directed mutagenesis library. In the 7SM selections, of the 114 variants analysed, only one double transformant was identified (7SM_N_39; Table 3.1). NfsA variants from other libraries were found to be contaminating the 8SM library, evident by the identification of amino acid substitutions occurring outside of the eight positions targeted in the 8SM library (Table 3.2). This contamination most likely occurred via aerosols, due to the widespread use of other *nfsA* libraries in the laboratory at the time of creation of the 8SM library. While this significantly decreased the number of 8SM derived NfsA variants analysed, we continued with the amount of data obtained, which we deemed sufficient to provide meaningful pilot data. The creation of the 8SM library also emphasised precautions which must be taken during library generation. Consequently, for the handling of future libraries all work was conducted in a lamina flow hood, using filter pipette tips and UV irradiation of consumables and equipment to prevent DNA contamination.

A	Growth inhibition (%)	600 uM MTZ	25 uM azomycin	500 uM CBI954	37.5 uM NFZ	600 uM 2,4-DNT	S41	F42	F83	K222	S224	R225	F227
7SM_N_2	65.9	74.8	29.8	43.5	39.8		Y	N	L	E	F	N	D
7SM_N_3	89.2	79.2	38.8	41.2	39.3		H	N	L	Y	L	N	S
7SM_N_4	100.0	63.5	56.1	45.4	46.9		R	Y	F	L	D	S	F
7SM_N_5	75.3	93.9	35.5	39.1	37.5		H	N	L	K	G	D	R
7SM_N_6	100.0	42.4	42.4	33.3	32.8		I	F	L	V	S	V	H
7SM_N_7	69.2	66.1	26.5	48.5	40.5		Y	H	F	S	S	S	V
7SM_N_8	96.9	71.7	37.5	26.9	24.2		H	Y	F	R	N	H	V
7SM_N_9	90.7	83.1	31.7	50.0	36.1		Y	F	L	I	R	H	N
7SM_N_10	87.9	59.6	49.5	54.8	52.4		N	N	F	K	F	H	H
7SM_N_11	56.4	52.8	20.0	40.7	34.4		Y	N	F	P	R	S	S
7SM_N_12	93.9	71.3	41.1	38.7	41.3		Y	F	R	E	Y	N	N
7SM_N_13	96.9	93.2	37.0	61.4	40.4		H	F	Y	S	S	L	S
7SM_N_14	100.0	67.2	72.4	42.9	71.5		N	F	V	T	R	N	S
7SM_N_15	96.8	96.8	34.0	65.9	37.6		H	L	F	R	L	N	R
7SM_N_16	84.5	81.2	42.7	50.4	39.3		H	Y	F	R	N	H	V
7SM_N_17	94.6	78.2	49.8	63.7	38.9		C	H	F	R	H	C	H
7SM_N_18	97.0	82.6	46.8	46.5	39.0		H	Y	L	R	G	N	L
7SM_N_19	85.9	78.6	17.5	50.6	30.6		H	C	F	Q	N	F	D
7SM_N_20	91.1	82.9	46.0	56.9	43.8		R	L	F	F	L	S	H
7SM_N_21	79.7	90.2	30.8	61.9	39.9		F	F	F	R	D	R	F
7SM_N_22	86.8	89.7	41.1	68.1	43.2		F	L	H	Q	I	N	H
7SM_N_23	84.1	75.9	29.1	60.1	40.3		Y	S	F	L	N	N	S
7SM_N_24	98.2	53.0	55.3	28.0	59.1		N	H	F	C	C	H	N
7SM_N_25	62.7	59.5	23.4	65.7	36.4		Y	H	F	N	S	F	N
7SM_N_26	74.6	88.3	20.0	55.4	35.2		Y	H	L	S	F	N	H
7SM_N_27	80.8	68.1	27.8	54.6	34.5		Y	Y	L	R	S	F	R
7SM_N_28	92.8	83.8	31.1	76.8	44.9		H	S	F	L	S	C	H
7SM_N_29	68.6	59.8	14.2	57.2	33.6		Y	N	F	G	D	L	L
7SM_N_30	72.5	74.6	16.1	34.1	28.6		R	N	Y	L	H	N	H
7SM_N_31	92.9	82.0	39.2	70.5	46.1		H	L	L	D	Y	F	D
7SM_N_32	100.0	83.6	34.6	70.9	38.5		H	F	I	Q	G	N	R
7SM_N_33	73.6	85.4	25.3	51.4	34.3		H	N	L	Q	H	I	G
7SM_N_34	69.5	77.4	18.6	53.5	34.2		Y	H	F	G	D	S	G
7SM_N_35	82.5	79.5	22.0	60.4	34.9		Y	F	C	L	H	R	S
7SM_N_36	62.5	63.1	12.3	35.5	31.9		Y	H	F	H	H	S	I
7SM_N_37	67.4	50.6	26.4	65.9	38.0		Y	H	F	F	C	L	H
7SM_N_38	100.0	58.8	73.7	44.9	71.3		N	Y	L	R	R	N	H
7SM_N_40	82.2	86.4	43.2	64.1	55.9		R	N	L	K	D	V	D
7SM_N_41	100.0	49.0	62.3	41.5	67.1		N	Y	Y	K	F	N	N
7SM_N_42	39.3	51.4	8.2	34.2	31.8		Y	H	L	R	S	N	Y
7SM_N_43	92.2	55.5	49.0	61.5	62.9		N	N	F	N	D	Y	H
7SM_N_44	100.0	87.6	41.2	65.8	48.2		H	Y	F	S	R	L	N
7SM_N_45	71.0	87.0	29.3	59.8	42.5		F	H	F	T	S	G	S
7SM_N_46	91.8	92.5	32.9	72.1	46.8		H	H	F	D	S	I	Y
7SM_N_47	93.8	83.3	43.8	72.5	51.5		R	H	L	S	S	N	H
7SM_N_48	78.8	86.1	35.9	54.1	42.8		F	S	L	L	S	N	H
7SM_N_49	68.1	70.7	17.8	58.0	36.8		Y	H	F	Q	H	I	R
7SM_N_50	59.5	52.3	22.0	42.1	40.3		Y	H	F	R	R	I	R
7SM_N_51	87.8	91.9	28.7	63.4	44.5		H	N	F	R	N	S	H
7SM_N_52	76.1	64.7	11.3	39.3	27.6		H	C	F	Q	H	C	L
7SM_N_53	86.7	76.2	40.1	64.8	46.5		R	L	F	P	H	G	S
7SM_N_54	64.5	25.7	22.0	16.9	27.9		Y	S	Y	W	N	F	D
7SM_N_55	95.5	60.2	47.7	58.1	50.2		R	Y	F	V	Y	G	N
7SM_N_56	80.8	76.5	37.0	52.7	46.6		F	F	C	S	R	R	D
7SM_N_57	96.4	31.0	53.9	51.0	53.4		G	F	Y	H	H	C	R
7SM_N_58	71.8	88.0	24.6	60.6	40.4		H	H	F	T	N	R	N
NfsA_Ec	50.6	51.8	44.6	46.0	59.9								
Empty pUCX	0.0	1.8	7.9	1.9	14.9								

Amino acid makes up
 ≥ 25% of amino acids
 at this position

A	Growth inhibition (%)	75 μ M							L43	C45	E99	L103	R133	N134	H215	T219
		MTZ	12 μ M	RB6145	250 μ M	CB1954	27.5 μ M	NFZ								
8SM_N_30	27.8	2.2	35.0	33.8	9.2				Y	C	E	L	R	D	Y	D
8SM_N_32	16.9	20.6	0.0	34.2	1.8				F	V	E	I	R	R	H	N
8SM_N_33	47.3	42.8	62.2	56.9	33.2				F	V	E	I	R	N	F	I
8SM_N_35	44.3	27.6	53.4	46.5	19.7				F	G	E	I	R	D	Y	V
8SM_N_36	18.2	29.6	32.7	33.9	5.8				C	V	E	V	R	D	L	A
8SM_N_40	77.9	36.6	53.9	51.3	20.7				F	V	L	V	Y	Y	Y	S
8SM_N_41	38.0	71.2	21.5	63.6	14.3				V	C	E	L	R	G	F	S
8SM_N_42	23.7	33.1	22.6	30.2	10.5				N	V	E	V	R	D	F	H
8SM_N_43	89.0	38.8	50.4	37.7	3.1				V	C	R	I	Y	N	H	A
8SM_N_45	64.8	30.5	48.2	60.7	16.4				F	C	L	I	V	R	Y	T
8SM_N_46	31.5	30.4	46.4	73.6	21.9				L	V	V	I	R	N	C	Y
8SM_N_47	70.4	9.6	25.8	38.7	0.0				V	C	R	I	S	I	H	T
8SM_N_48	0.0	19.7	26.5	73.4	4.3				N	F	R	N	Y	L	H	F
8SM_N_52	55.5	71.3	36.1	64.1	27.0				C	C	E	I	R	R	L	I
NfsA_Ec	50.1	41.4	59.5	41.7	34.2											
pUCX_empty	0.0	0.0	2.9	3.5	4.3											

B	Growth inhibition (%)	75 μ M							L43	C45	E99	L103	R133	N134	H215	T219
		MTZ	12 μ M	RB6145	250 μ M	CB1954	27.5 μ M	NFZ								
8SM_C_2	86.3	10.6	50.3	40.0	42.6				Y	C	E	L	R	D	Y	D
8SM_C_3	71.5	28.1	70.1	34.5	41.9				N	V	E	I	R	D	V	S
8SM_C_7	60.5	29.0	39.1	49.1	25.6				L	V	E	I	R	Y	S	L
8SM_C_9	20.9	51.9	34.9	41.9	17.0				F	I	E	I	R	H	Y	T
8SM_C_10	68.2	21.1	76.6	15.7	39.6				I	V	E	I	R	D	L	N
8SM_C_11	26.6	48.4	34.9	56.7	22.1				F	I	E	I	R	H	Y	T
8SM_C_12	66.0	10.3	32.8	11.2	26.5				N	C	E	I	R	D	G	D
8SM_C_14	32.5	11.6	28.1	64.3	15.2				F	V	E	V	R	D	C	D
8SM_C_15	43.0	40.4	58.3	51.6	44.3				F	V	E	V	R	C	Y	N
8SM_C_20	43.1	59.6	47.7	70.1	25.4				I	C	E	I	R	H	L	F
8SM_C_25	40.3	38.3	55.2	55.5	47.0				I	V	E	V	R	C	H	I
8SM_C_26	40.9	49.9	55.5	72.3	28.5				I	V	E	I	R	D	I	F
8SM_C_29	76.5	46.5	63.6	52.6	31.4				H	V	E	L	R	D	R	D
8SM_C_35	21.9	40.2	17.5	43.8	14.9				L	V	E	V	R	C	F	I
8SM_C_38	65.3	28.2	52.2	37.2	48.3				V	C	E	I	R	D	G	V
8SM_C_39	11.0	8.5	4.9	8.9	7.5				Y	Y	G	S	S	H	F	V
8SM_C_41	57.4	42.1	58.5	64.4	35.4				I	G	E	I	R	C	I	I
8SM_C_42	60.1	25.8	39.0	50.2	41.7				F	C	E	I	R	I	G	I
8SM_C_44	87.0	39.7	23.1	61.5	63.7				L	C	V	I	V	I	Y	A
8SM_C_45	85.6	35.1	38.7	62.4	55.9				L	C	G	V	R	V	S	A
8SM_C_46	36.8	38.5	50.7	49.8	52.6				V	G	E	I	R	N	Y	I
8SM_C_47	38.8	30.0	32.4	54.4	31.8				F	V	E	V	R	S	S	I
8SM_C_50	59.1	62.1	55.3	66.7	56.2				L	S	E	I	R	R	L	S
8SM_C_52	42.8	41.3	50.8	73.8	20.7				C	C	E	V	R	S	R	D
8SM_C_53	3.2	17.0	8.0	24.0	8.8				L	C	E	I	R	H	I	L
8SM_C_55	52.5	41.1	54.3	58.7	49.1				S	V	E	V	R	C	N	S
8SM_C_56	77.0	41.2	23.5	76.7	48.8				L	C	V	V	V	F	Y	V
NfsA_Ec	52.4	51.2	60.3	63.1	58.1											
pUCX_empty	0.0	6.8	1.2	0.7	7.1											

Table 3.2: Percentage growth inhibition and sequencing results for 8SM *nfsA* variants selected on niclosamide and chloramphenicol. (A) *NfsA* variants selected on 5 μ M niclosamide. (B) *NfsA* variants selected on 3 μ M chloramphenicol. Growth inhibition was monitored by measuring the OD₆₀₀ pre- and post 4-hour incubation with each prodrug. Percentage growth inhibition was determined relative to unchallenged controls. Data are the average of three biological replicates (standard error of the mean less than 5, omitted for clarity). The amino acid present at each of the positions mutated is shown. An amino acid is shaded if it makes up $\geq 25\%$ of all amino acids at any given position.

The key points from Table 3.1 and Table 3.2 were summarised in Table 3.3 and used to determine the final eight positions to target. It was noted how frequently a position had an amino acid substitution and whether these substitutions gave rise to more active NfsA variants, indicated by higher percentage growth inhibition values. Amino acids were split into the five standard types; non-polar, polar, acidic, basic and aromatic. This was used to calculate the percentage of amino acids of a different type to the wild-type amino acid at the 15 positions mutated in the 7SM and 8SM libraries. This gave an indication of the variation being selected for at each position.

Position	% Variation at each position (compared to the wild type residue)		
	Niclosamide Selection	Chloramphenicol Selection	Average
S41	87	77	82
F42	66	63	64
L43	85	48	66
C45	78	59	68
F83	37	47	42
E99	70	15	42
L103	52	4	28
R133	67	11	39
N134	89	70	79
H215	85	89	87
T219	81	74	77
K222	70	72	71
S224	64	81	72
R225	84	70	77
F227	93	93	93

Table 3.3: Percentage variation at the 15 positions targeted in the 7SM and 8SM nfsA libraries. Variation was calculated by determining the percentage of residues at each position that were a different type (i.e., non-polar, polar, acidic, basic or aromatic) to the wild-type residue.

Some amino acid positions were easily eliminated from the eight final target positions. These included F83, E99, L103, and R133 due to the high level of conservation of the wild-type amino acid or other similar amino acids, suggesting that change was not beneficial at these positions. Positions S41, H215 and F227 all had the highest sequence diversity (> 80%) so were selected as the first three positions. The remaining five positions were decided between F42, L43, C45, N134, T219, K222, S224 and R225 which all still showed a high level of sequence diversity, between 64-79%. Initially, the eight amino acid positions with the greatest sequence diversity were selected: S41, N134, H215, T219, K222, S224, R225 and F227. Following a quote from Lab Genius, the company approached to synthesise

and clone the library, we were informed the cloning was charged per section of the gene mutated, rather than total number of positions mutated. Position N134 was replaced by position L43 resulting in only two sections of *nfsA* being targeted for mutagenesis versus three sections. This significantly decreased the cost of creating this library without substantially changing the mutation conformation. Position L43 was chosen over C45 as NfsA variants exhibited a more diverse range of amino acid substitutions at this position (Table 3.2).

3.3.3 Generation of a large simultaneous site-directed mutagenesis *nfsA* library

From this point in my thesis the simultaneous site-directed mutagenesis *nfsA* library will be referred to as the “Marsden” library, signifying the source of project funding and to avoid confusion with the previous 8SM library which also contained eight-targeted mutagenesis positions.

The Marsden library was constructed using DNA sourced from two different companies. Both sources utilised degenerate codons in place of regular codons at the eight-targeted positions. The degenerate codon NDT was used at positions S41, L43, H215, S224, R225 and F227, NHT at position T219 and VNG (V = adenine, cytosine, guanine) at position K222 to ensure the wild-type amino acid could be encoded at each of the positions (Table 3.4). The Marsden library encoded approximately 430 million different codon combinations and approximately 394 million unique amino acid combinations.

Degenerate codon	Amino acids encoded	Total Codons	Total Unique Amino Acids
NDT	Arg, Asn, Asp, Cys, Gly, His, Ile, Leu, Phe, Ser, Tyr, Val	12	12
NHT	Ala, Asn, Asp, His, Ile, Leu, Phe, Pro, Ser, Thr, Tyr, Val	12	12
VNG	Ala, Arg, Arg, Gln, Glu, Gly, Leu, Lys, Met, Pro, Thr, Val	12	11

Table 3.4: *Degenerate codons utilised in the Marsden library. Amino acids encoded by each degenerate codon are listed followed by the number of unique amino acids.*

3.3.3.1 Lab Genius DNA

The first source of DNA was a synthetic gene, pre-cloned into the expression vector pUCX, ordered from the company Lab Genius (*London, United Kingdom*). This company claimed they could provide coverage of all 394 million unique NfsA variants cloned into the pUCX

vector. We believed ordering the cloned library would significantly reduce the time and number of transformations required to obtain a library of this size, as purified plasmid is taken up more efficiently by electrocompetent cells than a ligation reaction. Lab Genius sent a total of 400 ng of lyophilized plasmid, claiming the use of commercial electrocompetent cells would generate enough transformants to obtain coverage of the Marsden library. Commercially available *E. coli*® 10G Supreme Electrocompetent cells, with an advertised transformation efficiency of $> 4 \times 10^{10}$ colony forming units per μg of pUC19 plasmid were used. Despite the use of these electrocompetent cells, a library containing only 30 million transformants was generated. In further tests comparing the Lab Genius DNA to a comparable concentration of purified plasmid DNA, an approximate 20-fold decrease in the number of transformants was obtained when using the Lab Genius DNA compared to purified plasmid. This showed that the Lab Genius DNA was not being taken up as efficiently as purified plasmid a library containing all 394 million unique NfsA variants could not be obtained using only the provided DNA.

A whole plasmid PCR was performed using 35 ng of the remaining Lab Genius DNA as template. This amount of DNA represented a predicted 6.6×10^9 copies of the plasmid and 17-fold coverage of the 394 million possible unique variants. A standard PCR protocol was conducted (Section 2.5.2) using a low cycle number was used (15 cycles) to minimise the introduction of mutations. Using this DNA an additional 90 million transformants were obtained, bringing the total number of transformants to 120 million.

3.3.3.2 GenScript® DNA

A second source of the Marsden library DNA was ordered as a synthetic gene from the company GenScript® (*Nanjing, China*) that was subsequently cloned into the vector of interest, pUCX. The gene was cloned into pUCX using our standard restriction cloning protocol (Section 2.5.3). To increase the number of transformants obtained, a DNA precipitation method using yeast tRNAs was optimised (Section 2.5.6.1). This method was found to increase the number of transformants obtained by 10-fold, compared to a ligation which had not been precipitated. This method, combined with an improved method for generating electrocompetent cells (Section 2.5.9.3), enabled another 676 million transformants to be obtained.

The total number of transformants from both the Lab Genius and GenScript® DNA was 796 million variants. This value includes transformants containing out of frame variants, variants with additional mutations outside of the eight-targeted residues and pUCX vector containing no insert.

3.3.3.3 Estimating the number of unique NfsA variants in the Marsden library

In order to estimate the number of unique NfsA variants present in the Marsden library, the proportion of transformants containing out of frame variants, variants with erroneous mutations or pUCX_empty vector, was estimated and subtracted from the total number of transformants obtained (Table 3.5). For each transformation conducted, a small sample of the resulting recovery culture was plated on LB agar supplemented with Amp. The resulting colonies were: (A) counted to estimate the total number of transformants present in each transformation, (B) sequenced to estimate the total number of transformants containing a variant which was out of frame or contained erroneous mutations (30 variants sequenced from each DNA source) and (C) PCR screened with plasmid specific primers to estimate the portion of the transformants that contained empty pUCX vector. These steps were performed using a sample of transformants obtained from solid media and it was assumed these transformants were representative of the remaining recovery culture, which was grown in a larger liquid culture (Section 2.6.2).

	Total Number Transformants	Portion Out of Frame	Portion Additional Mutations	Portion pUCX_empty Vector	Total Number Transformants
Lab Genius Transformants	120,000,000	0.242	0.09	0	80,000,000
GenScript® Transformants	676,000,000	0.167	0.04	0.323	318,000,000

Number Transformants With No anomalies	398,000,000

Table 3.5: Number of transformants with no anomalies. The number of transformants with no anomalies was determined by estimating and subtracting the portion of transformants which contained out of frame mutations, erroneous mutations and empty pUCX vector.

Of the 796 million transformants obtained, 33.2% of the Lab Genius derived transformants and 53% of the GenScript® derived transformants contained a plasmid with some anomaly (primarily frameshift mutations for the former and pUCX_empty plasmid the latter). This meant the total number of transformants containing no anomalies was 398 million. While this number is larger than the number of unique NfsA variants possible in the Marsden

library (394 million), it is essentially impossible that full coverage of the library had been obtained. Using the GLUE web interface, a programme designed to estimate the completeness in protein-encoding libraries based on a random distribution of clones, the expected completeness of the Marsden library was found to be 64% (Patrick *et al.*, 2003). This meant the Marsden library had coverage of an estimated 252 million NfsA variants. In order to obtain 95% completeness of the Marsden library for screening purposes, it would have been necessary to create a library of greater than 1 billion transformants.

3.3.3.4 Growth advantage of transformants containing pUCX_empty plasmid

When creating the Marsden library, it was decided to grow transformants in a liquid culture as opposed to on solid media, due to the prohibitive number of agar plates which would be required to plate individual colonies on, and then scrape together to generate a pooled library. In order to achieve desired coverage of the Marsden library, the library would need to be plated over 40,000 agar plates (assuming a maximum number of 10,000 colonies per agar plate). As this was not feasible, the transformants were grown in a liquid culture for 6 hours to grow the library which was then spun down and pooled to generate the final Marsden library (Section 2.6.2).

When calculating the final size of the Marsden library, a key assumption was that the proportion of transformants containing pUCX_empty in library would remain the same following growth in the liquid culture. Unlike on solid media, in liquid media there is a direct competition between cells, therefore *E. coli* expressing a large or otherwise fitness-impairing insert have a growth disadvantage. It was subsequently discovered that the growth advantage *E. coli* 7NT cells expressing pUCX_empty had over *E. coli* 7NT cells expressing pUCX containing a *nfsA* variant had not been accounted for. When a sample of the final Marsden library, following growth in the liquid culture, was plated on solid agar it was found that the percentage of *E. coli* 7NT expressing pUCX_empty had increased from approximately 27% to 74%, substantially changing the overall library composition.

A retrospective experiment was performed to see how quickly the ratio of *E. coli* 7NT cells expressing pUCX_empty versus pUCX_nfsA could change in a 6-hour period. A 2 mL culture containing 27% *E. coli* 7NT expressing pUCX_empty and 73% *E. coli* expressing pUCX_nfsA was added to 50 mL LB and grown in a baffled flask for 6 hours, mimicking the conditions used when creating the library (Section 2.6.2). Samples of the culture were

taken at 1 hour intervals over a 6 hour period and plated on LB agar plates supplemented with Amp. The proportion of pUCX_empty and pUCX_NfsA was estimated by testing a sample of the resulting *E. coli* 7NT colonies (114 colonies plus controls) in growth inhibition assays with niclosamide. *E. coli* 7NT cells which grew when exposed to niclosamide, indicated that the cells were expressing pUCX_nfsA, while *E. coli* 7NT cells which died when exposed to niclosamide indicated that the cells were expressing pUCX_empty. The percentage of *E. coli* 7NT expressing pUCX_empty substantially changed over time, eventually making up nearly 100% of the culture (Figure 3.2).

Figure 3.2: Ratio of *E. coli* 7NT cells expressing pUCX_empty to pUCX_nfsA at different time points in a liquid culture. Samples of the culture were taken at t = 0, 1, 2, 3, 4, 5 and 6 and plated over LB agar plates supplemented with Amp. 114 colonies were selected from these plates and tested with 5 μ M niclosamide in growth inhibition assays alongside *E. coli* 7NT cells expressing pUCX_empty and pUCX_nfsA as controls to determine whether a cell was expressing pUCX_empty or pUCX_nfsA. This data is representative of one experiment.

To account for the high proportion of *E. coli* 7NT cells expressing pUCX_empty, adjustments were made to determine the number of colonies required to be plated in order to obtain coverage of the 252 million unique NfsA variants present in the Marsden library. Although the ratio of transformants containing pUCX_empty to pUCX_nfsA variants substantially changed due to the growth disadvantage of transformants bearing nfsA variants, we would still expect the total number of pUCX_nfsA variants present in the Marsden library to be 252 million. The final composition of the Marsden library following growth in the liquid culture was 74% pUCX_empty, 6.4% out of frame vector, 1.7% containing mutations outside of the eight-targeted residues and 17.8% containing desirable inserts (assuming that transformants containing pUCX_nfsA variants with or without mutations would have the same growth disadvantage). This meant if 1000 colonies were

plated, only 178 (on average) would be expressing a desirable *nfsA* variant, with the majority (740 colonies) containing *pUCX_empty*. Therefore, in order to plate 95% coverage of the 252 million variants in the Marsden library it would be necessary to plate a minimum of 4.2 billion clones ($252,000,000/0.178 = 1,400,000,000$ then used the GLUE web interface to determine 95% coverage of this number; Patrick *et al.*, 2003).

3.3.4 Using a positive selection to recover superior NfsA variants

3.3.4.1 Chloramphenicol versus niclosamide as a positive selection compound

The aim of this chapter was to evolve NfsA variants exhibiting superior detoxification of chloramphenicol or niclosamide which could be used as models in future experiments to understand enzyme evolution and evolutionary trajectories (Chapter 4). Initially we sought to recover and compare variants that had been selected for maximal activity with either niclosamide or chloramphenicol. However, pilot testing revealed that only chloramphenicol was able to provide a sufficiently stringent selection to distinguish the top few variants from the remainder of the library (Table 3.6). Due to limitations of solubility in aqueous solutions, the maximum achievable concentration of niclosamide was 50 μM , which was only sufficient to eliminate 98% of variants from the pool. Moving forward it was decided to focus on chloramphenicol as the sole selection agent, and this was used at 45 μM or 47.5 μM .

Concentration Chloramphenicol (μM)	% Selection	Concentration Niclosamide (μM)	% Selection
1.5	0.6	0.5	18
3.0	0.08	5	5
20	0.03	20	3
25	0.008	50	2
30	0.0006	Growth cut off wild-type NfsA	10 μM niclosamide
35	0.0003		
40	0.00008		
45	0.00004		
47.5	0.000006		
50	0		
Growth cut off wild- type NfsA	3 μM chloramphenicol		

Table 3.6: The percentage of the Marsden library able to grow on various selective concentrations of chloramphenicol or niclosamide. A sample of the Marsden library was resuspended in LB medium and serial dilutions were created as per Section 2.6.3. The concentrated cell resuspension was plated on the selection conditions above and a sample of diluted culture was plated on unselective media (LB agar supplemented with Amp). The number of colonies which grew on the unselective media were counted and multiplied by the dilution factor to estimate the total number of colonies plated on the selective media. This was then used to calculate the % selection.

3.3.4.2 Positive selection using chloramphenicol.

14.8 billion clones were plated over large agar plates containing 45 μM or 47.5 μM chloramphenicol. Using the GLUE interface, the set of 14.8 billion clones was predicted to comprise virtually 100% of the 252 million unique NfsA variants present in the Marsden library (Patrick *et al.*, 2003). A total of 365 colonies were recovered from the selection. These colonies were picked into the inner 60 wells of 96-well plates alongside *E. coli* 7NT expressing wild-type *nfsA* and *pUCX_empty* as controls. All selection tests were performed on media containing no IPTG as high concentrations of IPTG appeared toxic in solid media as substantially less colonies grew (a phenomenon that had been observed by other members of the Ackerley lab). As the pUCX vector used in the Marsden library has leaky expression, omission of IPTG from the solid media still resulted in substantial expression of the cloned *nfsA* variants (Figure 3.3).

Figure 3.3: SDS-PAGE showing the relative protein expression levels of ten *E. coli* 7NT strains expressing pUCX_nfsA variants in the absence of IPTG. Overnight *E. coli* 7NT cultures expressing pUCX_nfsA constructs were used to inoculate medium containing lacking IPTG and incubated for 4 hours at 30°C. Samples were normalised based on cell density. An *E. coli* 7NT strain expressing pUCX_empty was included as a negative control. Predicted molecular mass of NfsA = 26.8 kDa.

3.3.5 Characterisation of the top chloramphenicol reducing NfsA variants

3.3.5.1 Chloramphenicol screening and sequencing of NfsA variant

Standard growth inhibition assays were conducted on the 365 colonies recovered in Section 3.3.4.2, to test the ability of each NfsA variant to detoxify 30 μ M chloramphenicol (Section 2.8.1). Plasmids encoding the top 114 NfsA variants that protected *E. coli* 7NT against chloramphenicol to the greatest extent were sequenced (representing the inner wells two 96-wells plates with 3-wells reserved for controls).

	S41	L43	H215	T219	K222	S224	R225	F227
Non-polar	1	54	30	15	63	32	30	20
Polar	2	45	23	0	13	21	54	18
Acidic	0	0	17	0	10	9	8	28
Basic	9	0	18	4	14	32	4	26
Aromatic	89	1	13	82	0	6	4	8

Table 3.7: Percentage of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of the top 114 NfsA variants selected on 45 and 47.5 μ M chloramphenicol. The property of the wild-type amino acid at each position is boxed in red.

In the 114 *nfsA* sequences obtained, 15 duplicate sequences were recovered encoding seven different NfsA variants. This was surprising, as a predicted 11-fold coverage of the 252 million unique Marsden variants had been plated, it was anticipated that a larger proportion of duplicate sequences would have been identified. Despite the lack of duplicate sequences, clear sequence patterns were evident among the 114 *nfsA* variants, suggesting

an effective enrichment for the top chloramphenicol detoxifying variants had still occurred. Notable patterns included substitutions to aromatic residues at positions 41 and 219, with 89% and 82% of the NfsA variants sequenced having this substitution respectively (Table 3.7).

3.3.5.2 Chloramphenicol IC₅₀ assays with top *nfsA* variants expressed in *E. coli* 7NT

The top 30 unique NfsA variants that protected *E. coli* 7NT to the greatest extent against chloramphenicol were taken forward for more sensitive IC₅₀ growth assays (Section 2.8.2). Prior to this, plasmids encoding the NfsA variants were re-transformed into *E. coli* 7NT cells to ensure any changes in activity were due to mutations introduced into the nitroreductase gene and not acquired chromosomal mutations. An example of the raw IC₅₀ data is shown in the appendices (Appendix A).

NfsA Variant	IC ₅₀ (μM)	S41	L43	H215	T219	K222	S224	R225	F227
36_37	42 ± 7	Y	L	C	Y	V	R	V	G
20_39	40 ± 6	Y	L	N	Y	R	Y	D	H
36_3	39 ± 8	H	S	G	Y	V	H	C	H
19_22	38 ± 4	Y	S	V	F	Q	N	N	R
37_4	38 ± 9	Y	S	G	F	M	H	C	H
35_58	37 ± 5	Y	Y	N	L	E	R	D	H
22_8	37 ± 5	F	C	G	F	R	N	N	H
36_51	37 ± 7	Y	V	N	Y	A	N	G	D
35_45	37 ± 3	Y	V	G	Y	V	H	V	G
21_17	36 ± 8	Y	V	G	Y	A	Y	C	D
36_6	36 ± 9	F	V	F	H	R	D	F	D
21_22	36 ± 6	Y	V	G	F	E	R	N	D
37_5	36 ± 6	Y	C	D	F	L	V	N	G
19_49	35 ± 5	Y	L	D	F	K	I	S	G
22_4	35 ± 4	Y	S	N	F	A	I	N	Y
36_15	35 ± 7	F	L	V	Y	A	H	V	H
37_18	35 ± 5	Y	L	D	L	T	R	N	L
36_22	35 ± 4	Y	C	D	F	K	I	C	D
20_48	34 ± 5	Y	L	D	Y	R	F	N	L
20_5	34 ± 6	Y	C	F	F	Q	H	C	C
36_42	32 ± 8	Y	V	G	Y	T	R	C	D
36_17	32 ± 6	Y	V	D	F	L	V	N	F
35_38	32 ± 5	F	C	H	F	T	S	I	D
36_38	30 ± 6	F	L	D	F	V	R	N	G
36_29	30 ± 4	Y	C	F	F	V	S	C	F
21_49	29 ± 6	Y	C	H	Y	E	N	V	G
36_47	28 ± 5	Y	C	H	Y	A	D	V	S
21_24	28 ± 3	F	S	I	F	E	R	V	G
35_44	27 ± 5	Y	L	I	L	G	L	L	D
36_41	23 ± 6	F	L	Y	F	T	Y	N	Y
NfsA	4 ± 0.6								

Amino acid makes up
≥ 25% of all amino acids
at this position

Table 3.8: Chloramphenicol IC₅₀ values and amino acid substitutions for the *E. coli* 7NT cells expressing top 30 *nfsA* variants. *E. coli* 7NT cultures expressing *nfsA* variants were challenged for 4 hours across serial dilutions of chloramphenicol (0-100 μM). Growth inhibition was determined by measuring the OD₆₀₀ of cultures pre- and post the 4-hour challenge. Percentage growth relative to an unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represent the average of five biological repeats ± SD. The amino acid present at each of the eight-targeted positions in the Marsden library is shown. An amino acid is shaded in grey if it made up ≥ 25% of all amino acids at any given position.

All NfsA variants tested were able to protect *E. coli* 7NT cells to substantially higher concentrations of chloramphenicol than *E. coli* 7NT cells expressing wild-type *nfsA* (Table 3.8). The *E. coli* 7NT strain expressing *nfsA* 36_37 was found to have the highest IC₅₀ value, with an almost 10-fold improvement in IC₅₀ over the *E. coli* 7NT strain expressing wild-type *nfsA*. NfsA 36_37 had seven amino acid substitutions (S41Y, H215C, T219Y,

K222V, S224R, R225V and F227G). The second-best variant, NfsA 20_39 also contained seven amino acid substitutions, with two of these the same as NfsA 36_37 (S41Y, H215N, T219Y, K222R, S224Y, R225F, F227H).

3.3.5.3 Amino acid conservation patterns in the top chloramphenicol detoxifying NfsA variants

The most common amino acid substitutions evident in the top 30 chloramphenicol detoxifying NfsA variants (Table 3.8) are depicted in the modelled protein structure in Figure 3.4.

Figure 3.4: Enzyme model of NfsA depicting the most common substitutions present in the top 30 chloramphenicol detoxifying variants. (A) Wild-type amino acid at positions S41, T219, R225 and F227. (B) Model of the most common amino acid substitutions in the top 30 chloramphenicol detoxifying variants; S41Y, T219F, R225N and F227D. Images were created using the mutagenesis function in PyMOL, PDB accession number 1f5v. The rotamer showing the smallest amount of steric hindrance was chosen. Targeted residues are shown in pink and the FMN molecule is shaded yellow. Only one of the active sites present in the homodimer is shown, with each monomer shaded either blue or green.

The S41Y substitution introduces a large aromatic residue into the active site pocket of NfsA in place of the small polar amino acid serine. Modelling in PyMOL predicts that the tyrosine residue lies planar to the FMN cofactor, therefore could be involved in pi-pi stacking interactions, stabilising chloramphenicol between the aromatic rings of FMN and the tyrosine residue (Figure 3.4). Despite ongoing efforts in the Ackerley lab group and other groups, there are currently no solved crystal structures of NfsA with chloramphenicol or any other nitroaromatic substrate bound. Therefore, we can only predict how nitroaromatic substrates may bind in the active site of NfsA based on modelling. Crystal

structures for nitroreductases from other families have been solved with other nitroaromatic substrates bound such as nitrofurazone and CB1954 (Fu *et al.*, 2005; Race *et al.*, 2005; Ryan *et al.*, 2011). In these examples, the nitroaromatic substrate lies planar to the cofactor molecule, and it is possible that chloramphenicol will bind in a similar orientation in NfsA.

Position 225 is also located in a close proximity to the active site. Unlike at position 41, the large arginine residue is substituted to a smaller asparagine residue, enlarging the active site pocket (Figure 3.4). This may enable larger nitroaromatic substrates such as chloramphenicol to sit within the active site and also compensate for larger amino acid substitutions such as S41Y. Positions 219 and 227 are located further away from the active site and are unlikely to be involved direct in catalysis or substrate stabilisation (Figure 3.4). Substitutions at these positions may help to stabilise and facilitate other mutations such as S41Y and R225N, improving the overall stability of the enzyme. The epistatic interactions between these and other residues was examined in more detail in Chapter 4.

3.3.5.4 Reduction of chloramphenicol by purified His₆-tagged NfsA variants

The top five chloramphenicol reductases were purified as His₆-tagged recombinant proteins to assess whether the improved activity observed *in vivo* was evident *in vitro*. Genes encoding *nfsA* 36_37, 20_39, 36_3, 19_22 and 37_4 were cloned into the His₆-tag expression vector pET28a+ and purified using nickel affinity chromatography (Section 2.8.5, Appendix A). Michaelis-Menten steady state kinetic parameters were measured by monitoring the depletion of NADPH at 340 nm as an indirect measure of chloramphenicol reduction (Section 2.8.6.2). As nitroreductases exhibit a ping-pong bi-bi kinetic mechanism involving dual substrates (NADPH and chloramphenicol in this example), the rate of reduction of chloramphenicol varies depending on the concentration of NADPH. Therefore, the kinetic parameters presented in Table 3.9, describe apparent K_M and k_{cat} at 250 μ M NADPH (however, the k_{cat}/K_M is not ‘apparent’, as this ratio is not affected by varying NADPH concentrations).

	k_{cat} (s ⁻¹)*	K_M (μM)*	k_{cat}/K_M (s ⁻¹ /M ⁻¹)	NADPH oxidase activity (s ⁻¹)
NfsA 36_37	0.13 ± 0.02	180 ± 70	720 ± 310	0.07 ± 0.007
NfsA 20_39	0.034 ± 0.003	57 ± 20	600 ± 190	0.05 ± 0.002
NfsA 36_3	0.086 ± 0.003	130 ± 20	660 ± 90	0.04 ± 0.004
NfsA 19_22	0.12 ± 0.1	99 ± 30	1200 ± 380	0.1 ± 0.09
NfsA 37_4	0.048 ± 0.003	190 ± 90	250 ± 50	0.04 ± 0.001
Wild-type NfsA	0.82 ± 0.08	2200 ± 480	370 ± 90	0.07 ± 0.002

* apparent values as measured at 250 μM NADPH

Table 3.9: Kinetic parameters for the reduction of chloramphenicol by the top five NfsA variants. The initial rates of reduction at varying concentrations of chloramphenicol and 250 μM NADPH were monitored at 340 nm using an extinction coefficient of 12,440 M⁻¹cm⁻¹. Reactions were measured in triplicate. Apparent K_M and k_{cat} were calculated using Graphpad 8.0. NADPH oxidase activity was determined by measuring the rate of NADPH reduction in the absence of chloramphenicol. Michaelis-Menten plots shown in Appendix A.

The low turnover rate of chloramphenicol exhibited by the NfsA variants made measuring kinetic parameters difficult, due to the intrinsic NADPH oxidase activity of NfsA. This occurs when electrons are transferred directly from NADPH to dissolved O₂ present in the reaction medium. The intrinsic NADPH oxidase activity is normally negligible as the turnover rate of NfsA with many other nitroaromatic compounds is significantly higher than the intrinsic NADPH oxidase activity (Valiauga *et al.*, 2018). However, the intrinsic NADPH oxidase activity can confound reaction measurements when the substrate turnover is low, as was observed here for chloramphenicol. This has also been encountered as a problem when measuring the reduction of metronidazole and other 5-nitroimidazole compounds, for which NfsA also has a slow turnover rate (Olekhovich *et al.*, 2009; Williams, 2013). A NADPH regeneration method had been developed to measure kinetics with metronidazole, allowing the reduction of metronidazole (which also absorbs at 340 nm) to be measured independently of NADPH turnover (Section 2.8.6.3). However, this NADPH regeneration method could not be used for chloramphenicol kinetics, as the reduction of chloramphenicol cannot be measured directly (no substantial change in absorbance in the visible spectrum). As a consequence, the NADPH oxidase activity could not be nullified, leading to the relative high errors in k_{cat}/K_M evident in Table 3.9. Measurements of NADPH depletion in the absence of substrate confirmed that all NfsA variants exhibited a similar rate of intrinsic NADPH oxidase activity to wild-type NfsA.

All of the NfsA variants exhibited improved K_M values for chloramphenicol relative to wild-type NfsA, however exhibited substantially lower turnover rates (k_{cat} ; Table 3.9). While the overall catalytic efficiency (k_{cat}/K_M) for four out of five NfsA variants was

improved over wild-type NfsA, ranging from a 1.6 to 3.2-fold improvement, these improvements seemed surprisingly low relative to the substantial improvements in IC_{50} observed *in vivo* (Table 3.8). This trend of improved K_M for evolved substrate but decreased substrate turnover (k_{cat}) has been observed by several researchers in the Ackerley group following directed evolution of nitroreductases (Williams, 2013; Copp *et al.*, 2017; Rich, 2017; Sharrock, 2018). It has been postulated that substantially enhanced activities observed *in vivo* are not always a consequence of improved catalytic efficiency, but can also be caused by an enhanced specificity for the evolved substrate over competing intracellular metabolites.

3.3.5.5 Reduction of *p*-benzoquinone by purified His₆-tagged NfsA variants

The reduction of water-soluble quinones is believed to be one of the primary biological roles of the NfsA and NfsB family nitroreductases (Liochev *et al.*, 1999). *P*-benzoquinone is a water-soluble quinone known to be present in the *E. coli* intracellular environment. It has been theorised that directed evolution campaigns where screening is conducted *in vivo* may select for nitroreductase variants which have a diminished affinity for competing metabolites such as *p*-benzoquinone (Copp *et al.*, 2017). These variants may exhibit an apparent improved activity with the desired substrate *in vivo*, however when tested *in vitro* in the absence of competing metabolites, these variants no longer have an advantage over the parental form of the enzyme and do not appear substantially improved (or may even seem worse).

To test whether the same phenomenon was manifesting here, the ability of the top five purified His₆-tagged NfsA variants to reduce *p*-benzoquinone was assessed. Michaelis-Menten steady state kinetic parameters were measured by monitoring the depletion of NADPH at 340 nm as an indirect measure of *p*-benzoquinone reaction (Table 3.10).

	k_{cat} (s ⁻¹)*	K_M (μM)*	k_{cat}/K_M (s ⁻¹ /M ⁻¹)	Rate with 100 μM <i>p</i> -benzoquinone and 250 μM NADPH (s ⁻¹)
Wild-type NfsA	8.4 ± 0.3	18 ± 2	470,000 ± 54,000	7.0 ± 0.3
NfsA 36_37	N.D.	N.D.	N.D.	0.06 ± 0.01
NfsA 20_39	N.D.	N.D.	N.D.	0.02 ± 0.007
NfsA 36_3	N.D.	N.D.	N.D.	0.1 ± 0.001
NfsA 19_22	N.D.	N.D.	N.D.	0.06 ± 0.02
NfsA 37_4	N.D.	N.D.	N.D.	0.05 ± 0.003

*apparent values measured at 250 μM NADPH

Table 3.10: Kinetic parameters for the reduction of *p*-benzoquinone by the top five NfsA variants. The initial rates of reduction at varying concentrations of *p*-benzoquinone and 250 μM NADPH were monitored at 340 nm using an extinction coefficient of 6,220 M⁻¹cm⁻¹. Reactions were measured in triplicate. Apparent K_M and k_{cat} were calculated using Graphpad 8.0. Where the k_{cat} and K_M could not accurately be determined (N.D.) the rate of *p*-benzoquinone reduction was measured at a single high concentration (100 μM) with 250 μM of NADPH (additional column to right of main table). The Michaelis-Menten plot for wild-type NfsA is shown in Appendix A.

Accurate kinetic parameters for the reduction of *p*-benzoquinone by the top five NfsA variants could not be determined as the low turnover rate of *p*-benzoquinone could not be reliably measured over the high intrinsic NADPH oxidase activity. A low level of reduction was observed at a high concentration of *p*-benzoquinone (100 μM) and this reduction rate is presented as a measure of *p*-benzoquinone reduction (Table 3.10). All five evolved NfsA variants exhibited substantially lower reduction rates of *p*-benzoquinone, 70- to 350-fold lower than wild-type NfsA with 100 μM *p*-benzoquinone and 250 μM NADPH. This suggests these NfsA variants have evolved to discriminate for prodrugs over intracellular competing metabolites such as *p*-benzoquinone, therefore appeared to have overall improved activity when measured *in vivo*.

Nevertheless, the primary aim of this chapter was to identify the top chloramphenicol detoxifying NfsA variants, and it was not essential that this be driven by superior biochemical activity *in vitro*. Based on the results presented here and the relative ease of IC₅₀ assays compared to purified enzyme kinetics, it made sense to confine our analysis to chloramphenicol detoxification *in vivo* as the primary measure of evolved enzyme activity.

3.3.6 Mutations in the plasmid backbone of selected clones

Partway into completing work for Chapter 4, mutations were identified in the backbones of the plasmids bearing several of the top *nfsA* variants. Particularly prevalent were mutations in the lac repressor (*lacI*) region of the pUCX plasmid. Seven of the top ten *nfsA* plasmids sequenced were found to contain mutations in the *lacI* region. This region encodes the Lac

repressor, which usually binds tightly to the *lac* operon, preventing transcription of downstream elements. The addition of an inducer such as IPTG, prevents the Lac repressor from binding to the *lac* operon and can be used to control and induce expression (Jacob and Monod, 1961). It is not uncommon for mutations to occur in the *lac* repressor gene and these mutations have been shown to frequently decrease the stability of the Lac repressor, resulting in increased expression of downstream elements (Lewis, 2005). *E. coli* 7NT cells expressing plasmids bearing these mutations would have a profound advantage in conditions containing a low level of IPTG, such as the conditions used in the positive selection (Section 3.3.4.2). We had not considered the possibility of mutations in the backbone until after we had re-cloned these *nfsA* variants for work in Chapter 4. Prior to this discovery, the Marsden library had also been used as a source of NfsA variants in Chapters 5, 6 and 7. We wanted to determine the consequence this discovery would have on the work which had been conducted in these other chapters.

3.3.6.1 Investigating the selective advantage of *E. coli* 7NT cells expressing plasmids bearing mutations in the *lacI* gene

The genes encoding the top ten chloramphenicol detoxifying NfsA variants were re-cloned into fresh pUCX vector to remove the mutations present in the backbone of the plasmid. Equivalent plasmids with or without mutations in the *lacI* gene were expressed in *E. coli* 7NT cells and the relative expression levels of the resulting *nfsA* variants were compared when in the presence of both a low (5 μ M) and high (50 μ M) concentration of IPTG. The plasmids encoding NfsA variants 36_3, 37_4 and 35_45 did not contain mutations in the *lacI* gene and these genes were re-cloned as a controls. Chloramphenicol detoxification and enzyme expression levels were quantified by IC₅₀ assays (Figure 3.5) and SDS-PAGE gels (Figure 3.6) respectively.

Figure 3.5: Chloramphenicol IC₅₀ assays of *E. coli* 7NT cells expressing *nfsA* variants with or without mutations in the *lacI* region. *E. coli* 7NT cells expressing the originally recovered *nfsA* constructs were tested in IC₅₀ assays (dotted bar) alongside *E. coli* 7NT cells expressing the equivalent re-cloned construct (no pattern). Assays were conducted in the presence of both 5 μM IPTG (A) and 50 μM IPTG (B). Constructs expressing *nfsA* variants 36_3, 37_4 and 35_45 did not contain mutations in the *lacI* region however were re-cloned as a control. Data are the average of three biological repeats ± SD. Differences in IC₅₀ values were analysed for statistical significance using an unpaired t-test (*** = p-value ≤ 0.001, ** = p-value ≤ 0.01, * = p-value ≤ 0.05).

Figure 3.6: SDS-PAGE gel showing relative nitroreductase enzyme levels of *E. coli* 7NT strains expressing pUCX_ *nfsA* constructs with or without mutations in the *lacI* region. Overnight *E. coli* 7NT cultures expressing pUCX_ *nfsA* constructs with (*) and without mutations in the *lacI* region were used to inoculate medium containing 5 or 50 μM IPTG and incubated for 4 hours at 30°C. Samples were normalised based on cell density. An *E. coli* 7NT strain expressing pUCX_ empty was included as a negative control. NfsA = 26.8 kDa.

In the presence of 5 μM IPTG (Figure 3.5A) there was a significant difference in IC₅₀ value between *E. coli* 7NT strains expressing pUCX_ *nfsA* constructs containing mutations in the *lacI* gene (dotted bars) versus no mutations in the *lacI* gene (no pattern bars). The pUCX_ *nfsA* constructs in the presence of 5 μM IPTG rendered their *E. coli* hosts far less sensitive to chloramphenicol (i.e., have high chloramphenicol IC₅₀ values) compared to the sensitivities conferred at 50 μM IPTG (Figure 3.5B). This showed that the mutations

present in the *lacI* gene construct had indeed conferred a profound advantage, likely caused by a non-functional Lac repressor resulting in expression levels comparable to high levels of the inducer IPTG. The only variants which did not exhibit a significant difference in host cell sensitivities between the original and re-cloned constructs were *nfsA* 36_3, 37_4 and 35_45, all of which did not have mutations present in or close to the *lacI* gene. This correlates with the results seen in Figure 3.6, as more protein was present in the *E. coli* strains expressing *pUCX_nfsA* constructs with mutations in the *lacI* gene in the presence of 5 μ M IPTG compared to *E. coli* strains expressing the equivalent construct without mutations. This discrepancy, however, was not observed in the presence of 50 μ M IPTG.

All *nfsA* variants containing mutations in the *lacI* gene originated from the Lab Genius DNA. It is likely that during the synthesis of the Lab Genius DNA and/or during the PCR amplification of this DNA, mutations were introduced throughout the entire plasmid, as in each case the *nfsA* gene and vector were synthesised as one. In comparison, the GenScript® DNA was cloned using restriction cloning, which limits the likelihood of mutations occurring within the plasmid. As the Marsden library had been used in several downstream applications prior to the discovery of these mutations, we did not want to remake the library without the Lab Genius DNA. Instead, all growth assays of NfsA selected variants presented in this thesis were conducted at 50 μ M IPTG to negate the effect mutations in the *lacI* gene would have on activity. This includes the IC₅₀ data previously presented in this chapter (Table 3.8) which was ultimately performed the presence of 50 μ M IPTG to ensure that any improvement in nitroreductase activity was the result of mutation(s) in the *nfsA* variant not from mutations in the backbone leading to improved expression.

3.4 Discussion

The research described in this chapter involved the identification of several NfsA variants that exhibited substantial improvements in chloramphenicol detoxification over wild-type NfsA. The best NfsA variants were able to protect *E. coli* 7NT cells against an almost 10-fold higher concentration of chloramphenicol than *E. coli* 7NT strains expressing wild-type *nfsA*. These NfsA variants were isolated from the Marsden library, a simultaneous site-directed mutagenesis library. The residues targeted in the Marsden library were chosen from two preliminary libraries created in the Ackerley lab (7SM and 8SM libraries). Despite our best efforts to identify eight residues which exhibited high levels of variation, one position in particular, L43, was infrequently mutated in the top chloramphenicol

detoxifying NfsA variants (Table 3.7 and Table 3.8). The small-scale sequence analysis of *nfsA* variants recovered from the 8SM library showed that position 43 was more frequently mutated to other amino acid types in NfsA variants selected on niclosamide (85%) versus on chloramphenicol (48%). However, when deciding the final eight residues we had not decided which positive selection (niclosamide or chloramphenicol) was going to be used. Selecting an alternative eighth residue such as F42, C45 or N134 may have resulted in the generation of even better chloramphenicol detoxifying NfsA variants. The use of two separate preliminary libraries may have also been an imperfect strategy, as we may have negated the full potential of some positions. For example, a substitution from one library (e.g., at position E99, L103 or R133 from the 8SM library) may have provided a fitness benefit with a substitution from the 7SM library. However, it was necessary to keep the number of targeted positions at eight to keep the library sizes manageable.

One main problem encountered during library construction was the high portion of *E. coli* 7NT cells expressing the *pUCX_empty* construct generated during the liquid growth culture. To avoid this occurring in future, any *pUCX_empty* vector would need to be eliminated from the DNA source prior to the transformation step. A gel extraction of the digested *pUCX* vector could be used to remove undigested *pUCX_empty* vector. This was trialled, however invariably yielded too low an amount of DNA to make large-scale library generation possible. A restriction site unique to undigested *pUCX_empty* could also be included to ensure all *pUCX_empty* vector is removed prior to transformation. Recombinational cloning could also be used as an alternative to restriction enzyme-based cloning. This method is mediated by site-specific recombinase enzymes and eliminates the need for restriction enzymes and ligase (Festa *et al.*, 2013). Several examples using this technology for library construction have been reported in the literature, although only achieving library sizes with a maximum of 10^4 variants (Rolfs *et al.*, 2008; Wingler and Cornish, 2011). Further testing would need to be done to see if a library of substantially greater magnitude could be created using this method.

Despite an estimated 11-fold coverage of the 252 million unique *nfsA* variants created in the Marsden library being plated on selective media containing chloramphenicol, only 15 duplicate sequences were identified in the 114 *nfsA* variants sequenced. This may be due to the high number of *E. coli* plated on each selection plate, meaning not every clone had an equal opportunity to grow on the selective media. Although it was unlikely that the

all the *nfsA* variants from the Marsden library were included in the positive selection, clear sequence patterns were evident among the top chloramphenicol detoxifying NfsA variants. The contributions of these different amino acid substitutions to the overall activity of the enzyme were investigated further in Chapter 4, looking at whether each substitution was beneficial individually, or whether the substitutions were interactive (i.e., epistasis), to improve our understanding of enzyme evolution.

The top five chloramphenicol detoxifying variants were also purified as His₆-tagged recombinant proteins to assess chloramphenicol reduction *in vitro*. The activity measured *in vitro* appeared to be substantially less than the level of resistance conferred *in vivo*. It was postulated that this was the result of recovering NfsA variants using an *in vivo* selection, where NfsA variants had evolved a diminished affinity for competing metabolites, therefore exhibited an apparent improved activity with the desired substrate when tested *in vivo*. All five NfsA variants tested exhibited substantially less activity with *p*-benzoquinone, a predicted metabolite of NfsA, supporting this hypothesis. This shows the importance of considering the selection used to recover variants exhibiting a desired property(s), as directed evolution may find an alternative way to respond to a selection without improving catalytic activity. This reflects the frequently quoted first law of directed evolution “you get what you screen for” (Arnold, 1998; Schmidt-Dannert and Arnold, 1999). However, for the current study it was not essential that the selected NfsA variants exhibited improved reduction *in vitro*. Having determined that there was not necessarily a direct correlation between chloramphenicol detoxification *in vivo*, and chloramphenicol reductase activity measured *in vitro*, we decided to focus on the former as our primary measure of enzyme fitness when using these NfsA variants as a model to understand enzyme evolution and evolutionary progression (Chapter 4).

Chapter 4 Using NfsA to model evolutionary trajectories

4.1 Introduction

Most conventional directed evolution studies fail to address the role of epistasis and other such phenomena. Epistasis describes interactions between mutations, such that the phenotypic effect of one mutation is conditional on its genetic background. These interactions are critical to enzyme fitness. This was shown in a study looking at the β -lactamase gene, which found five mutations that jointly contributed to a ~100,000-fold increase in resistance to the antibiotic cefotaxime when expressed in *E. coli*. A closer examination found not all mutations individually conferred an increase in cefotaxime resistance in *E. coli* and each was dependant on the genetic background (Weinreich *et al.*, 2006). Weinreich and colleagues had previously refereed to this phenomena as “sign epistasis”, where a mutation is beneficial on some genetic backgrounds and deleterious on others (Weinreich *et al.*, 2005).

Epistatic interactions are known to be widespread, however are frequently overlooked during conventional stepwise directed evolution approaches. When mutations are examined in isolation they may appear neutral or deleterious in one genetic background, however may provide profound improvements in enzyme fitness when introduced into a different genetic background. This means a stepwise mutagenesis approach could disregard entire evolutionary pathways, decelerating fitness gains.

4.1.1 Superior chloramphenicol detoxifying NfsA variants

In Chapter 3, a simultaneous site-directed mutagenesis library was created. From this library several NfsA variants were recovered that were able to protect *E. coli* 7NT cells to an almost 10-fold higher concentration of chloramphenicol, compared to wild-type NfsA. We reasoned that these enzyme variants could be used to model an enzyme exhibiting peak activity. Having identified an enzyme exhibiting peak activity, all possible intermediates of this enzyme variant could be created. These intermediates could then be used to recreate all possible evolutionary pathways when progressing from wild-type NfsA to the evolved

NfsA variant. Assessing the activities of each of these variants would enable us to determine whether a stepwise mutagenesis pathway could have yielded these NfsA variants, or if these enzymes could only be recovered using a simultaneous site-directed mutagenesis approach. By recreating all possible enzyme intermediates, we could also comprehensively assess epistatic interactions occurring between different substitutions and identify combinations which may be key to the evolutionary history of the enzyme.

Weinreich *et al.* (2006) took a similar approach in assessing the contributions of each of the five mutations in their β -lactamase variant. All 32 possible combinations of this enzyme variant were generated, creating 120 possible evolutionary trajectories. They found 102 of these trajectories were inaccessible to Darwinian selection, meaning they would be unlikely to be selected for in a stepwise directed evolution approach. We planned to take a similar approach using two of the NfsA variants identified in Chapter 3, however on a much larger scale, having targeted more residues in the active site of NfsA. We considered that this might reveal substantially more complex epistatic interactions than have previously been explored.

4.2 Aims

- Create all possible stepwise amino acid substitution intermediates of the top two chloramphenicol detoxifying NfsA variants and assess the ability of each enzyme intermediate to detoxify chloramphenicol.
- Recreate all evolutionary trajectories to determine whether a stepwise mutagenesis approach could have yielded the same NfsA variant.
- Identify key epistatic interactions that may be restricting evolution.
- Counter-screen select enzyme intermediates against a panel of prodrugs to assess how evolving activity towards chloramphenicol detoxification affects the other promiscuous activities of NfsA.

4.3 Results

4.3.1 Creation and characterisation of intermediates of the top two chloramphenicol detoxifying NfsA variants

	41	43	215	219	222	224	225	227
NfsA 36_37	Y	L	C	Y	V	R	V	G
NfsA 20_39	Y	L	N	Y	R	Y	D	H
WT NfsA	S	L	H	T	K	S	R	F

Table 4.1: Amino acids at the eight-targeted positions in the top two chloramphenicol detoxifying NfsA variants.

Chloramphenicol IC₅₀ values were used as a measure of enzyme fitness throughout this chapter, with a higher IC₅₀ values indicating improved enzyme fitness. The top two NfsA variants which protected *E. coli* 7NT cells against the highest chloramphenicol concentration (i.e., highest IC₅₀ value), NfsA 36_37 and 20_39, were selected for all experiments going forward (Table 3.8). Both of these variants had substitutions at seven of the eight-targeted positions in the Marsden library. Some similarities were seen, with both variants containing S41Y and T219Y amino acid substitutions and the wild-type leucine residue at position 43 (Table 4.1).

The seven amino acid substitutions meant that there were 128 intermediates for each of these variants (i.e., 2⁷). The genes encoding the 128 enzyme intermediates of both NfsA 36_37 and 20_39 (256 total) were ordered as arrayed synthetic genes from Twist Biosciences (*San Francisco, California*). The genes were cloned into the pUCX vector and used to transform the screening strain *E. coli* 7NT using a high-throughput restriction cloning and transformation protocol (Section 2.5.9.1). IC₅₀ assays with chloramphenicol were subsequently conducted for all 256 *E. coli* 7NT strains that resulted. These individual IC₅₀ values are reported in the appendices (Appendix B).

4.3.2 Recreating the evolutionary trajectories of NfsA 36_37 and 20_39

4.3.2.1 Most plausible evolutionary trajectories for NfsA 36_37 and 20_39

We wanted to determine whether NfsA 36_37 and 20_39 could have been evolved using a conventional stepwise mutagenesis approach or whether the simultaneous site-directed

mutagenesis approach provided the only plausible way to attain these evolved variants. To mimic a conventional stepwise directed evolution experiment, substitutions were introduced one at a time and the substitution which resulted in the greatest improvement in chloramphenicol detoxification was selected (i.e., highest IC₅₀ value). If at any point no substitution resulted in an increase in activity, then the substitution which resulted in the smallest decrease in activity was selected.

Figure 4.1: Evolutionary trajectory of NfsA 36_37 and 20_39. Each trajectory shows the amino acid substitution which resulted in the greatest improvement in chloramphenicol IC₅₀ value at each step. (A) Evolutionary trajectory of NfsA 36_37, (B) Evolutionary trajectory of NfsA 20_39, (C) Overlay of the two evolutionary trajectories, (D) Evaluation of alternative amino acid substitutions at the fifth step in the evolutionary trajectory for NfsA 36_37 (green) and NfsA 20_39 (blue). All data is presented as the fold activity of wild-type NfsA. Data represents the average of a minimum of four biological repeats $\pm$ SD.

In both examples the evolved NfsA variant could have been attained using a conventional stepwise evolutionary approach, as a linear increase in enzyme fitness was observed with the incremental addition of single amino acid substitutions (Figure 4.1). The last two substitutions introduced into NfsA 20_39 did not contribute to enzyme fitness, while the last three substitutions introduced into NfsA 36_37 only minimally contributed to enzyme

fitness. The order of the residues that were substituted to yield the greatest improvements in chloramphenicol reduction were the same for each enzyme for the first four steps (Figure 4.1C). This provided confidence that each step reflected genuine improvements in activity, despite the large error bars. To gain greater certainty that the fifth step up reflected a genuine improvement for each enzyme, each of the remaining possible substitutions was added in turn to the four-substitution variant of NfsA 36_37 (S41Y/T219Y/R225V/F227G) or NfsA 20_39 (S41Y/T219Y/R225D/F227H) and tested side-by-side in additional IC₅₀ assays. Figure 4.1D confirms that the S224R and H215N substitutions in NfsA 36_37 and 20_39 respectively, provided the largest increase in enzyme fitness. Despite the small fitness increase yielded from the introduction of S224R, this intermediate had a consistently higher IC₅₀ value than the intermediate at the previous step.

One compelling observation is the structural differences of the amino acids introduced at positions 225 and 227, despite each substitution conferring almost identical improvements in enzyme fitness. At position 225 in NfsA 36_37 a valine residue was introduced in place of arginine, while in NfsA 20_39 an aspartic acid residue was introduced. Valine and aspartic acid possess substantially different structures (Table 4.2, Figure 3.4). A similar phenomenon was observed at position 227 where in NfsA 36_37 a glycine residue was introduced in place of phenylalanine, while in NfsA 20_39 a histidine residue was introduced. Again, both these introduced amino acids have considerably different structures (Table 4.2, Figure 3.4). As all of these substitutions confer an almost equal improvement in enzyme fitness, it suggests that eliminating the wild-type residue at these two positions is key to improving the ability to detoxify chloramphenicol. That said, not all substitutions at these positions were equally beneficial as 84% of the top 114 chloramphenicol detoxifying variants had a polar or non-polar residue, suggesting a preference for these amino acids at position 225. More redundancy was evident at position 227 with a roughly equal portion of the top 114 chloramphenicol detoxifying variants having a non-polar, polar, acidic or basic amino acid substitution (Table 3.7).

	Wild-type NfsA	NfsA_Ec 36_37	NfsA_Ec 20_39
Position 225	 Arginine	 Valine	 Aspartic acid
Position 227	 Phenylalanine	 Glycine	 Histidine

Table 4.2: Structure of the amino acids present at positions 225 and 227 in NfsA 36_37, NfsA 20_39 and wild-type NfsA.

4.3.2.2 All “permissible” evolutionary trajectories of NfsA 36_37 and 20_39

We wanted to investigate how many evolutionary trajectories were accessible to a stepwise mutagenesis approach. In particular, we wanted to look at how many pathways may have been precluded due to negative epistatic interactions (i.e., combinations of mutations that are deleterious). Dr Mark Calcott, a postdoctoral fellow in the Ackerley lab wrote a Python script to generate the 5040 possible evolutionary trajectories for each of NfsA 36_37 and 20_39 ($7! = 5040$ combinations). This script then delineated how many evolutionary trajectories existed where the introduction of each amino acid substitution resulted in a minimum 16% improvement in enzyme fitness (16% being greater than the average error for each set of 128 variants, NfsA 36_37 set = 14.2% and NfsA 20_39 set = 15.6%). To account for the diminishing returns in enzyme fitness that are typically observed at the end of an evolutionary trajectory, the minimum improvement in enzyme fitness was decreased to 8% for the last three steps. This allowed substitutions like S224R, which had been independently validated to provide a fitness benefit in the trajectory of NfsA 36_37, to be included in the trajectory. The phenomenon of diminishing returns in enzyme fitness has been reported in the literature, with Tokuriki *et al.*, (2012) reporting decreasing fitness gains through the evolutionary progression of a phosphotriesterase. I used the output of this Python script to generate Figure 4.2 and Figure 4.3. This Python script is publicly available on GitHub (https://github.com/MarkCalcott/Analyse_epistatic_interactions/blob/master/data_analysis_v2.py).

Figure 4.2: Positive evolutionary trajectories of NfsA 36_37 and 20_39. Each node represents an enzyme intermediate with the substitutions present in the enzyme denoted by + or - symbols corresponding to the presence (+) or absence (-) of the substitutions S41Y, H215C/N, T219Y, K222V/R, S224R/Y, R225V/D and F227G/H respectively. Numbers in bold within each circle represent the chloramphenicol IC_{50} value in μM . Evolutionary pathways are shown for every scenario where a substitution yielded a minimum 16% increase in enzyme fitness for the first four substitutions and a minimum 8% increase in enzyme fitness for the last three substitutions. The pathway showing the greatest improvement in enzyme fitness at each step shown in Figure 4.1 is shown by thick black arrows. This pathway for NfsA 20_39 involves one decrease in enzyme fitness (albeit well within error) denoted by a dashed thick line.

Of the 5040 possible evolutionary trajectories, only three for NfsA 36_37 and four for NfsA 20_39 are accessible through continuous positive selection (as defined by a minimum 16% increase in enzyme fitness for the first four amino acid substitution, and a minimum 8% increase in enzyme fitness for the last three amino acid substitutions; Figure 4.2). Most of the evolutionary trajectories are selectively inaccessible, and would not yield either NfsA 36_37 or 20_39. This is because not all of the amino acid substitutions confer an improvement in enzyme fitness on all allelic backgrounds.

4.3.2.3 All permissible evolutionary trajectories of NfsA 36_37* and 20_39*

The final two amino acid substitutions in NfsA 36_37 and 20_39 were not accessible using the rules defined in Section 4.3.2.2. This skewed the evolutionary trajectory analysis, as

reaching each final evolved variant required amino acid substitutions that did not substantially contribute to enzyme fitness. A simpler evolutionary trajectory was generated for each enzyme, excluding the H215C and K222V substitutions from NfsA 36_37 (denoted NfsA 36_37* with the code YYRVG) and K222R and S224Y from NfsA 20_39 (denoted NfsA_Ec 20_39* with the code YNYDH). This greatly reduced the number of enzyme intermediates that existed between wild-type NfsA and NfsA 36_37* or 20_39* to 32 (2^5).

Figure 4.3: Positive evolutionary trajectories of the NfsA variants 36_37* and 20_39*. Each node represents an enzyme intermediate with the substitutions presented in the enzyme denoted by + or - symbols corresponding to the presence (+) or absence (-) of substitutions. Amino acid substitutions introduced into NfsA 36_37*: S41Y, T219Y, S224R, R225V and F227G. Amino acid substitutions introduced into NfsA 20_39*: S41Y, H215N, T219Y, R225D and F227H. Numbers represent the chloramphenicol IC_{50} value in μ M. Evolutionary pathways are shown for every scenario where a substitution yielded a minimum 16% increase in enzyme fitness for the first four amino acid substitutions and a minimum 8% increase in enzyme fitness for the last amino acid substitution. The pathway showing the greatest improvement in enzyme fitness at each step shown in Figure 4.1 is shown by thick black arrows.

Reducing the number of amino acid substitutions also decreased the number of possible evolutionary trajectories from 5040 to 120 (i.e., 5!) for NfsA 36_37 and 20_39. Of these possible evolutionary trajectories, three were identified for NfsA 36_37* and four for NfsA 20_39* that resulted in a minimum 16% increase in enzyme fitness with the addition of each of the first three substitutions and a minimum 8% increase in enzyme fitness for the

final substitution (Figure 4.3). Like in Figure 4.2, most of the evolutionary trajectories were selectively inaccessible.

4.3.3 Epistatic interactions between amino acids in NfsA 36_37 and 20_39

The majority of evolutionary trajectories in the evolution of NfsA 36_37 and 20_39 were deemed inaccessible. This is because the introduction of many substitutions resulted in neutral or even deleterious effects on enzyme fitness, depending on the other substitutions that had already been introduced into each enzyme. This provided evidence of epistatic interactions occurring. To investigate this further, the average effect each substitution had on overall enzyme fitness was calculated. A Python script was written by Dr Mark Calcott to determine the number of times each substitution had a positive, negative or negligible effect on enzyme fitness, as well as its mean proportional increase on enzyme fitness (Table 4.3). This Python Script is publicly available on GitHub (https://github.com/MarkCalcott/Analyse_epistatic_interactions/blob/master/data_epistatic_v2.py).

NfsA_Ec 36_37	Positive	Negative	Negligible	Mean Proportional Increase
S41Y	47	4	13	2.3
H215C	13	23	28	1.0
T219Y	35	3	26	1.5
K222V	24	6	34	1.3
S224R	18	29	17	1.0
R225V	59	2	3	5.0
F227G	40	6	18	1.7

NfsA_Ec 20_39	Positive	Negative	Negligible	Mean Proportional Increase
S41Y	41	2	21	3.1
H215N	8	34	22	0.8
T219Y	32	8	24	1.4
K222R	3	10	51	0.9
S224Y	4	24	36	0.9
R225D	53	8	3	4.2
F227H	34	10	20	1.5

Table 4.3: Summary of the effect each amino acid substitution in NfsA 36_37 or NfsA 20_39 had on overall enzyme fitness. The effect each substitution had on enzyme fitness was determined in the 64 (i.e., 2^6) different allelic backgrounds which exist for each substitution. An unpaired two-tailed t-test was used to determine statistical significance when comparing each substitution to the corresponding allelic background without the substitution. Positive = introduction of the substitution resulted in significantly improved enzyme fitness. Negative = introduction of the substitution resulted in significantly decreased enzyme fitness. Negligible = introduction of the substitution resulted in no significant difference in enzyme fitness. A ratio was generated for the 64 allelic backgrounds which exist for each substitution (IC_{50} variant with substitution/ IC_{50} variant without substitution), this was averaged to determine the mean proportional increase.

All seven substitutions conditionally increased enzyme fitness in at least three allelic backgrounds. Interestingly, all substitutions also had a deleterious effect on enzyme fitness in at least two allelic backgrounds. This suggests complex epistatic interactions are occurring between these substitutions, with activity highly dependent on the other

substitutions that have been introduced. Some key observations demonstrating these complex epistatic interactions are outlined in the following sections.

4.3.3.1 Introduction of individual amino acid substitutions into NfsA

Although all seven amino acid substitutions were selected simultaneously in the evolved variants, only one substitution from NfsA 36_37 or 20_39 (R225V/D) conferred an improvement in enzyme fitness when each was introduced individually (Figure 4.4). This is why all of the trajectories presented in Figure 4.1, Figure 4.2 and Figure 4.3 began with this substitution, as the remaining 86% of trajectories began with a substitution that did not contribute to enzyme fitness. Several of the substitutions (H215C, S224R and F227H) significantly decreased enzyme fitness when introduced individually (Figure 4.4). Although in isolation these substitutions are neutral or deleterious to enzyme fitness, in combination with other substitutions they can provide a positive effect on enzyme fitness as shown in Table 4.3. This illustrates the influence epistatic interactions can exert on enzyme evolution, as looking at these substitutions in isolation would disqualify them as worthy candidates.

Figure 4.4: The fitness effect of the single amino acid substitutions present in NfsA 36_37 and 20_39. Chloramphenicol IC₅₀ values of *E. coli* 7NT cells expressing pUCX:nfsA constructs encoding the amino acid substitutions S41Y, H215C/N, T219Y, K222V/R, S224R/Y, R225V/D or F227G/H. An unpaired t-test was used to determine whether introduction of a substitution was significantly different to wild-type NfsA (*** = p -value ≥ 0.001 , ** = p -value ≥ 0.01). Data presented is the average of a minimum of four biological repeats $\pm$ SD.

4.3.3.2 Epistatic interactions between S41Y, T219Y and R225V/D

S41Y provided the second highest mean proportional increase in enzyme fitness of the seven substitutions, providing a positive effect on enzyme fitness in 64-73% of allelic backgrounds (Table 4.3). In Chapter 3, a strong selection for the S41Y substitution was

observed in NfsA variants recovered from a stringent chloramphenicol selection, with 73% of the top 30 chloramphenicol reductase variants having this substitution (Table 3.8). This suggested S41Y should substantially improve enzyme fitness, yet on its own it did not confer any improvement in enzyme fitness. A similar scenario was seen with the T219Y substitution. T219Y was also highly selected for in the top 30 chloramphenicol detoxifying variants, with 37% of variants having this substitution. Although T219Y did not confer an improvement in enzyme fitness on its own, it provided a positive effect on enzyme fitness in 50-55% of allelic backgrounds. The epistatic interactions between these substitutions is summarised in Figure 4.5.

Figure 4.5: Epistatic interactions of S41Y, T219Y and R225V/D. All raw IC_{50} values were normalised to wild-type NfsA (e.g., IC_{50} value S41Y (5.1) – IC_{50} WT NfsA (4.8) = 0.3). The additive effect (plain bars) was calculated by adding the normalised IC_{50} values of individual amino acid substitutions (e.g., normalised IC_{50} value of S41Y (0.3) + normalised IC_{50} value of R225V/D (7.8) = 8.1). The additive error was derived using an error propagation equation ($error = \sqrt{(\delta X^2) + (\delta Y^2) + (\delta Z^2)}$ where δX , δY , δZ is the error of IC_{50} values X, Y and Z). The observed effect is the experimentally determined IC_{50} value normalised to wild-type NfsA (striped bars). The activity of enzyme variants containing R225V and R225D were averaged, as the IC_{50} values were comparable in all instances. An unpaired t-test was used to determine whether the additive effect was significantly different to the observed effect (***) = p -value ≥ 0.001 , ** = p -value ≥ 0.01 , * = p -value ≥ 0.05). The experimental data presented is the average of a minimum of four biological repeats $\pm$ SD.

Despite S41Y conferring no improvement in activity when introduced on its own, in combination with R225V/D the observed activity was much larger than the additive effect of these two substitutions (Figure 4.5, column A). S41Y and R225V/D also had a positive epistatic interaction with the T219Y substitution (Figure 4.5, column B). Interestingly, no positive effect was observed if S41Y or R225V/D were removed from this combination (Figure 4.5, columns C and D), showing any enzyme fitness benefit attributed to T219Y is conditional on the presence of both R225V/D and S41Y. This again illustrates why very

few evolutionary trajectories are selectively accessible, as the introduction of these substitutions must occur in a specific order to maximise fitness output.

4.3.3.3 F227G/H in different allelic backgrounds

The F227G/H substitution appeared to have a similar effect to S41Y or T219Y, providing a positive effect on enzyme fitness in 53-62% of allelic backgrounds (Table 4.3). However, unlike S41Y and T219Y, the F227G/H substitutions were mildly deleterious to enzyme fitness when introduced individually (Figure 4.4). The F227G/H substitutions could also be introduced earlier into the evolutionary progression of NfsA and still provide a fitness benefit (Figure 4.2 and Figure 4.3). The addition of the F227G/H substitution was the only selectively permissible alternative second step in the evolutionary trajectory to that presented in Figure 4.1.

4.3.3.4 H215C/N in different allelic backgrounds

Unlike many of the other substitutions, the H215C/N substitution had a predominantly negative effect on enzyme fitness, providing a positive effect in only a few allelic backgrounds (Table 4.3). The complexity of some of these epistatic interactions are outlined in Figure 4.6 and Figure 4.7.

Figure 4.6: Epistatic interactions of the H215C and H215N substitutions. All raw IC_{50} values were normalised to wild-type NfsA (e.g., IC_{50} value H215C (3.7) – IC_{50} WT NfsA (4.8) = -1.1). The additive effect (plain bars) was calculated by adding the normalised IC_{50} values of individual amino acid substitutions (e.g., normalised IC_{50} value of H215C (-1.1) + normalised IC_{50} value of R225V (8.2) = 7.1). The additive error was derived using an error propagation equation (error = $\sqrt{(\delta X^2) + (\delta Y^2) + (\delta Z^2)}$ where δX , δY , δZ is the error of IC_{50} values X, Y and Z). The observed effect is the experimentally determined IC_{50} value normalised to wild-type NfsA (striped bars). An unpaired t-test was used to determine whether the additive effect was significantly different to the observed effect (*** = p-value $\geq$ 0.001, ** = p-value $\geq$ 0.01, * = p-value $\geq$ 0.05). The experimental data presented is the average of a minimum of four biological repeats $\pm$ SD.

Introducing the H215C/N substitution in combination with R225V/D had a profoundly negative effect on enzyme fitness, generating an enzyme that had substantially less activity than wild-type NfsA (Figure 4.6). In contrast, the combination of the H215C and F227G substitutions resulted in a substantial improvement in enzyme fitness despite both these substitutions having a negative effect on enzyme fitness when introduced individually (Figure 4.4 and Figure 4.6).

In the evolutionary trajectory presented for NfsA 20_39, H215N was the last substitution to be introduced which contributed a positive effect on enzyme fitness (Figure 4.1). The effect of introducing H215N into the four prior enzyme intermediates is presented in Figure 4.7.

Figure 4.7: Evolutionary progression of NfsA 20_39 with and without the addition of H215N. The plain bars represent the enzyme intermediates identified in the evolutionary progression of NfsA 20_39 shown in Figure 4.1. The striped bars represent IC₅₀ values of the enzyme intermediates with the addition of H215N. These IC₅₀ values were experimentally determined as these enzyme intermediates were four of the 128 intermediates created of NfsA 20_39 (Appendix A). An unpaired *t*-test was conducted to determine whether the addition of H215N significantly changed activity from the enzyme intermediate without the H215N substitution (*** = *p*-value ≥ 0.001 , ** *p*-value ≥ 0.001 , * = *p*-value ≥ 0.05). Data presented is the average of a minimum of four biological repeats $\pm$ SD.

Introducing the H215N substitution into the first or second step in the evolutionary trajectory had a profoundly deleterious effect on enzyme fitness, disrupting the positive effect of R225D and S41Y (Figure 4.7). It was not until after the R225D, S41Y, T219Y and F227H substitutions had been introduced, that introducing H215N resulted in a significant improvement in enzyme fitness. This again illustrates the complexity of epistatic interactions between these substitutions and the importance of introducing these

substitutions in the correct combinations to maximise fitness gains and avoid deleterious interactivities.

4.3.4 Consequence of evolving NfsA variants towards chloramphenicol detoxification

In this study, the modelled evolutionary trajectories of NfsA 36_37 and 20_39 showed incremental improvements in chloramphenicol detoxifying activity with the addition of each amino acid substitution (Figure 4.1). We saw an opportunity to assess the effect that evolution towards chloramphenicol detoxifying activity had on the other promiscuous activities of NfsA. *E. coli* 7NT strains expressing the seven enzyme intermediates that were identified in the evolutionary trajectory presented in Figure 4.1, as well as wild-type *nfsA*, were tested in IC₅₀ growth assays to assess their ability to reduce a panel of nitroaromatic prodrugs (metronidazole, 2,4-DNT, CB1954, RB6145 and nitrofurazone).

Figure 4.8: Activity of enzyme intermediates of NfsA 36_37 and 20_39 with a range of nitroaromatic prodrugs. *E. coli* 7NT cells expressing each of the seven enzyme intermediates of *nfsA* 36_37 and 20_39 together with wild-type *nfsA* were tested in IC₅₀ assays with chloramphenicol, metronidazole, 2,4-DNT, CB1954, nitrofurazone and RB6145. Data is presented as the fold activity of wild-type NfsA, with an average of four biological repeats $\pm$ SD. Green = NfsA 36_37, blue = NfsA 20_39. Chloramphenicol and metronidazole activity are plotted on a different scale due to a larger fold-increase in activity.

A similar trend was observed with chloramphenicol and metronidazole, with an approximate 10-fold improvement in activity observed with both of these compounds (Figure 4.8). The introduction of the first substitution (R225V/D) resulted in the greatest improvement in metronidazole activity (> 5-fold). Slight improvements in 2,4-DNT and CB1954 reductase activity were evident with improving chloramphenicol detoxification. The ability of NfsA to reduce 2,4-DNT steadily increased with the sequential addition of substitutions, while the increase observed in CB1954 reductase activity was mainly attributed to the first amino acid substitution (R225V/D). No obvious changes in activity were observed with RB6145 or nitrofurazone. This shows that evolving NfsA to exhibit improved chloramphenicol detoxification can have varying effects on the other promiscuous activities of this enzyme.

4.4 Discussion

In this chapter we sought to recreate the evolutionary pathways of NfsA 36_37 and 20_39 to determine whether a stepwise mutagenesis pathway could have yielded the same enzyme variants, or whether a simultaneous mutagenesis approach was required. We also investigated epistatic interactions occurring between substitutions in these variants to gain a better understanding of the role epistasis can play during enzyme evolution.

Both NfsA 36_37 and 20_39 could potentially have been generated using a conventional stepwise evolutionary approach, with incremental improvements in enzyme fitness observed with successive amino acid substitutions (Figure 4.1). The improvement in enzyme fitness plateaued at the end of both evolutionary trajectories. Diminishing returns in enzyme fitness have been reported in the literature, with smaller benefits per mutation achieved as an enzyme nears its fitness optimum (Tokuriki *et al.*, 2012). NfsA 36_37 and 20_39 exhibited almost identical evolutionary trajectories in terms of fitness gains and the order of the positions mutated, despite differences in the amino acid substitutions (Table 4.1). This showed the order of substitutions was more important than the specific amino acid incorporated, with several positions exhibiting a substantial tolerance for an alternative amino acid to be introduced. For example, at position 225, 36% of the top 114 chloramphenicol detoxifying variants had an asparagine residue, however many of the other amino acids which could be encoded by the NDT codon at this position (e.g., histidine, aspartic acid, glycine, valine, leucine, isoleucine and cysteine) were also highly prevalent. This was also seen at position 227 with 28% of the top 114 chloramphenicol

detoxifying variants having an aspartic acid residue, however a histidine and glycine residue were the second and third most prevalent, despite obvious structural differences. Despite this redundancy, not all amino acids substitutions were tolerated at these two positions. At position 225, only two of the top 114 chloramphenicol detoxifying variants had a serine residue, indicating that not any polar residue is tolerated at this position. This was also seen at position 227 with only one of the top 114 chloramphenicol detoxifying variants having a valine residue, contrasting with the high prevalence of glycine residues at this position.

Although all amino acid substitutions were simultaneously selected for in NfsA 36_37 and 20_39, only one substitution (R225V/D) conferred an improvement in enzyme fitness when introduced individually (Figure 4.4). The other substitutions (S41Y, T219Y, H215C/N and F227G/H) exhibited sign epistasis, only conditionally improving enzyme fitness depending on which substitutions had been previously introduced. Some of these interactions could potentially have been predicted by analysing the active site of NfsA, and are consistent with previous observations reported in Chapter 3, Section 3.3.5.3. For example, the NfsA active site image presented in Figure 3.4 showed that steric hindrance likely prevents the S41Y substitution from co-existing with an unmutated arginine residue at position 225. Mutating the arginine residue at position 225 to any smaller residue opens up the active site, subsequently allowing the S41Y substitution. This model explains why in all evolutionary trajectories the R225V/D substitution must be introduced before the S41Y substitution to maximise the fitness benefit of this substitution. In contrast, other epistatic interactions such as those involving the H215C/N substitution would have been much harder to anticipate using structural data, as this position is located further away from the positions it has potential to exhibit epistasis with (Figure 3.1).

While a stepwise evolutionary trajectory was identified for both NfsA 36_37 and 20_39, the vast majority of evolutionary trajectories were selectively inaccessible using a conventional stepwise approach (Figure 4.2 and Figure 4.3). This was found to be a consequence of complex epistatic interactions occurring between the amino acid substitutions, which were key to the evolutionary history of these enzyme variants. In isolation, most of the substitutions introduced into NfsA 36_37 and 20_39 would not be selected for, as only R225V/D was found to be beneficial on its own. Rationally, we would never think to introduce many of these substitutions in combination with one another, e.g.,

using an approach such as iterative saturation mutagenesis, which involves identifying individually important residues which are then sequentially randomised (Reetz and Carballeira, 2007). Using this approach, only one of the seven sites targeted in the Marsden library (R225) would have been considered as an important residue. The effects of epistasis should also be considered when taking a simultaneous site-directed mutagenesis approach, particularly when selecting residues to target based off previous mutagenesis studies. Although a particular substitution may occur frequently in different enzyme variants, it does not mean a fitness benefit is derived from this substitution alone. This substitution may only be beneficial when selected in combination with another substitution, as was demonstrated at position 41. This may be a flaw within this study as previously discussed in Chapter 3, with the targeted positions being selected from two non-overlapping libraries (Section 3.4).

Finally, we wanted to investigate the consequences that evolving enzyme variants towards improved chloramphenicol detoxification had on the other promiscuous activities of NfsA. The substitutions which resulted in improved chloramphenicol detoxification were found to also consistently improve metronidazole reductase activity, with lesser improvements evident in 2,4-DNT and CB1954 reductase activities. In contrast, no improvement was seen with the RB6145 and nitrofurazone reducing activities. This shows that evolving NfsA towards chloramphenicol detoxification had a variety of off-target effects, in one case improving activity with an alternate substrate by 10-fold. The diverse structures of chloramphenicol and metronidazole (Figure 1.4 and Figure 1.7) makes it hard to predict why improving the ability to reduce chloramphenicol would also substantially improve metronidazole reduction. Some of the substitutions introduced (e.g., R225V/D) are predicted to be involved in enlargement of the active site pocket which may accommodate large substrates like chloramphenicol. This enlargement may also allow for the relatively small metronidazole substrate to easily enter the active site and subsequently become reduced. This does not explain why the ability to reduce other small substrates like 2,4-DNT (Figure 1.6) are not also substantially improved. This data led us to believe that NfsA could be used as a model to understand the “black-box” effect often encountered in directed evolution. This is the phenomenon where evolving an enzyme towards one activity can have off-target and possibly detrimental effects on other metabolic processes with a cell. This concept was investigated further in Chapter 5, using the promiscuous nature of NfsA to model off-target effects that can occur under varying selection pressures.

Chapter 5 Using NfsA to investigate the off-target effects of directed evolution

---

5.1 Introduction

Enzymes are traditionally considered to be specific catalysts, with the notion that one enzyme catalyses one reaction. This definition dominates high school and university textbooks where we are taught about the high specificity and accuracy of enzymes. Despite this well entrenched notion, it is estimated that 37% of enzymes in *E. coli* can act on multiple substrates (Nam *et al.*, 2012). Promiscuous enzymes can have profound mechanistic and evolutionary implications on enzyme evolution. While promiscuous enzymes can provide a crucial starting point in enzyme evolution (Jensen, 1976; Toscano *et al.*, 2007), they can be a hindrance due to their off-target effects (Pitera *et al.*, 2007; Bantscheff *et al.*, 2009; Kim and Copley, 2012). Despite this, scientists infrequently address this additional layer of complexity when evolving enzymes.

In Chapter 4 we found substitutions improving chloramphenicol detoxification also improved metronidazole reduction, while the ability to reduce other substrates such as nitrofurazone and RB6145 was unaffected by the drive towards an alternative substrate. This experiment brought to light the “black-box” theory of directed evolution. This is the phenomenon where scientists can blindly evolve an enzyme towards a desired function without examining how this evolution might be constrained by unknown limitations within the “black-box” e.g., by interfering with other metabolic pathways within a cell. Promiscuous enzymes can potentially interact with a vast number of substrates within a cell. As discussed in Chapter 1, a study from Huang *et al.*, (2015) identified a haloacid dehydrogenase which was found to interact with 143 of 169 substrates tested. This stresses how complicated and interconnected the different metabolic pathways of a cell are. There could be thousands of potential interactions between evolved enzymes and existing metabolic networks.

5.1.1 Examples of metabolic interference

Although examples are scarce, some researchers have reported the effect that introducing a new enzyme or pathway can have on both the evolved enzyme/pathway and existing reactions within a cell. Kim and Copley, (2012) looked at the off-target effects of introducing a novel pathway for the synthesis of pyridoxal 5'-phosphate (PLP) into the existing *E. coli* metabolic network. They identified several unanticipated reactions between metabolites and enzymes from the novel pathway and the existing metabolic network. Problems included; (1) interference by metabolites in the novel PLP synthesis pathway with metabolic processes in the existing *E. coli* network, (2) interference by metabolites in the existing *E. coli* network with the function of the novel PLP synthesis pathway and (3) diversion of metabolites from the novel PLP synthesis pathway by promiscuous enzymes in the existing *E. coli* network. A similar effect was also reported by Pitera *et al.*, (2007) who introduced a mevalonate pathway into *E. coli* to improve isoprenoid production. It was later discovered that the introduction of this pathway led to the accumulation of intermediate metabolites, limiting flux through the pathway. Another example was shown by Hansen *et al.*, (2009) who engineered yeast to produce vanillin from glucose. Three genes were introduced into *Saccharomyces cerevisiae* to establish this *de novo* pathway. However, initial experiments found the resulting vanillin was converted to a toxic alcohol form by a promiscuous alcohol dehydrogenase. This was resolved by modifying the existing metabolic network to prevent the build-up of the toxic compound. A comprehensive understanding of the conflicting interaction that was occurring was required to do this, however it can be difficult to anticipate *a priori* what these inhibitory interactions will be.

5.1.2 NfsA as a model for metabolic interference

We sought to use NfsA as a model to improve our understanding of the off-target effects of directed evolution. The experiments conducted in Chapter 4, were on two individual enzyme variants (NfsA 36_37 and 20_39), looking at how the activity profile of these enzyme variants changed through the evolutionary progression towards chloramphenicol detoxification. We wanted to build on this, and investigate at a population level how the promiscuous activities of a population of NfsA variants evolves in response to different selection conditions. We used a panel of nitroaromatic compounds to represent other metabolites that could exist in a metabolic network, to examine any prohibitive off-target interactions during the directed evolution of NfsA. Our aim was to gain an understanding

of how enzymes can be evolved towards a user-defined goal and allow for the successful integration of these enzymes into pre-existing metabolic networks.

5.2 Aims

- Investigate the effect a positive selection has on the promiscuous activities of a population of NfsA enzyme variants by counter-screening with a panel of prodrugs (metronidazole, RB6145, CB1954, 2,4-DNT and nitrofurazone).
- Investigate the effect a dual positive-negative selection has on the promiscuous activities of a population of NfsA enzyme variants by counter-screening with a panel of prodrugs (metronidazole, RB6145, CB1954, 2,4-DNT and nitrofurazone).

5.3 Results

5.3.1 Effects of subjecting the Marsden library to positive selection pressures

The previously created Marsden library was subjected to three different positive selection conditions using various combinations and concentrations of chloramphenicol and niclosamide to create both stringent and weak selections. The first selection condition chosen was 45 μ M chloramphenicol, mimicking the selection conditions used in Chapter 3 (Section 3.3.4.2). The second selection condition used 50 μ M niclosamide (the maximum concentration soluble in agar) and 40 μ M chloramphenicol, which together allowed for a similar percentage of the Marsden library to grow. The third selection condition used 1 μ M niclosamide. A stringent niclosamide only selection could not be obtained, therefore niclosamide was used at a low concentration allowing us to investigate a weak positive selection. These positive selection conditions and the resulting percentage selection are presented in Table 5.1.

Selection Conditions	% Selection
45 μ M chloramphenicol	0.000036%
50 μ M niclosamide + 40 μ M chloramphenicol	0.000054%
1 μ M niclosamide	2.4%

Table 5.1: Three positive selection conditions and percentage of the Marsden library able to grow under these conditions. A sample of the Marsden library was resuspended in LB medium and serial dilutions were created as per Section 2.6.3. The concentrated cell resuspension was plated on the selection conditions above and a sample of diluted culture was plated on unselective media (LB agar supplemented with Amp). The number of colonies which grew on the unselective media were counted and multiplied by the dilution factor to estimate the total number of colonies plated on the selective media. This was then used to calculate the % selection.

5.3.1.1 Growth inhibition assays of *E. coli* 7NT cells expressing *nfsA* variants recovered from positive selection conditions

The Marsden library was plated on the positive selection conditions described in Table 5.1. Fifty-seven colonies which grew under each selection condition were picked into the inner 60 wells of 96-well plates, alongside one pUCX_empty and two wild-type *nfsA* controls. Replicate plates of strains overexpressing *nfsA* variants were challenged with a panel of nitroaromatic compounds (metronidazole, RB6145, CB1954, nitrofurazone, 2,4-DNT, chloramphenicol and niclosamide). Concentrations were chosen for each prodrug which resulted in ~50% growth inhibition for *E. coli* 7NT cells expressing wild-type *nfsA*. This meant growth inhibition greater than 50% would indicate a more active variant and growth inhibition less than 50% a less active variant. For niclosamide, a percentage growth of ~75% was obtained. Increasing the niclosamide concentration to greater than 4 μ M did not decrease the percentage growth, therefore 4 μ M niclosamide was used as we believed this was the highest concentration which could give us reliable data. The 57 *nfsA* variants were sequenced to determine the amino acid substitutions at the eight-targeted positions. Only 54 *NfsA* variants were included from the 1 μ M niclosamide selection as two variants were found to be double transformants (*E. coli* bearing two plasmids) and one contained pUCX_empty. The data obtained from the growth inhibition assays were translated into dot plot graphs (Figure 5.1). *E. coli* 7NT cells expressing *nfsA* variants picked from unselective media (LB agar supplemented with Amp) were also included as controls.

Figure 5.1: Growth inhibition assay data of *E. coli* 7NT strains expressing *nfsA* variants selected on positive selection conditions. The Marsden library was plated on either 45 μ M chloramphenicol (1st column), 50 μ M niclosamide + 40 μ M chloramphenicol (2nd column), 1 μ M niclosamide (3rd column) or unselective media (LB agar supplemented with Amp, 4th column). Replica plates of *E. coli* 7NT colonies expressing 57 unique *nfsA* variants were selected under each condition (except 1 μ M niclosamide, where only 54 unique variants were recovered) and counter-screened in growth inhibition assays with 75 μ M metronidazole (red), 18 μ M RB6145 (orange), 250 μ M CB1954 (yellow), 27.5 μ M nitrofurazone (green), 375 μ M 2,4-DNT (blue), 4 μ M chloramphenicol (purple) or 4 μ M niclosamide (pink) and compared to *E. coli* 7NT strains expressing wild-type *nfsA* (dark grey, 5th column) and pUCX_empty (light grey, 6th column) controls. The % growth inhibition is reported for prodrugs (metronidazole, RB6145, CB1954, 2,4-DNT and nitrofurazone) and the % growth is reported for niclosamide and chloramphenicol. Data are the average of four biological replicates with error bars omitted for clarity (standard error of the mean less than 5). The median growth inhibition is shown as a black horizontal line on each data set.

The NfsA variants recovered from the 45 μ M chloramphenicol selection exhibited an improved reduction of all nitroaromatic compounds tested, except RB6145, compared to wild-type NfsA. The greatest improvement was evident with metronidazole, where all 57 NfsA variants exhibited almost 100% growth inhibition at the metronidazole concentration tested (red plot, column 1). This was consistent with the results seen in Chapter 4, where both NfsA 36_37 and 20_39 exhibited a large increase in metronidazole reductase activity, in addition to improved chloramphenicol detoxification (Figure 4.8). Also consistent with the results observed in Chapter 4, the chloramphenicol selection recovered NfsA variants which exhibited an improved reduction of both CB1954 (yellow plot, column 1) and 2,4-DNT (blue plot, column 1), but not RB6145 (orange plot, column 1). One stark difference between the NfsA variants identified in the chloramphenicol selection and NfsA variants 36_37 and 20_39, was the overall improved reduction of nitrofurazone. Every NfsA variant selected on chloramphenicol exhibited improved reduction of nitrofurazone compared to wild-type NfsA (green plot, column 1), whereas NfsA 36_37 and 20_39 did not exhibit improved nitrofurazone reduction, despite being selected on the same concentration of chloramphenicol (Figure 4.8). This may be because IC_{50} assays were used when testing the activities of NfsA 36_37 and 20_39, whereas growth inhibition assays were used for the population of NfsA variants. Growth inhibition assays are measured at a fixed concentration of nitrofurazone, therefore an increase on a growth inhibition scale may only translate to a slight improvement in IC_{50} value (e.g., 80% growth inhibition at 45 μ M nitrofurazone may equate to an IC_{50} value of 40 μ M, a slight improvement in activity).

Almost identical observations were evident in the population selected on the second positive selection condition, 50 μ M niclosamide and 40 μ M chloramphenicol. However, NfsA variants recovered from the combined niclosamide/chloramphenicol selection exhibited a significantly higher mean percentage growth inhibition with RB6145 (p-value <0.001). This suggests that the ability to reduce RB6145 is linked to the detoxification of niclosamide, but not chloramphenicol. Consistent with this assumption, NfsA variants recovered under the weak niclosamide-only selection were further improved in their mean RB6145 reducing activity (orange plot, column 3). Nitrofurazone was the only compound for which the dual niclosamide/chloramphenicol selection gave the highest median activity (i.e., greater than in the chloramphenicol or niclosamide only selections). All other compounds appeared biased towards either chloramphenicol or niclosamide, suggesting nitrofurazone reduction is improved by generic nitroreductase improvement. In general, the

NfsA variants recovered under the weak niclosamide-only selection exhibited improved activity with all prodrugs tested. However, the level of improvement was not as great as for the stringent positive selection conditions, with the exception of RB6145, which is clearly favoured by niclosamide selection relative to chloramphenicol selection.

5.3.1.2 Amino acid analysis of NfsA variants selected under positive selection conditions

		S41	L43	H215	T219	K222	S224	R225	F227
45 μ M chloramphenicol	Non-polar	0	70	21	23	60	40	21	19
	Polar	2	30	16	2	18	14	67	18
	Acidic	0	0	37	0	11	4	7	21
	Basic	9	0	19	2	12	37	5	33
	Aromatic	89	0	7	74	0	5	0	9
		S41	L43	H215	T219	K222	S224	R225	F227
50 μ M niclosamide + 40 μ M chloramphenicol	Non-polar	2	51	32	21	54	32	18	30
	Polar	5	40	26	2	18	14	70	21
	Acidic	0	0	14	0	9	12	7	14
	Basic	7	0	16	2	19	32	5	23
	Aromatic	86	9	12	75	0	11	0	12
		S41	L43	H215	T219	K222	S224	R225	F227
1 μ M niclosamide	Non-polar	19	43	43	33	44	31	50	24
	Polar	9	28	28	26	35	30	24	46
	Acidic	0	0	6	2	2	7	6	2
	Basic	28	4	17	7	19	17	7	17
	Aromatic	44	26	7	31	0	15	13	11

Table 5.2: Percentage occurrence of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library for NfsA variants recovered from the positive selection conditions. *E. coli* 7NT strains expressing 57 unique nfsA variants were selected on each of the following conditions; 45 μ M chloramphenicol, 50 μ M niclosamide + 40 μ M chloramphenicol and 1 μ M niclosamide. These nfsA variants were sequenced to determine the amino acids present at the eight-targeted positions. The property of the wild-type amino acid at each position is boxed in red.

A summary of the amino acid preferences at each residue position for the 57 unique NfsA variants selected under the three positive selection conditions is shown in Table 5.2. Clear amino acid preferences were evident in the NfsA variants recovered from the stringent positive selections (45 μ M chloramphenicol and 50 μ M niclosamide + 40 μ M chloramphenicol). These amino acid preferences were almost identical with those seen in

Chapter 3, in the 114 NfsA variants also recovered from a stringent chloramphenicol selection (Table 3.7). Some amino acid differences were evident between the NfsA variants recovered in the weak niclosamide selection compared to both stringent positive selections. The most obvious difference was a much lower prevalence of aromatic residues at positions 41 and 219. This likely contributed to the activity differences in NfsA variants recovered from the weak niclosamide selection. The sequence and growth inhibition assay data for each individual NfsA variant is shown in Appendix C.

5.3.2 Effects of subjecting the Marsden library to a dual positive-negative selection pressure

To expand on the results obtained in Section 5.3.1, we wanted to investigate the effects a dual positive-negative selection would have on the promiscuous activities of NfsA. In this selection condition, a positive selection compound such as niclosamide or chloramphenicol was included in the media alongside a prodrug (Figure 5.2). As prodrugs are cytotoxic after nitro-reduction, they can act like a suicide switch, directly selecting against any cell expressing a *nfsA* variant that is able to reduce the prodrug. In contrast, niclosamide and chloramphenicol are toxic prior to nitro-reduction, therefore including one of these compounds in the media ensures the ability to detoxify these compounds is retained.

Figure 5.2: Schematic of a positive selection and dual positive-negative selection.

This system was used to model a scenario where a prodrug represented a metabolite in an essential metabolic pathway in a cell. In this scenario, if an evolved enzyme gained an off-target activity with this essential metabolite then it could interfere with an essential metabolic pathway, and have detrimental consequences for the host. Thus, these enzymes might never be recovered from a directed evolution experiment, no matter how active they

might become with the primary screening substrate. The dual positive-negative selection was used to recover NfsA variants exhibiting retained or improved activity with the desired substrate (chloramphenicol or niclosamide), while excluding activity with the essential metabolite (a nitroaromatic prodrug). We believed this selection could be used to identify more specialised enzymes, allowing for fewer adventitious interactions upon introducing the enzyme into an existing metabolic network.

Metronidazole and RB6145 were selected as the two prodrugs to be included alongside niclosamide and/or chloramphenicol in the dual positive-negative selection. These two prodrugs were chosen as NfsA variants recovered from the stringent positive selection conditions exhibited differing activities with these two compounds (Figure 5.1, red and orange plots). Almost 100% of the NfsA enzyme variants identified from the stringent positive selection conditions were improved with metronidazole, compared to wild-type NfsA, while only around 50% of the NfsA enzyme variants recovered were improved with RB6145 compared to wild-type NfsA.

The previously created Marsden library was used as a source of NfsA variants to subject to a dual positive-negative selection. We wanted to recover NfsA variants which were able to protect *E. coli* 7NT cells to a similar or higher chloramphenicol or niclosamide concentration than wild-type NfsA, while also excluding either metronidazole or RB6145. In order to achieve this, we needed to use concentrations of chloramphenicol and niclosamide that were higher than *E. coli* 7NT cells expressing wild-type *nfsA* were able to grow on ($>1\ \mu\text{M}$ chloramphenicol or $>0.4\ \mu\text{M}$ niclosamide). The Marsden library was first plated over varying concentrations of chloramphenicol (2-10 μM) in combination with metronidazole (100-200 μM). Despite plating the Marsden library on low concentrations of both compounds, all selection conditions recovered colonies which were small and sickly and were unable to be revived when picked into unselective liquid media. During a chloramphenicol-only selection at 1 μM , *E. coli* 7NT cells expressing wild-type *nfsA* take two days to appear, therefore it was concluded the addition of a prodrug added too much stress to the already slow growing colonies. We therefore turned to the use of niclosamide. The Marsden library was plated on media containing 1 μM niclosamide in combination with either 500 μM metronidazole or 40 μM RB6145. A low concentration of niclosamide was used, as increasing the concentration recovered unhealthy NfsA variants, which when picked into liquid media were unable to protect *E. coli* 7NT cells against even low

concentrations of niclosamide. Concentrations of metronidazole and RB6145 were selected that resulted in a similar selection stringency to one another (Table 5.3).

Condition	% Selection
45 μ M chloramphenicol	0.000036%
50 μ M niclosamide + 40 μ M chloramphenicol	0.000054%
1 μ M niclosamide	2.4%
1 μ M niclosamide + 500 μ M metronidazole	0.0012%
1 μ M niclosamide + 40 μ M RB6145	0.0013%

Table 5.3: The five positive and dual positive-negative selections used with the resulting percentage selection. A sample of the Marsden library was resuspended in LB medium and serial dilutions were created as per Section 2.6.3. The concentrated cell resuspension was plated on the selection conditions above and a sample of diluted culture was plated on unselective media (LB agar supplemented with Amp). The number of colonies which grew on the unselective media were counted and multiplied by the dilution factor to estimate the total number of colonies plated on the selective media. This was then used to calculate the % selection.

5.3.2.1 Growth inhibition assays of *E. coli* 7NT cells expressing *nfsA* variants recovered from dual positive-negative conditions

The Marsden library was plated on each of the dual positive-negative selection conditions described in Table 5.3. Fifty-seven colonies which grew on each selection condition were picked into the inner 60 wells of 96-well plates alongside one pUCX_empty and two wild-type NfsA controls. Replicate plates of strains overexpressing *nfsA* variants were challenged with niclosamide (the positive selection compound), metronidazole and RB6145 (the negative selection prodrug, simulating an essential metabolite), and also CB1954, nitrofurazone, 2,4-DNT and chloramphenicol to assess whether activity with these other nitroaromatic compounds had been collaterally affected by the dual positive-negative selection (simulating other non-essential metabolites within a metabolic network). As previously (Section 5.3.1.1), concentrations of each compound were chosen which resulted in ~50% growth inhibition for *E. coli* 7NT cells expressing wild-type *nfsA* (with the exception of niclosamide). The 57 *nfsA* variants were sequenced prior to growth inhibition assays to ensure 57 unique *nfsA* variants had been identified. The data obtained from the growth inhibition assays were translated into dot plot graphs (Figure 5.3). *E. coli* 7NT cells expressing *nfsA* variants picked from unselective media (LB agar supplemented with Amp) and 1 μ M niclosamide shown in Figure 5.1 were also included for comparison.

Figure 5.3: Growth inhibition assay data of *E. coli* 7NT expressing *nfsA* variants recovered from dual positive-negative selection conditions. The Marsden library was plated on either 1 μ M niclosamide + 500 μ M metronidazole (1st column) or 1 μ M niclosamide + 40 μ M RB6145 (2nd column). *NfsA* variants recovered from the 1 μ M niclosamide selection (3rd column) and unselective media (4th column) from Figure 5.1 are included for comparison. Replica plates of *E. coli* 7NT expressing 57 unique *nfsA* variants selected under each condition (except 1 μ M niclosamide, where only 54 unique variants were recovered) were counter-screened in growth inhibition assays with 75 μ M metronidazole (red), 18 μ M RB6145 (orange), 250 μ M CB1954 (yellow), 27.5 μ M nitrofurazone (green), 375 μ M 2,4-DNT (blue), 4 μ M chloramphenicol or 4 μ M niclosamide and compared to *E. coli* 7NT expressing wild-type *NfsA* (dark grey, 5th column) and pUCX_empty (light grey, 6th column) controls. The % growth inhibition is reported for prodrugs (metronidazole, RB6145, CB1954, 2,4-DNT and nitrofurazone) and the % growth is reported for niclosamide and chloramphenicol. Data are the average of four biological replicates with error bars omitted for clarity (standard error of the mean less than 5). The median growth inhibition is shown as a black horizontal line on each data set.

In general, inclusion of metronidazole or RB6145 in the media had similar effects on overall nitroreductase activity (Figure 5.3). Unsurprisingly, including metronidazole in the media recovered NfsA variants which had a decreased capacity to reduce metronidazole (red plot, column 1), compared to wild-type NfsA. More surprisingly, this also affected the ability of the NfsA variants to reduce RB6145 (orange plot, column 1), albeit not as strongly as was the case for metronidazole. The loss in RB6145 activity was contrary to our expectations based on Figure 5.1, where the recovered NfsA variants exhibited opposing activities with metronidazole and RB6145. The effect was even stronger in the selection containing RB6145, where NfsA variants were recovered that had an equally decreased ability to reduce both metronidazole (red plot, column 2) and RB6145 (orange plot, column 2).

In terms of collateral activities, both dual positive-negative selections also yielded NfsA variants that had a decreased ability to reduce CB1954 (yellow plot, columns 1 and 2), 2,4-DNT (blue plot, columns 1 and 2) and chloramphenicol (purple, columns 1 and 2) compared to NfsA variants recovered from the 1 μ M niclosamide selection in the absence of prodrugs (column 3, yellow, blue and purple plot). In the case of CB1954 and 2,4-DNT the levels of activity obtained were comparable to wild-type NfsA, while ability to detoxify chloramphenicol was substantially lower than wild-type NfsA. The presence of metronidazole or RB6145 did not affect the ability of NfsA variants to reduce nitrofurazone. NfsA variants isolated from both dual positive-negative selection conditions (green plot, columns 1 and 2) were able to reduce nitrofurazone at a comparable level to NfsA variants selected in the absence of the prodrug (1 μ M niclosamide, green plot column 3).

A summary of the effect this evolutionary constraint had on the evolution of niclosamide detoxification is outlined in Table 5.4. In brief, inclusion of metronidazole or RB6145 in the selection media had a negative impact on the evolution of niclosamide detoxification. The median percentage growth dropped from 90% in NfsA variants recovered from the 1 μ M niclosamide selection to 63% and 68% in NfsA variants recovered from the dual positive-negative selections (niclosamide/metronidazole and niclosamide/RB6145, respectively). This shows that the evolutionary constraints imposed diminished the ability to recover NfsA variants that could protect *E. coli* 7NT cells against niclosamide, with the median activity falling below the starting level of activity (Table 5.4).

Niclosamide Challenge	Median % Growth	Activity Ratio to WT NfsA
1 μ M niclosamide + 500 μ M metronidazole	63	0.80
1 μ M niclosamide + 40 μ M RB6145	68	0.87
1 μ M niclosamide	90	1.15
Wild-type NfsA	78	1.00

Table 5.4: Median percentage growth of NfsA variants recovered from various selection conditions challenged with niclosamide. The median percentage growth is presented for the 57 nfsA variants recovered from each selection condition, challenged with 4 μ M niclosamide under the various selection conditions tested in Figure 5.3. The ratio of activity to *E. coli* 7NT cells expressing wild-type nfsA is also reported.

5.3.2.2 Amino acid analysis of NfsA variants selected under dual positive-negative conditions

		S41	L43	H215	T219	K222	S224	R225	F227
1 μ M niclosamide + 500 μ M metronidazole	Non-polar	2	18	30	19	30	30	21	21
	Polar	14	51	23	72	25	16	26	30
	Acidic	0	0	2	2	0	2	4	9
	Basic	0	2	33	4	46	39	33	28
	Aromatic	84	30	12	4	0	14	16	12
		S41	L43	H215	T219	K222	S224	R225	F227
1 μ M niclosamide + 40 μ M RB6145	Non-polar	0	18	19	19	39	18	35	51
	Polar	2	60	28	47	16	32	26	12
	Acidic	0	2	5	9	5	2	5	14
	Basic	0	4	39	9	39	32	5	9
	Aromatic	98	18	9	16	2	18	28	14

Table 5.5: Percentage occurrence of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of NfsA variants selected on dual positive-negative selection conditions. *E. coli* 7NT expressing 57 unique nfsA variants were selected on each of the following conditions; 1 μ M niclosamide + 500 μ M metronidazole or 1 μ M niclosamide and 40 μ M RB6145. These nfsA variants were sequenced to determine the amino acids present at the eight-targeted positions. The property of the wild-type amino acid at each position is boxed in red.

The NfsA variants recovered from each of the dual positive-negative selections displayed very similar amino acid preferences, likely explaining the similarities observed in the growth inhibition assays (Table 5.5). Many differences were evident between NfsA variants recovered from the positive selections versus NfsA variants recovered from the dual

positive-negative selections. The most prominent difference was a preference for a polar residue at position 219 in NfsA variants selected on media containing metronidazole or RB6145, compared to an aromatic residue in NfsA variants recovered from the positive selection conditions. Specifically, a proline residue was the most common polar residue at this position, with 67% of NfsA variants recovered from the niclosamide/metronidazole selection and 42% of NfsA variants recovered from the niclosamide/RB6145 selection containing this residue (Appendix C). This was a stark contrast to the three positive selection conditions, as only one of the other *nfsA* variants sequenced (168 in total) encoded a proline residue at position 219 (Appendix C). Another notable sequence difference was a decreased preference for polar residues at position 225 compared to NfsA variants recovered from the positive selection conditions (Table 5.5). Specifically, 30% of NfsA variants recovered from the niclosamide/metronidazole selection had the wild-type residue arginine at position 225. This was a stark contrast from the four other selection conditions, as only two other *nfsA* variants sequenced (225 total) encoded an arginine residue at position 225 (Appendix C).

5.3.3 Summary of the five selection conditions

A summary graph was created to enable a side-by-side comparison of the 57 NfsA variants selected under each of the six different selection conditions with their corresponding activity (except the 1 μ M niclosamide selection, where only 54 unique variants were recovered). The six selection conditions correspond to; (1) 45 μ M chloramphenicol, (2) 50 μ M niclosamide and 40 μ M chloramphenicol, (3) 1 μ M niclosamide, (4) 1 μ M niclosamide and 500 μ M metronidazole, (5) 1 μ M niclosamide and 40 μ M RB6145 and (6) unselective media containing no nitroaromatic compound.

Figure 5.4: Summary graph of the 57 NfsA variants recovered under each of the six selection conditions. The selection conditions are listed 1-6 and correspond to; (1) 45 μ M chloramphenicol, (2) 50 μ M niclosamide and 40 μ M chloramphenicol, (3) 1 μ M niclosamide, (4) 1 μ M niclosamide and 500 μ M metronidazole, (5) 1 μ M niclosamide and 40 μ M RB6145 and (6) unselective media containing no nitroaromatic compounds. Each dot represents a NfsA variant counter-screened in a growth inhibition assay with either; 75 μ M metronidazole (red), 18 μ M RB6145 (orange), 250 μ M CB1954 (yellow), 27.5 μ M nitrofurazone (green), 375 μ M 2,4-DNT (blue), 4 μ M chloramphenicol or 4 μ M niclosamide. Wild-type NfsA is included as a control in dark grey in each column. All data is reported as the % growth inhibition with the exception of chloramphenicol and niclosamide which are represented as % growth. Data represent the average of four biological repeats $\pm$ SD with error omitted for clarity (standard error of the mean less than 5).

The exclusively positive selection conditions including chloramphenicol and/or niclosamide recovered NfsA variants that had an improved ability to reduce chloramphenicol and niclosamide as well as all nitroaromatic prodrugs tested, with the exception of RB6145 whenever chloramphenicol was used (Figure 5.4). This suggests that chloramphenicol is antagonistic toward RB6145 activity. Inclusion of metronidazole or RB6145 in the selective media had a negative impact on the reduction of metronidazole and RB6145 plus all other nitroaromatic compounds tested with the exception of nitrofurazone. NfsA variants recovered from all five selection conditions exhibited a roughly equal capacity to reduce nitrofurazone (Figure 5.4). Consistent with results previously described in Chapter 4 (Figure 4.8), NfsA variants exhibited comparable activity profiles with metronidazole and chloramphenicol, in all selection conditions tested. This suggests that similar residues are involved in the reduction of these two compounds, and these two activities are closely linked.

5.3.4 Kinetic analysis of the promiscuous activities of His₆-tagged NfsA

We wanted to investigate why some promiscuous activities of NfsA appeared more “evolvable” (e.g., the especially large increase in metronidazole reduction in NfsA variants recovered from the positive selection conditions). To assess this, kinetic assays were conducted testing the *in vitro* activity of purified His₆-tagged NfsA with the panel of prodrugs tested above (metronidazole, RB6145, CB1954, nitrofurazone and 2,4-DNT). We wanted to determine if there was a correlation between the kinetic parameters of wild-type NfsA with a particular prodrug and how frequently NfsA variants were recovered exhibiting improved *in vivo* activity with the same prodrug. The single electron reduction potential (E^1_7) was also identified for each prodrug. The single electron reduction potential quantifies the ease with which a compound can accept a single electron. The higher and more positive the value, the better the compound is at competing with molecular oxygen for electrons, while a compound with a lower value will be less sensitive to the presence of available electrons (Wiebe, 1999). It has been shown that the reactivity of NfsA with some nitroaromatic compounds exhibits a positive correlation with the E^1_7 value of the compound (Valiauga *et al.*, 2017). However, of the prodrugs tested in this study, only CB1954 was tested by Valiauga *et al.*, (2017).

Michaelis-Menten steady state kinetic parameters measuring CB1954 reduction were determined by measuring the appearance of the 2- and 4-hydroxylamine at 420 nm (Section

2.8.6.2). The reduction of nitrofurazone, RB6145 and 2,4-DNT cannot be measured directly (no substantial change in absorbance in the visible spectrum), therefore steady state kinetic parameters were determined indirectly, by measuring the decrease of NADPH at 340 nm (Section 2.8.6.2). Kinetics were measured at a fixed concentration of NADPH (250 μ M) and used to determine the apparent k_{cat} and K_M at that concentration.

As previously discussed in Chapter 3 (Section 3.3.5.4), determining steady state kinetic parameters of NfsA and other family members with metronidazole is notoriously difficult. This is due to the intrinsic NADPH oxidase activity of NfsA and slow turnover rate of metronidazole, making it difficult to determine accurate kinetic parameters when monitoring the depletion of NADPH (Olekhovich *et al.*, 2009; Valiauga *et al.*, 2018). The reduction of metronidazole can be monitored directly at 340 nm, however this is the same wavelength that NADPH is monitored at. This meant the direct measurement of metronidazole reduction was not possible using the standard assay. A NADPH regeneration system was developed to determine steady state kinetic parameters of metronidazole (Section 2.8.6.3). This system used glucose dehydrogenase from *Bacillus subtilis* to convert D-glucose to D-glucono-1,5-lactone, in the process regenerating reduced NADPH from NADP⁺. A form of this system had been previously used by members of the Ackerley lab allowing for the regeneration of NADH in a cell-free system for remediation of hexavalent chromium (Robins *et al.*, 2013). The extinction coefficient of metronidazole at 340 nm was experimentally determined to be 1,980 M⁻¹ cm⁻¹ using the protocol outlined in Section 2.8.6.1.

I first validated this system by determining steady state kinetic parameters for CB1954 with and without the regeneration system, to confirm the system was not affecting the rate of prodrug reduction. CB1954 was chosen for this as the reduced form can be monitored at 420 nm, independent of NADPH reduction. As no substantial change in k_{cat}/K_M was observed (Table 5.6), the system was deemed suitable for determining kinetic parameters with metronidazole. This system has since been used in another published study from the Ackerley group (Rich *et al.*, 2018).

	k_{catapp} (s ⁻¹)	K_{Mapp} (μM)	k_{cat}/K_M (M ⁻¹ .s ⁻¹)
CB1954 (without regeneration)	5.5 ± 0.3	160 ± 27	34,000 ± 6,000
CB1954 (with regeneration)	9.6 ± 0.4	180 ± 22	53,000 ± 7,000

Table 5.6: Kinetic parameters for the reduction of CB1954 by His₆-tagged NfsA with and without the NADPH regeneration system. The initial rate of reduction across 0-800 μM CB1954 was monitored at 420 nm either at a fixed concentration of 250 μM NADPH or using the NADPH regeneration system described in Section 2.8.6.3. Data is the average of three technical repeats. Data measured November 2016.

Following optimisation of a method to quantify metronidazole activity, Michaelis-Menten steady state kinetic parameters for the reduction of CB1954, nitrofurazone, 2,4-DNT, RB6145 and metronidazole were derived for His₆-tagged NfsA (Table 5.7). Also shown is the single electron reduction potential (E^1_7) for each prodrug. A linear regression was also plotted to assess whether there was a relationship between the E^1_7 value and the catalytic efficiency of these compounds in a NfsA catalysed reaction (Figure 5.5).

		E^1_7 (mV)	k_{catapp} (s ⁻¹)	K_{Mapp} (μM)	k_{cat}/K_M (M ⁻¹ .s ⁻¹)
1	Nitrofurazone	-257 ^a	10 ± 0.4	15 ± 2	670,000 ± 110,000
2	CB1954	-385 ^b	7.7 ± 0.2	140 ± 13	55,000 ± 5,000
3	2,4-DNT	-397 ^c	4.0 ± 0.2	58 ± 10	69,000 ± 12,000
4	RB6145	-398 ^d	2.8 ± 0.2	410 ± 60	6,800 ± 1000
5	Metronidazole	-486 ^a	0.5 ± 0.05	540 ± 100	930 ± 200

Table 5.7: Kinetic parameters for the reduction of CB1954, nitrofurazone, 2,4-DNT, RB6145 and metronidazole by His₆-tagged NfsA. The initial rate of reduction across 0-800 μM CB1954 and 250 μM NADPH was monitored at 420 nm. The initial rate of reduction across 0-200 μM nitrofurazone, 0-400 μM 2,4-DNT and 0-1200 μM RB6145 at 250 μM NADPH was monitored at 340 nm. The initial rate of reduction across 0-800 μM metronidazole was determined using the NADPH regeneration system and monitoring of metronidazole depletion at 340 nm. Data is the average of three technical replicates. E^1_7 values obtained from the following references, ^a = Smith and Edwards, (1995), ^b = Stratford et al., (1981), ^c = Riefler and Smets, (2000) and ^d = Fielden et al., (1992). Data measured August 2018. Michaelis-Menten plots shown in Appendix D.

Figure 5.5: Linear regression of the $\log k_{\text{cat}}/K_M$ of nitroaromatic compounds in a NfsA catalysed reaction and their single-electron reduction potential (E^I_7). The numbering of the nitroaromatic compounds, catalytic efficiency (k_{cat}/K_M), and single-electron reduction potential (E^I_7) corresponds to the data presented in Table 5.7. Linear regression analysis calculated using Prism 8.0 software.

The reduction of nitrofurazone by NfsA exhibited the highest catalytic efficiency, at least 10-fold higher than the reduction of all other substrates tested (Table 5.7). This activity is likely promoted by the substantially higher E^I_7 value for nitrofurazone, indicating this prodrug is more receptive to electrons, contributing to the high substrate turnover (k_{cat}) and overall higher catalytic efficiency (k_{cat}/K_M). This likely explains why in all five selection conditions tested, NfsA variants were able to reduce nitrofurazone better than wild-type NfsA, as reducing nitrofurazone is much easier to achieve than the other prodrugs. CB1954 and 2,4-DNT exhibited almost identical activity patterns in the five different selection conditions tested (Figure 5.4). The reduction of both of these compounds by NfsA also exhibited very similar kinetic parameters and E^I_7 values (Table 5.6). The reduction of RB6145 and metronidazole by NfsA exhibited the lowest catalytic efficiency, almost 300-fold lower than the reduction of nitrofurazone (Table 5.7). For metronidazole, this low catalytic efficiency is likely explained by the low E^I_7 value, indicating metronidazole is less able to accept electrons compared to nitrofurazone. The data presented in Table 5.7 and Figure 5.5 provides evidence that E^I_7 values are an important factor in NfsA reactivity, supporting the results published by Valiauga *et al.*, (2017), however, they are not the only factor. The ability to accommodate a substrate in the active site is also likely to be very important to allow for efficient substrate reduction. This may explain why NfsA is less active with RB6145 than either CB1954 or 2,4-DNT, despite all three compounds sharing similar E^I_7 values. It may also explain why fewer NfsA variants exhibiting improved reduction of RB6145 were recovered from the positive selection conditions (Figure 5.1),

despite a higher E^l_7 value and k_{cat}/K_M to metronidazole, however structural analyses would be required to back this up.

5.4 Discussion

The research described in this chapter involved the recovery of NfsA variants from different selection conditions. *E. coli* 7NT strains expressing these *nfsA* variants were tested in growth inhibition assays, to assess the ability of each variant to reduce the compound they were selected for or against, plus additional nitroaromatic compounds which were not involved in the selection. Although not commonly found in nature, we wanted to use these nitroaromatic compounds to represent other metabolites that could exist in a network. We then used this model system to examine the impact of having prohibitive off-target interactions during the directed evolution of NfsA.

It was previously known in the Ackerley group that inclusion of chloramphenicol or niclosamide in solid media could identify nitroreductase variants exhibiting an improved ability to reduce other nitroaromatic substrates. In previous studies, this has allowed for the identification of nitroreductase variants which are able to reduce nitroaromatic compounds of biotechnological importance (Little, 2015; Rich, 2017; Sharrock, 2018). However, in the current study we used this knowledge to show how activity with structurally distinct compounds can be collaterally affected during directed evolution. We showed how a selection involving different combinations of chloramphenicol and/or niclosamide, recovered NfsA variants which not only exhibited an improved reduction of chloramphenicol and niclosamide, but also improved reduction of a panel of structurally distinct nitroaromatic compounds (metronidazole, RB6145, CB1954, nitrofurazone and 2,4-DNT; Figure 5.1). However, the NfsA variants did not exhibit an equal improvement with all compounds. This was shown by the substantial difference between the reduction of metronidazole and RB6145 (Figure 5.4). Kinetic parameters of NfsA determined *in vitro* with metronidazole and RB6145 did not provide insight as to why more NfsA variants were recovered exhibiting improved reduction of metronidazole compared to RB6145 (Table 5.7). This suggested structural data may be important in predicting off-target interactions.

A dual positive-negative selection was also conducted using combinations of niclosamide with metronidazole or RB6145. We used this to simulate a scenario where a prodrug

represented a metabolite within an essential metabolic pathway in a cell. The dual positive-negative selection was used to recover NfsA variants exhibiting improved or retained activity with niclosamide, while excluding activity with the essential metabolite (as modelled using metronidazole or RB6145). These evolutionary constraints came at a cost for niclosamide detoxification, as NfsA variants were recovered with a decreased capacity to reduce niclosamide relative to NfsA variants recovered from the positive selection (Table 5.4). However, activity with both metronidazole and RB6145 was substantially decreased regardless of which compound was included in the selection. In particular, the inclusion of metronidazole in the selection impacted the reduction of both metronidazole and RB6145 equally. This would have been hard to predict in advance by looking at *in vitro* kinetic data, E^l_7 values or even the structure of these two distinct compounds. This suggested that substitutions crucial for the reduction of metronidazole may also be important for RB6145 reduction. This shows that while the correlation between catalytic efficiency and the E^l_7 value of a substrate is important, it should not be used as the only predictor of activity. In this example, structural information with metronidazole or RB6145 bound in the active site of could have provided insights into how the ability of NfsA to reduce both these substrates is closely linked, despite the differences highlighted above. In contrast, the ability to reduce nitrofurazone was not affected by the presence of either metronidazole or RB6145. In this example, the high activity observed did correlate with the E^l_7 value of nitrofurazone.

As previously stated, it is estimated that 37% of enzymes in *E. coli* act on multiple substrates (Nam *et al.*, 2012). These promiscuous and adventitious interactions have implications when introducing evolved enzymes into existing metabolic networks due to their potential to disrupt essential metabolic pathways. This phenomenon of metabolic interference is of ever-increasing importance given the current rise in synthetic biology and the need for bacteria to efficiently express these recombinant proteins. These results display the unpredictability and high frequency of adventitious interactions between an evolved enzyme (NfsA) and existing metabolites in our model network. Nonetheless, using a dual positive-negative selection we were able to recover NfsA variants exhibiting a specialised activity profile, exhibiting a decreased affinity for an essential metabolite (metronidazole or RB6145), albeit at a slight cost to the desired activity (niclosamide). The results presented in this chapter may be of relevance for synthetic biology studies, providing insight into how we can evolve enzymes with decreasing off-targets effects to improve their utility in an existing metabolic network.

Chapter 6 Identifying improved NfsA variants as tools for targeted cell ablation

---

6.1 Introduction

Zebrafish are an important genetic model for providing insights into cell regeneration, due to their ability to regenerate tissues such as the heart, fins, nervous system and eyes (Gemberling *et al.*, 2013). Targeted cell ablation is a tool that scientists can use to gain an understanding of the mechanisms of cell regeneration. Nitroreductase mediated cell ablation is one of the leading methods for achieving targeted cell ablation (White and Mumm, 2013). This method involves the expression of a nitroreductase under a tissue specific promoter. An otherwise innocuous prodrug can be administered which is converted to a cytotoxic form by the nitroreductase, allowing for targeted ablation of cells expressing the nitroreductase. This allows temporal and spatial control of cell ablation. The expression of the nitroreductase can also be linked to a fluorescence reporter transgene, allowing cell ablation and regeneration to be monitored. Nitroreductase mediated cell ablation was first achieved in NIH3TC cells, whereby administering the prodrug CB1954, resulted in ablation of cells expressing a bacterial nitroreductase (Bridgewater *et al.*, 1995). This system was then adapted for use in zebrafish using the nitroreductase NfsB_Ec and the 5-nitroimidazole prodrug metronidazole (Curado *et al.*, 2007). Metronidazole is the most commonly used prodrug in nitroreductase mediated cell ablation due to the wide commercial availability of this compound and the cell-entrapped nature of the reduced product.

6.1.1 NfsB_Ec mediated cell ablation

NfsB_Ec and metronidazole was the first nitroreductase/prodrug pairing used for targeted cell ablation in zebrafish. This system was first reported in studies by Curado *et al.*, (2007) and Pisharath *et al.*, (2007) where zebrafish larvae expressing *nfsB_Ec* were administered metronidazole, resulting in ablation of pancreatic β -cells expressing the nitroreductase. Since these studies, several transgenic zebrafish lines have been generated targeting a broad range of cell types, enabling successful selective ablation in cells such as cardiomyocytes (Curado *et al.*, 2007), hepatocytes (Curado *et al.*, 2007), pineal photoreceptors (Fernandes *et al.*, 2012; Li *et al.*, 2012) podocytes (Zhou and Hildebrandt, 2012), skin cells (Chen *et*

al., 2011), gonad cells (Hu *et al.*, 2010) and retinal cells (Zhao *et al.*, 2009; Ariga *et al.*, 2010).

The concentration of metronidazole used in treatments and the metronidazole exposure time varies depending on the targeted cell type. This variation is due to a number of factors including nitroreductase expression levels, cell metabolic rate and accessibility of metronidazole to the cell (Mathias *et al.*, 2014). The current standard treatment for nitroreductase mediated cell ablation is 10 mM metronidazole for 24 hours (White and Mumm, 2013). This concentration is very close to the general toxicity of metronidazole, as zebrafish lethality is observed after 24 h at 15-20 mM metronidazole (Curado *et al.*, 2007). Some quiescent cell types such as the lens, notochord and macrophages are difficult to ablate, with macrophages requiring 15 mM metronidazole over a 15-hour period, approaching the inherent toxicity of metronidazole (Gray *et al.*, 2011). Zebrafish lethality has even been observed at concentrations as low as 5 mM metronidazole following exposure times > 48 hours (Mathias *et al.*, 2014).

6.1.2 NfsB_Ec triple mutant mediated cell ablation

A NfsB_Ec triple mutant (T41Q/N71S/F124T) was identified as being nearly 50-fold more catalytically efficient with CB1954 (Guise *et al.*, 2007; Jaberipour *et al.*, 2010). NfsB_Ec T41Q/N71S/F124T was subsequently found to also be effective at reducing metronidazole and transgenic zebrafish lines were generated expressing this nitroreductase (Chen *et al.*, 2011; Mathias *et al.*, 2014). However, the activity of NfsB_Ec T41Q/N71S/F124T in zebrafish was not as great as had been observed in *E. coli*, enabling only a ~2-3-fold lower concentration of metronidazole to be used (*personal communication, Assoc. Prof. Jeff Mumm*). This variant also did not solve the problem of achieving cell ablation with quiescent cells such as macrophages (*personal communication, Assoc. Prof. Jeff Mumm*).

6.1.3 NfsB_Vv F70A/F108Y

Identifying nitroreductase variants that exhibit improved activation of metronidazole has been an ongoing goal in the Ackerley group. Dr Abigail Sharrock, a previous PhD student in the Ackerley lab group, engineered a promising candidate by rationally introducing two substitutions into a NfsB homologue from *Vibrio vulnificus* (NfsB_Vv) (F70A and F108Y). The equivalent substitutions had previously been found to improve the activity of NfsB_Ec with SN33623, a 5-nitroimidazole PET probe, as well as enhancing activity with

metronidazole (Williams, 2013). *E. coli* overexpressing *nfsB_Vv F70A/F108Y* exhibited a 40-fold lower IC₅₀ value than *E. coli* overexpressing *nfsB_Ec*, and a 12-fold lower IC₅₀ value than *E. coli* expressing *nfsB_Ec T41Q/N71S/F124T*. This improved activity was also evident in human cell lines. Human HEK-293 cells expressing *nfsB_Vv F70A/F108Y* exhibited a > 500-fold lower IC₅₀ value than HEK-293 cells expressing *nfsB_Ec* or *nfsB_Ec T41Q/F71S/F124T* (Sharrock, 2018). Based on this, it was considered that NfsB_Vv F70A/F180Y might have utility as an alternative nitroreductase for prodrug induced cell ablation, however prior to the research described in this chapter its activity in zebrafish had yet to be assessed.

6.1.4 Evolving NfsA variants for use in targeted cell ablation

In both *E. coli* and HEK-293 cells, wild-type NfsA is significantly better at reducing metronidazole than either wild-type NfsB_Ec or NfsB_Ec T412Q/N71S/F124T (Rich, 2017; Sharrock, 2018). NfsA is a highly mutable enzyme, therefore is a logical starting point for a directed evolution campaign to further improve this activity. NfsA is also broadly active with a range of 5- and 2-nitroimidazole nil-bystander prodrugs such tinidazole, misonidazole and RB6145, which might also have applications as alternative prodrugs in a cell ablation system. It is possible that alternative prodrugs might be able to ablate more stubborn cell types such as the lens, notochord and macrophages that the standard NfsB_Ec/metronidazole pairing struggles to ablate. With this in mind, new NfsA variants were sourced from the Marsden library as well as a previously generated 2-nitroimidazole combination library that had been designed to improve activity with 2-nitroimidazole substrates.

6.1.4.1 2-nitroimidazole combination library

A 2-nitroimidazole combination library was designed and created by a previous PhD student Dr Elsie Williams (Williams, 2013). Ten amino acid substitutions were selected which had been previously identified as contributing to improved reduction of 2-nitroimidazole prodrugs (R12S, S40T, S41C, S41Y, F42I, C45L, C45M, L103V, N134A, H215R). A degenerate *nfsA* gene library was designed to encode every possible combination of these ten amino acid substitutions or the corresponding wild-type residues, yielding a library of 576 different NfsA variants.

6.2 Aims

- Characterise *NfsB_Vv F70A/F108Y* in zebrafish
- Recover and characterise *NfsA* variants exhibiting improved activation of the 5-nitroimidazole prodrugs, metronidazole and tinidazole
- Recover and characterise *NfsA* variants exhibiting improved activation of the 2-nitroimidazole prodrugs, RB6145 and misonidazole.
- Develop a transient cell ablation assay in zebrafish to enable rapid assessment of nitroreductase mediated ablation.

6.3 Results

6.3.1 Creation of stable zebrafish lines

The work presented in Sections 6.3.1, 6.3.2, 6.3.3 and 6.3.8 was carried out during a 10-week research trip to the Wilmer Eye Institute, Johns Hopkins University, Maryland, USA. This work was supervised by Associate Professor Jeff Mumm and was completed in conjunction with then PhD student Abigail Sharrock. In advance of our arrival at the Wilmer Eye Institute, *nfsB_Vv F70A/F108Y* had been cloned into the zebrafish expression vector pTOL2 and sent to our collaborators. Using the protocol described in Section 2.10.4, Dr Meera Saxena and Danielle Meir-Levi, generated transgenic zebrafish lines expressing *nfsB_Vv F70A/F108Y*. Using the same protocol, I later assisted in the creation of putative founders (P₀) expressing the *nfsA* mutants identified and described in Sections 6.3.4 and 6.3.6.

Transgenic zebrafish expressing *nfsB F70A/F108Y* (and later, *nfsA* mutants) were generated via *Tol2* transposon mediated transgenesis. The *Tol2* element is a naturally occurring element originally found in the genome of the medaka fish (Koga *et al.*, 1996). It was later demonstrated that vectors containing minimal *cis*-sequences of the *Tol2* element were capable of mediating transposition in zebrafish embryos (Urasaki *et al.*, 2006). A vector (pTOL2) was designed by members of the Mumm lab containing these *Tol2* recognition sequences flanking the targeted DNA to be integrated into the zebrafish genome. The corresponding pTOL2 vector and *Tol2* transposase mRNA were co-injected into fertilized zebrafish eggs at the one-cell stage. In order to achieve stable transgenesis, the construct must be integrated into cells which give rise to germ cells to ensure the transgene construct

is passed on to the next generation. For this reason, fertilized eggs were injected as early as possible to ensure little to no differentiation had occurred prior to injection.

Expression of *nfsB_Vv F70A/F108Y* (and later, *nfsA* mutants) in zebrafish was achieved using the yeast Gal4/UAS expression system (Figure 6.1). This system involves the use of a Gal4 transcriptional activator (KalTA4) which is placed downstream of a cell-type specific promoter (Halpern *et al.*, 2008). In this study, neuron restrictive silencer elements (NRSE) and a *cfos* minimal promoter were used to robustly suppress nitroreductase expression in non-neuronal cells without compromising desired expression in the neuronal cells (Xie *et al.*, 2012). To activate nitroreductase expression, the Gal4 transcriptional activator binds to the UAS element (upstream activating sequence), initiating transcription of the downstream nitroreductase. This system was developed *in trans*. The genes encoding the Gal4 transcription activator (KalTA4) and tissue specific promoter (NRES and *cfos* promoter) were located on one construct and expressed in a “driver line”, and the genes encoding the UAS binding elements and nitroreductase were located on another construct and expressed in a “reporter” line (Figure 6.1). The use of this system *in trans* allows for multiple driver and reporter line combinations to be generated, as a reporter line can be crossed with a different driver line, allowing for expression in a different cell type.

Figure 6.1: Schematic of the Gal4/UAS trans expression system. A driver line expresses the Gal4 transcription activator under the control of a tissue specific promoter. A second reporter line is generated containing the 5 × UAS element upstream of the TagYFP and nitroreductase (NTR). When the driver and reporter lines are crossed the Gal4 is able to bind to the 5 × UAS element, allowing for expression of downstream tagYFP:NTR genes.

In this study, the pTOL2 constructs containing the genes encoding *nfsB_Vv F70A/F108Y* or *nfsA* mutants were injected directly into a driver line (gmc617). This was done in preliminary experiments to gauge how well *nfsB_Vv F70A/F108Y* or *nfsA* variants

expressed in the neuronal cells of the zebrafish and whether these cells could be ablated following administering a prodrug.

6.3.2 Characterisation of NfsB_Vv F70A/F108Y in zebrafish.

6.3.2.1 Assessing metronidazole induced ablation mediated by NfsB_Vv F70A/F108Y

Stable zebrafish lines expressing *nfsB_Vv F70A/F108Y* were characterised by then PhD student Abigail Sharrock and me to assess the ability of NfsB_Vv F70A/F108Y to reduce metronidazole resulting in targeted cell ablation. Our primary aim was to determine whether the metronidazole reducing activity of NfsB_Vv F70A/F180Y seen in *E. coli* and HEK-293 cells would translate into zebrafish, identifying a new “gold-standard” nitroreductase. NfsB_Vv F70A/F108Y would also become the benchmark for assessing new nitroreductase variants in this chapter.

Due to the initially limited number of fish stably expressing *nfsB_Vv F70A/F108Y*, these lines were out-crossed with wild-type zebrafish lines. Larvae exhibiting TagYFP fluorescence indicative of *nfsB_Vv F70A/F108Y* expression were used in cell ablation assays at 4 dpf (Section 2.10.6). Both the prodrug concentration and prodrug challenge time were varied in different experiments. This ensured that the immune system of the zebrafish had time to clear any fluorescent debris following cell ablation (*personal communication, Assoc. Prof. Jeff Mumm*). In some cases, cell ablation was also visualised by confocal microscopy.

The ablation capability of NfsB_Vv F70A/F108Y was first tested at 5 mM metronidazole, half the concentration which is currently used for ablation studies with the “gold-standard” NfsB_Ec enzyme (White and Mumm, 2013). This was tested over 48-, 24- and 6-hour challenge times (Figure 6.2).

Figure 6.2: Metronidazole induced ablation in transgenic zebrafish larvae expressing *nfsB_Vv F70A/F108Y* across various challenge times. Zebrafish larvae 4 dpf expressing *nfsB_Vv F70A/F108Y* were exposed to 5 mM metronidazole (green) or the equivalent volume of DMSO (blue). Larvae were challenged with metronidazole for either (A) 48-hours, (B) 24- or 6-hours and incubated at 28.5 °C for a total of 48 hours. Following prodrug challenge, TagYFP fluorescence was measured using TECAN Infinite M1000 plate reader. TagYFP fluorescence in the DMSO (control) and 5 mM metronidazole (treatment) groups were compared to determine whether significant cell ablation had occurred. A Mann-Whitney *t*-test was used to determine whether there was a statistically significant difference between the control and treatment group (*** = *p*-value ≤ 0.001).

Near complete cell ablation was observed at all of the above prodrug challenge times (48-, 4- and 6-hours), compared to the DMSO control group (Figure 6.2). This showed that *NfsB_Vv F70A/F108Y* could achieve near complete cell ablation at prodrug challenge times substantially less than what was required for wild-type *NfsB_Ec* (White and Mumm, 2013). Different clutches of zebrafish larvae exhibited varying levels of TagYFP fluorescence, therefore zebrafish treated with 5 mM metronidazole for 48 hours were plotted on a graph with a differing Y-axis (Figure 6.2A). This natural variation in TagYFP fluorescence between different clutches of larvae likely explains the variation seen in the 24-hour treatment group (Figure 6.2B).

We also sought to determine the lowest concentration of metronidazole needed to induce complete cell ablation at both a 24- and 6-hour drug exposure. The lowest concentration found to induce near complete cell ablation with a 24-hour challenge period was 0.1 mM metronidazole (Figure 6.3A). This concentration of metronidazole (0.1 mM) is 100-fold less than the concentration of metronidazole used with the “gold-standard” *NfsB_Ec* ablation system, and at half the challenge time. The lowest concentration found to induce near complete cell ablation following a 6-hour challenge was 0.25 mM metronidazole (Figure 6.3C). This was also improved over the *NfsB_Ec* triple mutant (T41Q/N71S/F124T) which exhibited cell ablation with only a 4-hour challenge time but

required 10 mM metronidazole, a substantially higher concentration than what is achieved here (Mathias *et al.*, 2014).

Figure 6.3: Metronidazole induced cell ablation in transgenic zebrafish larvae expressing *nfsB_Vv F70A/F108Y* at decreased metronidazole concentrations challenged for a 24- or 6-hour period. Zebrafish larvae 4 dpf expressing *nfsB_Vv F70A/F108Y* were exposed to metronidazole (green) or the same volume of DMSO (blue). (A) Larvae exposed to metronidazole for 24 hours were treated with 1 mM or 0.1 mM metronidazole. Larvae exposed to metronidazole for 6-hours were treated with either (B) 0.5 mM, (C) 0.25 mM or 0.1 mM metronidazole. Following prodrug challenge, the TagYFP fluorescence was measured using TECAN Infinite M1000 plate reader. TagYFP in the DMSO (control, blue) and metronidazole (treatment, green) groups were compared to determine whether significant cell ablation had occurred. A Mann-Whitney t-test was used to determine whether there was a statistically significant difference between the control and treatment group (***) = p -value ≤ 0.001 .

Confocal imaging was used to visualise the extent and localisation of cell ablation (Figure 6.4). Zebrafish larvae expressing *nfsB_Vv F70A/F108Y* were treated with 0.1 or 0.25 mM metronidazole for a 24-hour challenge time as detailed above. Following prodrug challenge, larvae were prepared for confocal imaging as described in Section 2.10.7.

Figure 6.4: Confocal imaging of transgenic zebrafish. Zebrafish larvae 6 dpf expressing *nfsB_Vv F70A/F108Y* were treated for 24-hours with DMSO-only, 0.1 or 0.25 mM metronidazole. A non-transgenic control was included which was not treated with metronidazole or DMSO. Images show an overlay of the transmitted light and TagYFP fluorescence z-stack images. The images in the right-hand panel are a detailed close-up of the eye of the zebrafish larvae to show TagYFP remnants.

The DMSO challenged larva exhibited broad TagYFP fluorescence, reporting on *nfsB_Vv F70A/F108Y* expression. This fluorescence was localised to spinal motor neurons and neuronal subsets in the eye and brain. Treatment with 0.25 mM or 0.1 mM metronidazole resulted in a substantial loss of TagYFP fluorescence, indicative of widespread cell ablation in the spinal motor neurons and neuronal subsets in the eye and brain. Closer examination of the head of the zebrafish larvae revealed some TagYFP remnants in the 0.1 mM metronidazole treated larva. This may indicate neuronal cells which have not been ablated by metronidazole, as has been previously observed (White *et al.*, 2017).

6.3.2.2 RB6145 and misonidazole induced cell ablation mediated by NfsB_Vv F70A/F108Y

Figure 6.5: Chemical structures of the prodrugs RB6145 and misonidazole.

E. coli expressing *nfsB_Vv F70A/F108Y* have been shown to activate the 2-nitroimidazole nitroaromatic compound EF5, albeit not as strongly as NfsA (Sharrock, 2018). Based on the promising results obtained from metronidazole induced cell ablation experiments, this transgenic zebrafish line was also tested with the 2-nitroimidazole prodrugs RB6145 and misonidazole (Figure 6.5). It was considered that these two prodrugs may prove to be effective in ablating more stubborn cell types and in research described later in this chapter, I sought to test alternative prodrug/NTR pairings for ablating different cell types. The successful creation of a NfsB_Vv F70A/F108Y zebrafish line provided an early opportunity to define a benchmark of the level of cell ablation which could be obtained by these two prodrugs (Figure 6.6).

Figure 6.6: RB6145 and misonidazole induced ablation in transgenic zebrafish larvae expressing *nfsB_Vv F70A/F108Y*. Zebrafish larvae 4 dpf expressing *nfsB_Vv F70A/F108Y* were exposed to 0.5 mM RB6145 (red), 0.5 mM misonidazole (orange) or an equivalent volume of DMSO (blue) for a 24-hour period. Following incubation, the TagYFP fluorescence was measured using TECAN Infinite M1000 plate reader. The control and treatment groups were compared to see if significant cell ablation had occurred. A Mann-Whitney t-test was used to determine whether there was a statistically significant difference between the control and treatment group (***) = p -value ≤ 0.001 .

Zebrafish larvae expressing *nfsB_Vv F70A/F108Y* were challenged with 0.5 mM RB6145 or 0.5 mM misonidazole for a 24-hour period. As depicted in Figure 6.6, significant cell ablation of target cells was observed with both RB6145 (red) and misonidazole (orange) challenged groups compared to the control group. Further testing at lower concentrations of RB6145 and misonidazole, and the potential for reduced challenge times is ongoing.

6.3.3 RB6145 toxicity in non-transgenic zebrafish larvae

RB6145 is inherently more potent than the other prodrugs used for cell ablation in this chapter. In wild-type HEK-293 cells this compound induces toxicity at 20-fold lower concentrations than misonidazole and 200-fold lower concentrations than metronidazole (Table 6.3 and Table 6.6). It was therefore important to determine how toxic RB6145 was to non-transgenic zebrafish, to establish whether targeted cell ablation was achievable at non-lethal doses. Toxicity screens were conducted with non-transgenic zebrafish larvae to determine the maximum dosage that could be administered to transgenic fish without encountering off-target toxicity or lethality (Section 2.10.5). Non-transgenic zebrafish larvae 3 dpf were challenged with varying concentrations of RB6145 (0-2.5 mM) for either 6-, 24- or 48- hours. Following prodrug challenge, the media containing prodrug was replaced with E3 medium and zebrafish larvae were monitored until 5 days following the initial challenge point to assess longer-term toxicity. At each time point the concentration of RB6145 that displayed no lethality to the non-transgenic fish was identified and selected as the maximum concentration which could be used in future ablation assays (Figure 6.7). The following concentrations of RB6145 were deemed to be the highest concentrations that could be used for each duration without inducing cell death; 6-hours = 1.5 mM, 24-hours = 0.25 mM and 48-hours = 0.3 mM. This showed that the cell ablation induced with 0.5 mM RB6145 (Figure 6.6), may have resulted in some off-target toxicity. Further testing is ongoing at a lower concentration of RB6145.

Figure 6.7: RB6145 toxicity of non-transgenic zebrafish larvae. Non-transgenic larvae were challenged with varying concentrations of RB6145 (0-2.5 mM RB6145) for either 6-, 24- or 48-hours. Following prodrug challenge, the larvae were moved into E3 medium and monitored until 5 days post the initial prodrug challenge to monitor longer-term lethality. The number of viable zebrafish larvae was calculated as a percentage of the total larvae for each concentration and time challenge.

6.3.4 Recovery and characterisation of NfsA variants exhibiting improved activation of 5-nitroimidazole prodrugs in *E. coli*

Figure 6.8: Chemical structures of 5-nitroimidazole prodrugs metronidazole and tinidazole.

Although NfsB_Vv F70A/F108Y is the current lead nitroreductase for targeted cell ablation in partnership with metronidazole, I considered that the top variants from the NfsA Marsden library might yield further gains in efficacy with metronidazole or alternative nitroimidazole prodrugs. I initially focused on variants exhibiting enhanced reduction with the 5-nitroimidazole prodrugs metronidazole and tinidazole (Figure 6.8). In Chapter 5, the Marsden library was plated over two stringent positive selection conditions containing chloramphenicol and/or niclosamide. Colonies recovered from these selection conditions

were counter-screened in growth inhibition assays with a range of nitroaromatic prodrugs, including the 5-nitroimidazole prodrug, metronidazole. Counter-screening with metronidazole revealed that 100% of the variants recovered from these conditions exhibited substantially higher growth inhibition compared to wild-type NfsA (Figure 5.1). These two positive selection conditions were chosen to recover NfsA variants exhibiting improved activation of the 5-nitroimidazole prodrugs metronidazole and tinidazole.

The Marsden library was plated on these two stringent positive selection conditions (45 μ M chloramphenicol or both 50 μ M niclosamide and 40 μ M chloramphenicol). Colonies which grew under these selection conditions were picked into the inner 60 wells of 96-well plates alongside one pUCX_empty and two wild-type NfsA controls. A total of six 96-well plates (representing 342 colonies) were picked for each selection condition. Strains overexpressing *nfsA* variants were challenged with metronidazole (3 μ M) and tinidazole (3 μ M) in growth inhibition assays. All assays were conducted in the presence of 50 μ M IPTG.

The ten NfsA variants that sensitised *E. coli* 7NT to the greatest extent with metronidazole and tinidazole were taken forward for DNA sequencing and more sensitive IC₅₀ growth assays (following validation by re-transformation of *E. coli* 7NT cells and re-testing, to ensure any changes in activity were due to mutations introduced into the nitroreductase gene and not acquired chromosomal mutations). Sequence data and IC₅₀ values of *E. coli* 7NT cells expressing the top ten metronidazole-reducing *nfsA* variants are shown below (Table 6.1). *E. coli* 7NT strains expressing wild-type *nfsA* and *nfsB_Vv F70A/F108Y* were included as controls. IC₅₀ data rather than enzyme kinetics were used as the primary measure of activity, as the goal was to achieve effective cell ablation.

Variant	Metronidazole IC ₅₀ (μM)	S41	L43	H215	T219	K222	S224	R225	F227
NfsA 41_18	1.7 ± 0.4	Y	V	D	F	M	V	N	N
NfsA 40_34	1.9 ± 0.3	H	I	C	Y	V	V	N	H
NfsA 40_51	1.9 ± 0.2	Y	L	V	Y	G	C	V	N
NfsA 36_57	1.9 ± 0.3	H	L	L	Y	L	V	N	D
NfsA 21_19	2.0 ± 0.5	H	V	D	Y	V	R	N	F
NfsA 43_51	2.3 ± 0.8	F	V	I	F	R	V	N	H
NfsA 36_38	2.1 ± 0.5	Y	L	C	Y	V	R	V	G
NfsA 39_15	2.3 ± 0.1	I	L	H	I	L	L	N	H
NfsA 36_25	2.3 ± 0.8	Y	L	N	Y	G	C	L	G
NfsA 22_7	2.4 ± 0.2	Y	V	G	F	R	D	I	S
NfsB_Vv F70A/F108Y	8.8 ± 1.3								
Wild-type NfsA	53.7 ± 4.2								

Table 6.1: Metronidazole IC₅₀ values and sequences of the top 10 *nfsA* variants expressed in *E. coli* 7NT. *E. coli* cultures expressing *nfsA* variants were challenged for 4 hours across serial dilutions of 0-100 μM metronidazole including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ of cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represents the average of five biological repeats ± SD. The amino acid present at each of the eight-targeted positions in the Marsden library is shown. The most prevalent amino acid at each position is shaded in grey if it makes up ≥ 50% of all amino acids at any given position.

All NfsA variants tested sensitised *E. coli* 7NT cells to a substantially lower metronidazole concentrations than wild-type NfsA and NfsB_Vv F70A/F108Y, with IC₅₀ values ranging from 1.7 to 2.4 μM metronidazole (Table 6.1). The *E. coli* 7NT strain expressing *nfsA* 41_18 had a 32-fold lower IC₅₀ value than the *E. coli* 7NT strain expressing wild-type *nfsA*, and a 5-fold lower IC₅₀ value than the *E. coli* 7NT strain expressing *nfsB_Vv* F70A/F108Y. This variant contained amino acid substitutions at all of the eight-targeted positions in the Marsden library. Despite only obtaining the amino acid sequence of ten NfsA variants, there were clear amino acid preferences evident at some positions. Unsurprisingly, these amino acid preferences are similar to those observed in Chapter 3 and Chapter 5, in NfsA variants also recovered from selections containing chloramphenicol and/or niclosamide (Table 3.8 and Table 5.2).

Sequence data and IC₅₀ values for *E. coli* 7NT cells expressing the top ten tinidazole-reducing *nfsA* variants are shown in Table 6.2. *E. coli* 7NT strains expressing wild-type *nfsA* and *nfsB_Vv* F70A/F108Y were also included as controls.

Variant	Tinidazole IC ₅₀ (μM)	S41	L43	H215	T219	K222	S224	R225	F227
NfsA_Ec 41_18	2.6 ± 0.7	Y	V	D	F	M	V	N	N
NfsA_Ec 42_57	3.1 ± 0.5	Y	V	F	Y	Q	H	C	L
NfsA_Ec 21_19	3.2 ± 0.5	H	V	D	Y	V	R	N	F
NfsA_Ec 41_7	3.4 ± 0.4	H	C	G	Y	M	I	N	H
NfsA_Ec 36_31	3.6 ± 0.7	H	C	V	F	K	H	N	N
NfsA_Ec 39_15	3.6 ± 0.2	I	L	H	I	L	L	N	H
NfsA_Ec 41_3	3.6 ± 0.5	F	V	G	F	M	H	C	H
NfsA_Ec 43_44	3.7 ± 1.1	Y	I	N	F	T	L	N	N
NfsA_Ec 36_38	3.8 ± 0.7	Y	L	C	Y	V	R	V	G
NfsA_Ec 43_51	3.8 ± 1.4	F	V	I	F	R	V	N	H
NfsB_Vv F70A/F108Y	26.6 ± 3.9								
Wild-type NfsA_Ec	30.1 ± 1.3								

Table 6.2: Tinidazole IC₅₀ values and sequences of the top 10 *nfsA* variants expressed in *E. coli* 7NT cells. *E. coli* cultures expressing *nfsA* variants were challenged for 4 hours across serial dilutions of 0-100 μM tinidazole including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ of cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represents the average of four biological repeats ± SD. The amino acid present at each of the eight-targeted positions in the Marsden library is shown. The most prevalent amino acid at each position is shaded in grey if it makes up ≥ 50% of all amino acids at any given position.

All NfsA variants tested sensitised *E. coli* 7NT cells to substantially lower tinidazole concentrations than wild-type NfsA and NfsB_Vv F70A/F108Y, with IC₅₀ values ranging from 2.6 to 3.8 μM tinidazole (Table 6.2). *E. coli* 7NT cells expressing *nfsA* 41_18 were found to have a 10-fold lower tinidazole IC₅₀ value than *E. coli* expressing either wild-type *nfsA* or *nfsB_Vv F70A/F108Y*. NfsA 41_18 was also the lead metronidazole-reducing variant, which likely reflects the structural similarities between metronidazole and tinidazole (Figure 6.8). Although not quite as overt as was observed with the top ten metronidazole-activating variants, very similar sequence patterns were nevertheless evident among the top tinidazole reducing NfsA variants.

6.3.5 Expression and characterisation of *nfsA* variants with improved activation of metronidazole and tinidazole in HEK-293 cells.

6.3.5.1 Assessing fluoroprobe activation of the top metronidazole and tinidazole activating NfsA variants in *E. coli* 7NT cells.

The ultimate goal of this research is to express the top evolved *nfsA* variants in zebrafish to mediate targeted cell ablation and generate disease models. As it takes a minimum of six

months to generate a stable transgenic zebrafish line (Section 2.10.4.2), we wanted to maximise the likelihood that any nitroreductase variants would be tolerated and express in zebrafish. The human cell line HEK-293 was used as an alternative eukaryote model; zebrafish cell lines not being available to this study as they are difficult to culture and work with (*personal communication*, Assoc. Prof. Jeff Mumm). In contrast, HEK-293 cells are easy to work with, amenable to plasmid transfection and established protocols are in place within the Ackerley lab (Condon, 2013; Williams, 2013; Sharrock, 2018). Although our sample size is still quite small and the correlation not absolute, data to date suggest that functional nitroreductase expression in HEK-293 cells is generally predictive of nitroreductase functionality in zebrafish (*personal communication*, Prof. David Ackerley), as was the case for NfsB_Vv F70A/F108Y.

The nitro-quenched fluoroprobe FSL61 was used to assess nitroreductase expression and function in transiently transfected cell lines. Upon nitro-reduction, FSL61 is converted to a fluorescent form which can be detected under a fluorescence microscope using the DAPI filter (excitation/emission 355/460 nm). Pilot tests were carried out in *E. coli* 7NT strains expressing the *nfsA* variants to evaluate whether evolution had impaired the ability of each NfsA variant to activate the fluoroprobe as per Section 2.8.3. Following FSL61 incubation, most *E. coli* 7NT strains exhibited a substantially greater fluorescence than the *E. coli* 7NT strain expressing *pUCX_empty*, however not as great as the *E. coli* 7NT strain expressing wild-type *nfsA* (Figure 6.9).

Figure 6.9: Activation of FSL61 by *E. coli* 7NT strains expressing the top metronidazole and tinidazole reducing *nfsA* variants. Fluorescence of *E. coli* 7NT strains expressing *nfsA* variants or wild-type *nfsA* or pUCX_empty as positive and negative controls respectively was determined following a 1 hour incubation with 20 μ M FSL61. Fluorescence was quantified on a fluorescent plate reader (ex/em 355/460 nm) and is reported as relative fluorescence units (RFU). Data representative of two technical repeats $\pm$ SD.

6.3.5.2 Generation of HEK-293 cell lines transiently expressing *nfsA* variants

HEK-293 cell lines transiently expressing the top *nfsA* variants were generated to gain an initial indication of whether expression would be tolerated. The mammalian expression vector F279-V5 was used to generate transgenic cell lines, as previously described by Prosser *et al.*, (2013). F279-V5 constructs were generated containing the genes encoding *nfsA* 41_18, 40_34, 40_51, 36_57, 42_57, 21_19 and 41_7. Forty-eight hours following transfection, the HEK-293 cell lines were tested for transient nitroreductase expression using the nitro-quenched fluoroprobe FSL61 (Section 2.9.5.1). All HEK-293 cell lines showed some level of fluorescence brighter than wild-type HEK-293 cells, indicative of transient *nfsA* expression (images not shown).

6.3.5.3 Generation of HEK-293 cell lines stably expressing *nfsA* variants

HEK-293 cell lines stably expressing these *nfsA* variants were generated, to assess whether expression of the *nfsA* variants could be tolerated long-term, expressing from the host cell chromosome, rather than transient expression directly from the transfected F279-V5 construct. Integration of the *nfsA* gene can occur at multiple sites throughout the HEK-293 cell genome, resulting in differing expression levels of the nitroreductase depending on whether it had been integrated in region under high or low expression in genome. Therefore, polyclonal HEK-293 cell lines were generated to prevent this bias, creating an averaged

population. This approach is consistent with other studies conducted in the Ackerley lab (Prosser *et al.*, 2013; Copp *et al.*, 2017). Transfections of each construct were carried out in triplicate. Cells which had integrated the *nfsA* gene were selected by repeated sub-culturing of media containing increasing concentrations of puromycin (1-3 μ M) over a period of several weeks (Section 2.9.4). Triplicate transfections were then pooled together prior to downstream analysis, to minimise bias from any dominant chromosomal integration event. Cell lines were frozen down and stored as per Section 2.10.4.2. To gain an initial indication of nitroreductase expression, all HEK-293 cell lines were incubated with 50 μ M FSL61 and subsequently visualised by fluorescence microscopy as per Section 2.9.5.1 (Figure 6.10).

Figure 6.10: HEK-293 cell lines stably expressing *nfsA* variants visualised by the activated FSL61 fluorophore. HEK-293 cell lines stably expressing *nfsA* variants were challenged with 50 μ M FSL61 fluorophore for 1 hour before visualisation on a fluorescence microscope using the DAPI filter. Cells lines were compared to wild-type HEK-293 cells to determine the background level of cell auto-fluorescence. Exposure times were kept consistent between different samples.

Most cell lines exhibited comparable fluorescence to HEK-293 cells expressing wild-type *nfsA*, suggesting these variants are being expressed at a high level (*nfsA* 40_51, 36_57, 42_57 and 21_19). HEK-293 cells expressing *nfsA* 41_18, 40_34 and 41_7 exhibited lower fluorescence, suggesting these *nfsA* variants are expressed at a lower level. The fluorescence intensity observed in HEK-293 cells (Figure 6.10) does not correlate to what

was observed in *E. coli* 7NT cells (Figure 6.9). This may reflect varying nitroreductase expression levels in HEK-293 cells.

6.3.5.3.1 Optimisation of a flow cytometry protocol using the fluorophore FSL61

Flow cytometry using FSL61 was tested to quantitatively determine the portion of cells expressing a *nfsA* variant in each cell line. It has been noted in the past that the fluorophore FSL61 is not fully cell entrapped, meaning upon activation, FSL61 can leave the cell that it was activated in and enter another non-nitroreductase expressing cell (Horvat, 2012). This effect can give a false representation of the proportion of cells expressing a nitroreductase in a population. While high concentrations of FSL61 are needed for fluorescence microscopy in order to visualise the cells, much lower concentrations can be used for flow cytometry, as this instrument is much more sensitive. I sought to determine whether lower concentrations of FSL61 would minimise leeching of activated FSL61, to provide an accurate representation of the level of nitroreductase expression present in a transgenic cell line.

Mixed populations of wild-type HEK-293 (non-transgenic) and HEK-293 cells expressing wild-type *nfsA* were incubated with 5 μ M or 50 μ M FSL61. Following incubation, the cells were analysed via flow cytometry to determine the extent of leeching of the activated FSL61 (Figure 6.11).

Figure 6.11: Flow cytometry of mixed populations of wild-type HEK-293 cells and HEK-293 cells expressing wild-type *nfsA* incubated with FSL61. Mixed cultures of 90% of wild-type HEK-293 cells with 10% HEK-293 cells expressing wild-type *nfsA* were challenged with either (A) 5 μ M FSL61 or (B) 50 μ M FSL61 for 1 hour (red). A population not challenged with any FSL61 was included as a control to determine the level of background cell auto-fluorescence (blue). The percentage of cells present in each fluorescent peak of the red population was determined using FlowJo software.

Two distinct groups were evident in both Figure 6.11A and Figure 6.11B, where 10% of cells (expressing wild-type *nfsA*) exhibited a greater fluorescence than the remaining 90% of cells (wild-type HEK-293 cells). A greater fluorescent shift was evident in the wild-type HEK-293 cell population when incubated with 50 μ M FSL61 (Figure 6.11B), suggesting wild-type HEK-293 cells had taken up some of the leached activated FSL61 fluoroprobe. This was not evident in the population incubated with a lower concentration of 5 μ M FSL61, showing a lower concentration of FSL61 resulted in substantially less leeching of the activated fluoroprobe (Figure 6.11A).

6.3.5.3.2 Flow cytometry analysis of HEK-293 cell lines stably expressing *nfsA* variants

The method described in the previous section was used to quantify nitroreductase expression in the HEK-293 cell lines generated (Figure 6.12). HEK-293 cells stably expressing wild-type *nfsA* and wild-type HEK-293 cells were included as a positive and negative control respectively.

Figure 6.12: Flow cytometry of HEK-293 cells stably expressing the top metronidazole and tinidazole activating *nfsA* variants. Cell lines were incubated with 5 μ M FSL61 for 1 hour. Following the incubation, the cells were analysed by flow cytometry and compared to a non-transgenic population (red) to determine the background level of cell auto-fluorescence. The mean fluorescence for each cell line is presented. Image generated using FlowJo software.

Most of the HEK-293 cell lines exhibited similar fluorescent shifts and had a mean fluorescence comparable to HEK-293 cells expressing wild-type *nfsA* (NfsA 40_34, 40_51, 36_57, 41_7, 21_19 and 41_18). HEK-293 cells expressing *nfsA* 42_57 had two distinct fluorescent peaks with the majority of the population appearing non-fluorescent. These findings are consistent with the diminished ability of NfsA 42_57 to activate FSL61 as observed in *E. coli* 7NT (Figure 6.9). The bright fluorescence observed in HEK-293 cells expressing *nfsA* 42_57 under the fluorescence microscope may be due to the higher

concentration of FSL61 that was used to visualise the cells under the microscope (50 μ M), resulting in leeching of activated FSL61. The use of FSL61 in fluorescence microscopy still provides a useful indication of whether a nitroreductase is expressing, as flow cytometry is a time-consuming process.

6.3.5.3.3 Western blot analysis of HEK-293 cell lines expressing *nfsA* variants

As an additional means to assess nitroreductase expression in the HEK-293 cell lines, western blots were generated. An anti-NfsA antibody was generously donated by Dr Peter Searle (*University of Birmingham*) which enabled probing of cell lines expressing *nfsA* variants. Expression of the house-keeping protein α -tubulin was also detected and used as a sample loading control. ImageJ^{TM} software densitometry analysis (version 1.8.0) was used to determine the relative expression levels of *nfsA* variants compared to HEK-293 cells expressing wild-type *nfsA*, normalised to the density of the α -tubulin loading controls (Figure 6.13).

Figure 6.13: Western blot depicting the expression levels of *nfsA* variants in HEK-293 cells for variants exhibiting improved activation with metronidazole and tinidazole. HEK-293 cell samples were lysed with RIPA buffer and probed for NfsA using a NfsA polyclonal primary antibody raised in sheep and a donkey anti-sheep IgG HRP conjugate secondary antibody. The expression levels of α -tubulin in samples were used as loading controls and were detected via immunofluorescence.

HEK-293 cells expressing *nfsA* 42_57, 41_7 and 40_51 showed the lowest expression relative to HEK-293 cells expressing wild-type *nfsA* (Figure 6.13). The other *nfsA* variants exhibited varying degrees of expression (65-110%), with HEK-293 cells expressing *nfsA* 36_57 at a greater level than wild-type *nfsA* (110%).

6.3.5.4 IC₅₀ assays of HEK-293 cell lines stably expressing *nfsA* variants

HEK-293 cell lines stably expressing *nfsA* variants were tested in IC₅₀ assays to quantify the relative abilities of the various NfsA variants to sensitise HEK-293 cells to either metronidazole or tinidazole (section **Error! Reference source not found.**). HEK-293 cell lines were seeded at a density of 13,000 cells per well of a treated 96-well plate. Cells were left to adhere to the plate for 24 hours before being challenged with 0-100 µM metronidazole or tinidazole for 48 hours. Percentage cell survival was determined at each prodrug concentration and GraphPad 8.0 software was used to determine the IC₅₀ value (Table 6.3).

Variant	Metronidazole IC ₅₀ (µM)	Variant	Tinidazole IC ₅₀ (µM)
NfsA 41_18	5.3 ± 1.1	NfsA 21_19	2.7 ± 0.5
NfsA 40_34	10.1 ± 3.6	NfsA 41_18	5.1 ± 1.1
NfsA 40_51	11.8 ± 1.2	NfsA 41_7	9.4 ± 1.5
NfsA 36_57	15.4 ± 4.9	Wild-type NfsA	46 ± 12
Wild-type NfsA	50.9 ± 13.5	NfsA 42_57	93 ± 24
NfsB_Vv F70A/F108Y	2.1 ± 0.7	NfsB_Vv F70A/F108Y	11.1 ± 3.6
Wild-type HEK293	>10000	Wild-type HEK293	>10000

Table 6.3: Metronidazole and tinidazole IC₅₀ values of HEK-293 cell lines stably expressing *nfsA* variants. Replicate cultures of HEK-293 cells expressing *nfsA* variants were grown in 96-well plates and challenged with 0-100 µM metronidazole or tinidazole for 48 hours. MTS reagent was added to the wells and incubated for an additional 2-3 hours. The absorbance of unchallenged wells were measured at 490 nm were compared to corresponding challenged wells to determine the percentage viability at a given prodrug concentration. The IC₅₀ value was calculated using GraphPad 8.0 software. Data represents the average of four biological replicates ± SD.

All NfsA variants sensitised HEK-293 cells to a lower concentration of metronidazole than wild-type NfsA. The best NfsA variant, 41_18, sensitised HEK-293 cells to an almost 10-fold lower concentration of metronidazole than wild-type NfsA. However, in contrast with assays in *E. coli* 7NT cells, none of the NfsA variants tested were able to sensitise HEK-293 cells to a lower concentration of metronidazole than NfsB_Vv F70A/F108Y. NfsA 41_18, 40_34 and 36_57 were expressed at a comparable level to wild-type *nfsA* (Figure 6.10, Figure 6.12 and Figure 6.13), therefore decreased expression does not explain the activity disparity seen between *E. coli* and HEK-293 cells. This phenomenon may be a consequence of these NfsA enzyme variants being selected in *E. coli* cells and having lost activity with competing metabolites that are more prominent in *E. coli* than in HEK-293 cells. In contrast, NfsB_Vv F70A/F108Y was rationally engineered specifically for

enhanced metronidazole catalysis, an attribute that would be expected to improve activity in both cell types. This phenomenon has been previously discussed in Section 3.4.

Three of the four NfsA variants sensitised HEK-293 cells to a lower concentration of tinidazole than wild-type NfsA and NfsB_Vv F70A/F108Y (NfsA 21_19, 41_18 and 41_7). The best NfsA variant, 21_19 sensitised HEK-293 cells to a 17-fold lower concentration of tinidazole than wild-type NfsA and a 7-fold lower concentration than NfsB_Vv F70A/F108Y. NfsA 42_57 was the least active variant with tinidazole. This is consistent with flow cytometry (Figure 6.12) and western blot analyses (Figure 6.13), which indicated that *nfsA* 42_57 is expressed at a substantially lower level than wild-type *nfsA* in HEK-293 cells.

The genes encoding the top two *nfsA* variants with each prodrug, *nfsA* 41_18, 40_34 and 21_19, were cloned into the zebrafish expression vector pTOL2 and sent to our collaborators at the Wilmer Eye Institute to be tested in zebrafish. Although none of the NfsA variants identified in this study exhibited improved metronidazole reduction over NfsB_Vv F70A/F108Y, most of the variants expressed highly in HEK-293 cells, so are likely to express in zebrafish and may show more promising prodrug activation in this different model. Several NfsA variants were also identified which exhibited improved tinidazole reduction over NfsB_Vv F70A/F108Y and wild-type NfsA. This work is ongoing in the Mumm lab.

6.3.6 Recovery and characterisation of NfsA variants exhibiting improved activation of 2-nitroimidazole prodrugs in *E. coli*

RB6145 and misonidazole were chosen as two additional nil-bystander prodrugs that could potentially be paired with a nitroreductase for targeted cell ablation in zebrafish (Figure 6.5). We consider it possible that alternative prodrug-nitroreductase pairings may have utility in the ablation of different and more difficult cell types, which cannot be achieved with the current “gold-standard” NfsB_Ec/metronidazole ablation system. However, to test this hypothesis it was first necessary to recover nitroreductases that were highly active with the alternative prodrug candidates. The Marsden library was used as a resource to recover NfsA variants capable of reducing RB6145 and misonidazole. A second library source known as the “2-nitroimidazole combination library” was also used. This library contains a different set of targeted residues and was originally designed with the intent of evolving

NfsA variants exhibiting improved activation of 2-nitroimidazole prodrugs (Section 6.1.4.1).

In Chapter 5, *E. coli* 7NT cells expressing the Marsden library were plated on stringent positive selections containing chloramphenicol and/or niclosamide. Colonies which grew under these selection conditions were then challenged in growth inhibition assays with a range of nitroaromatic prodrugs, including the 2-nitroimidazole prodrug, RB6145. This recovered a large number of NfsA variants which were better at sensitising *E. coli* 7NT cells to RB6145 than wild-type NfsA (Figure 5.1). Replicates of the 96-well plates of colonies selected from 45 μ M chloramphenicol or 50 μ M niclosamide and 40 μ M chloramphenicol, generated in Section 6.3.4, were challenged with RB6145 (5 μ M) or misonidazole (75 μ M) in growth inhibition assays. Fourteen 96-well plates containing an estimated 90% coverage of the 2-nitroimidazole combination library were also challenged with the same concentrations of RB6145 and misonidazole. All assays were conducted in the presence of 50 μ M IPTG.

The ten NfsA variants that sensitised *E. coli* 7NT cells to the greatest extent with RB6145 and misonidazole were taken forward for DNA sequencing and more sensitive IC₅₀ growth assays. Prior to this, plasmids encoding the NfsA variants were re-transformed into *E. coli* 7NT cells to ensure any changes in activity were attributed to mutations introduced into the nitroreductase gene and not due to acquired chromosomal mutations. Sequence data and IC₅₀ values of *E. coli* 7NT cells expressing the top ten RB6145 reducing *nfsA* variants are shown below (Table 6.4). *E. coli* 7NT cells expressing wild-type *nfsA* and *nfsB_Vv F70A/F108Y* included as controls. Through chance, five NfsA variants had originated from each library.

Marsden Variant	RB6145 IC ₅₀ (μM)	S41	L43	H215	T219	K222	S224	R225	F227
NfsA 39_37	5.6 ± 0.5	Y	L	H	Y	R	I	Y	S
NfsA 35_48	10.2 ± 2.1	Y	L	F	Y	V	R	N	F
NfsA 43_26	10.3 ± 1.9	H	I	C	Y	V	V	N	H
NfsA 30_27	15.2 ± 1.7	F	C	N	Y	R	D	L	D
NfsA 22_43	17.1 ± 1.8	F	I	N	Y	V	R	N	N

2-nitroimidazole Combo Variant	RB6145 IC ₅₀ (μM)	R12	S40	S41	F42	C45	L103	N134	H215
NfsA 12_63	6.5 ± 1.2	R	S	Y	F	C	V	A	H
NfsA 2_67	9.2 ± 1.3	R	T	Y	F	C	L	A	H
NfsA 11_95	10.1 ± 1.1	S	S	C	F	C	V	N	R
NfsA 4_55	10.7 ± 0.9	R	S	Y	F	C	V	N	H
NfsA 13_60	11.2 ± 1.4	S	T	Y	F	C	V	N	H
Wild-type NfsA	15.8 ± 0.8								
NfsB_Vv F70A/F108Y	30.1 ± 2.4								

Table 6.4: RB6145 IC₅₀ values and sequences of the top ten nfsA variants expressed in *E. coli* 7NT. *E. coli* cultures expressing nfsA variants were challenged for 4 hours across serial dilutions of RB6145 (0-200 μM) including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ of cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate IC₅₀ values. Data represents the average of five biological repeats ± SD. The amino acid present at each of the eight-targeted positions in the Marsden and 2-nitroimidazole combination libraries is shown. The most prevalent amino acid at each position is shaded in grey if it makes up ≥ 50% of all amino acids at any given position.

The proportional improvement in RB6145 reduction was not as profound as previously seen with the metronidazole or tinidazole improved NfsA variants (Table 6.1 and Table 6.2). This is consistent with the results seen in Chapter 5 (Figure 5.1), where stringent positive selection conditions yielded a greater proportion of NfsA variants exhibiting improved metronidazole activation over RB6145. The *E. coli* 7NT strain expressing the best variant, nfsA 39_37, had a 2.8-fold lower IC₅₀ value than the *E. coli* 7NT strain expressing wild-type nfsA, and a 5-fold lower IC₅₀ value than the *E. coli* 7NT strain expressing nfsB_Vv F70A/F108Y (Table 6.4).

Only five of the ten NfsA variants originated from the Marsden library, despite the Marsden library containing 252 million unique variants compared to 576 variants in the 2-

nitroimidazole combination library. This confirms that the predefined set of amino acid substitutions introduced into the 2-nitroimidazole library did generate NfsA variants exhibiting improved 2-nitroimidazole prodrug activation. In contrast, the amino acid substitutions introduced to the Marsden library were not targeted towards 2-nitroimidazole prodrugs. Despite the small sample size, some patterns appeared evident within each library. In the NfsA variants derived from the 2-nitroimidazole library there was a preference for S41Y and L103V substitutions, while at positions 42, 45, 134 and 215 there was a preference for the wild-type amino acid over the mutated residue. The NfsA variants derived from the Marsden library exhibited a similar pattern of amino acid substitutions to the NfsA variants which exhibited improved metronidazole and tinidazole reduction.

Sequence data and IC₅₀ values of *E. coli* 7NT strains expressing the top ten misonidazole-reducing *nfsA* variants are shown below (Table 6.5). *E. coli* 7NT strains expressing wild-type *nfsA* and *nfsB_Vv F70A/F108Y* were included as controls. The top ten variants were again coincidentally made up of five variants from each library.

Marsden Variant	Misonidazole IC ₅₀ (μM)	S41	L43	H215	T219	K222	S224	R225	F227
NfsA 40_34	31.8 ± 5.5	H	I	C	Y	V	V	N	H
NfsA 37_10	36.0 ± 7.0	Y	S	S	L	L	R	N	H
NfsA 19_13	45.1 ± 6.7	C	L	I	V	L	V	N	H
NfsA 43_56	46.5 ± 6.0	Y	C	H	F	M	R	N	F
NfsA 20_31	53.5 ± 9.4	Y	C	H	Y	L	Y	N	R

2-nitroimidazole Combo Variant	Misonidazole IC ₅₀ (μM)	R12	S40	S41	F42	C45	L103	N134	H215
NfsA 2_70	38.4 ± 6.7	R	S	Y	I	C	L	A	H
NfsA 12_57	39.2 ± 7.4	R	S	C	I	C	V	N	H
NfsA 13_60	40.7 ± 5.2	S	T	Y	F	C	V	N	H
NfsA 13_65	46.7 ± 10.2	R	T	Y	F	C	V	N	H
NfsA 4_52	51.7 ± 12.3	S	S	Y	F	C	V	N	H
Wild-type NfsA	81.2 ± 13.2								
NfsB_Vv F70A/F108Y	127 ± 12.0								

Table 6.5: Misonidazole IC₅₀ values and sequences of the top ten *nfsA* variants expressed in *E. coli* 7NT. *E. coli* cultures expressing *nfsA* variants were challenged for 4 hours across serial dilutions of misonidazole (0-300 μM) including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ of cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represents the average of five biological repeats ± SD. The amino acid present at each of the eight-targeted positions in both the Marsden and 2-

nitroimidazole libraries is shown. The most prevalent amino acid at each position is shaded in grey if it makes up $\geq 50\%$ of all amino acids at any given position.

The *E. coli* 7NT strain expressing the top variant, *nfsA* 40_34, had a 2.6-fold lower IC₅₀ value than the *E. coli* 7NT strain expressing wild-type *nfsA*, and a 4-fold lower IC₅₀ value than the *E. coli* 7NT strain expressing *nfsB*_Vv F70A/F108Y (Table 6.5). This variant was also the second-best metronidazole-reducing variant (Table 6.1). As previously discussed, it is predicted that the evolved NfsA variants have lost activity with competing metabolites, as a result of being evolved in *E. coli*. This may explain why a NfsA variant appears improved with multiple prodrugs, as it has lost activity with competing metabolites, enabling it to be more active with a variety of alternative substrates when expressed in *E. coli* 7NT.

There was substantial overlap in the amino acid substitutions evident between the top RB6145 and misonidazole reducing NfsA variants. However, one main difference was apparent at position 219 where the preference for aromatic residues was not as clear in the top misonidazole-reducing variants (Table 6.5). In contrast, NfsA variants exhibiting activation of all of the other prodrugs tested (metronidazole, tinidazole and RB6145) had a clear preference for aromatic residues at this position.

6.3.7 Expression and characterisation of *nfsA* variants with improved activation of RB6145 and misonidazole in HEK-293 cells.

6.3.7.1 Assessing fluoroprobe activation of the top RB6145 and misonidazole activating NfsA variants in *E. coli* 7NT cells.

Prior to generating transfected human cell lines, the ability of *E. coli* 7NT strains expressing to the top RB6145 and misonidazole activating *nfsA* variants to activate the nitro-quenched fluoroprobe FSL61 was evaluated as per Section 2.8.3. This was necessary to establish whether nitroreductase expression could be monitored using FSL61 or whether evolution had impaired the ability of each NfsA variant to activate this fluoroprobe. All *E. coli* 7NT strains exhibited a substantially higher fluorescence than the *E. coli* 7NT strains expressing *pUCX_empty*, with most comparable to the *E. coli* 7NT strain expressing wild-type *nfsA* (Figure 6.14).

Figure 6.14: Activation of FSL61 by *E. coli* 7NT strains expressing the top RB6145 and misonidazole activating *nfsA* variants. Fluorescence of *E. coli* 7NT strains expressing *nfsA* variants or wild-type *nfsA* or pUCX_empty as positive and negative controls respectively following a 1 hour incubation with 20 μ M FSL61. Fluorescence was quantified on a fluorescent plate reader (ex/em 350/460 nm) and is reported as relative fluorescence units (RFU). Data representative of two technical repeats $\pm$ SD.

6.3.7.2 Generation of HEK-293 cell lines transiently expressing *nfsA* variants

As previously (Section 6.3.5.2), HEK-293 cell lines transiently expressing the *nfsA* variants were first generated as an initial test to see if the expression of these *nfsA* variants could be tolerated in HEK-293 cells prior to generating stably expressing cell lines. The top four misonidazole and RB6145 reducing *nfsA* variants (*nfsA* 39_37, 12_63, 2_67, 11_95, 40_34, 37_10, 2_70 and 12_57) were cloned into the human expression vector F279-V5. As a HEK-293 cell line expressing *nfsA* 40_34 had previously been created (Section 6.3.5.3), this cell line was not created again. Forty-eight hours following transfection, transient nitroreductase expression was tested by addition of FSL61 (Section 2.9.5.1). All HEK-293 cell lines exhibited substantially brighter fluorescence than wild-type HEK-293 cells, indicative of transient *nfsA* expression (images not shown).

6.3.7.3 Generation of stably expressing HEK-293 cell lines

Polyclonal HEK-293 cell lines were generated expressing the *nfsA* variants described above using the procedure described in Section 2.9.4. Cell lines were frozen down and stored as per Section 2.9.2. To assess nitroreductase expression, all stable cell lines were tested with the nitro-quenched fluoroprobe FSL61 and assessed by both fluorescence microscopy (Figure 6.15) and flow cytometry (Figure 6.16). HEK-293 cells expressing *nfsA* 40_34 were also tested to ensure expression levels had not decreased over time.

Figure 6.15: HEK-293 cell lines stably expressing *nfsA* variants visualised using the nitro-quenched fluorophore FSL61. HEK-293 cell lines stably expressing *nfsA* variants were incubated with 50 μ M FSL61 fluorophore for 1 hour before visualisation on a fluorescence microscope using the DAPI filter. Cells lines were compared to wild-type HEK-293 cells to determine the background level of cell auto-fluorescence. Exposure times were kept consistent between different samples.

All HEK-293 cell lines exhibited high fluorescence levels, albeit slightly lower than HEK-293 cells expressing wild-type *nfsA* (Figure 6.15). As most of the NfsA variants exhibited comparable activation of FSL61 in *E. coli*, this suggested these variants were being expressed at a slightly lower level to wild-type *nfsA*.

The flow cytometry protocol optimised in Section 6.3.5.3.1, was used to quantify nitroreductase expression in the HEK-293 cell lines. HEK-293 cell lines expressing *nfsA* variants were incubated with 5 μ M FSL61 for 1 hour, followed by analysis on the flow cytometer. HEK-293 cells expressing wild-type *nfsA* and wild-type HEK-293 cells were included as positive and negative controls respectively.

Figure 6.16: Flow cytometry of HEK-293 cells stably expressing the top RB6145 and misonidazole activating *nfsA* variants. Cell lines were incubated with 5 μ M FSL61 for 1 hour. Following incubation, the cells were analysed by flow cytometry and compared to a non-transgenic population (red) to determine the background level of cell auto-fluorescence. The mean fluorescence for each cell line is presented. Image generated using FlowJo software.

All of the HEK-293 cell lines expressing *nfsA* variants exhibited similar fluorescence shifts comparable to HEK-293 cells expressing wild-type *nfsA* (Figure 6.16). As an additional means to assess nitroreductase expression in the HEK-293 cell lines, western blots were generated (Section 2.9.5.3). The HEK-293 cells lines expressing *nfsA* 2_67, 12_63 and 40_34 exhibited comparable or greater expression than HEK-293 cell lines expressing wild-type *nfsA* (Figure 6.17). Unsurprisingly, the *nfsA* variants which expressed the best in HEK-293 cells lines (*nfsA* 2_67, 12_63 and 40_34) were able to sensitise HEK-293 cells to either RB6145 or misonidazole to the greatest extent (Table 6.6).

Figure 6.17: Western blot depicting the expression levels of *nfsA* variants when expressed in HEK-293 cells. HEK-293 cell samples were lysed with RIPA buffer and probed for NfsA using a sheep anti-NfsA polyclonal primary antibody and a donkey anti-sheep IgG HRP conjugate secondary antibody. The expression level of α -tubulin in samples was used as a loading control and was detected via immunofluorescence.

6.3.7.4 Characterisation of HEK-293 stably expressing *nfsA* variants

HEK-293 cell lines stably expressing *nfsA* variants were challenged with RB6145 or misonidazole and tested in IC₅₀ assays to quantify the ability of the NfsA variants to sensitise HEK-293 cells to these prodrugs (Table 6.6).

Variant	RB6145 IC ₅₀ (μM)	Variant	Misonidazole IC ₅₀ (μM)
NfsA 2_67	0.2 ± 0.06	Wild-type NfsA	3.6 ± 0.6
Wild-type NfsA	0.2 ± 0.05	NfsA 40_34	5.0 ± 0.7
NfsA 12_63	0.2 ± 0.02	NfsA 2_70	5.1 ± 1.0
NfsA 39_37	0.5 ± 0.2	NfsA 37_10	5.5 ± 1.5
NfsA 11_95	0.6 ± 0.2	NfsA 12_57	6.1 ± 0.9
NfsB_Vv F70A/F108Y	0.4 ± 0.09	NfsB_Vv F70A/F108Y	4.5 ± 1.1
Wild-type HEK293	55 ± 2	Wild-type HEK293	>1000

Table 6.6: RB6145 and misonidazole IC₅₀ values of HEK-293 cell lines stably expressing *nfsA* variants. Replicates of HEK-293 cells expressing *nfsA* variants were grown in 96-well plates and challenged with varying concentrations of either RB6145 (0-6.3 μM) or misonidazole (0-200 μM) for 48 hours. MTS reagent was added to the wells and incubated for an additional 2-3 hours. The absorbance of unchallenged wells were measured at 490 nm and were compared to that of challenged wells to determine the percentage viability at each prodrug concentration. The IC₅₀ value was calculated using GraphPad 8.0 software. Data represents the average of four biological replicates ± SD.

HEK-293 cells expressing the top *nfsA* variants had comparable or slightly higher RB6145 or misonidazole IC₅₀ values than HEK-293 cells expressing wild-type *nfsA*. The two best variants, NfsA 2_67 and 12_63, were able to sensitise HEK-293 cells to the same concentration of RB6145 (0.2 μM) as wild-type NfsA. NfsA 2_67 was expressed at a higher level than wild-type NfsA (136%) in HEK-293 cells, suggesting why this variant may appear more active than others (Figure 6.17). None of the NfsA variants were able to sensitise HEK-293 cells to a lower concentration of misonidazole, however did sensitise HEK-293 cells to a similar concentration (5 – 6.1 μM vs. 3.6 μM). As previously discussed, it is likely these NfsA variants have lost activity with competing metabolites, improving activity in *E. coli* but not in the HEK-293 cell environment. Nevertheless, as it was unknown how these variants might perform in zebrafish. The genes encoding the top two RB6145 and top two misonidazole reducing *nfsA* 2_67, 12_63, 40_34 and 2_70 were cloned into the zebrafish expression vector pTOL2 and sent to our collaborators at the Wilmer Eye Institute to be tested in zebrafish (wild-type *nfsA* having already been provided to them). NfsA 40_34 may be of particular interest as it is capable of sensitising HEK-293 cells to both misonidazole and metronidazole.

6.3.8 Development of a transient cell ablation assay in zebrafish.

A complete summary of the different *nfsA* variants generated in this work and sent to our collaborators is provided in Table 6.7.

Variant	Library Origin	Prodrug Pairing
NfsA 41_18	Marsden	Metronidazole & Tinidazole
NfsA 40_34	Marsden	Metronidazole & Misonidazole
NfsA 21_19	Marsden	Tinidazole
NfsA 2_67	2-nitroimidazole combo	RB6145
NfsA 12_63	2-nitroimidazole combo	RB6145
NfsA 2_70	2-nitroimidazole combo	Misonidazole

Table 6.7: Summary of the *nfsA* variants cloned into the zebrafish *pTOL2* vector.

No zebrafish lines expressing any of these *nfsA* variants had been generated prior to my visit to the Wilmer Eye Institute. Due to the long time it takes to generate stably expressing zebrafish lines (a minimum of six months), and the brief duration of our research visit (10 weeks), Abigail Sharrock and I looked at developing a transient cell ablation assay to enable rapid evaluation of new nitroreductase variants in zebrafish. We envisaged that this assay would involve injecting driver line embryos (gmc617 or gmc614) with a *pTOL2* construct containing a nitroreductase, and rely on transient nitroreductase expression to mediate cell ablation. Any larvae exhibiting TagYFP fluorescence at 4-6 dpf would be challenged with the desired prodrug and cell ablation monitored using the same plate reader-based assay. If this transient based assay were successful, it would avoid the need to generate large numbers of stable transgenic nitroreductase expressing zebrafish lines, allowing for a more high-throughput characterisation of nitroreductases in zebrafish and prioritisation of candidates to generate stably expressing zebrafish lines.

In order for this transient assay to work, it would require broad transient expression, capable of detection on the plate reader in order for cell ablation to be quantified. To achieve this, the following strategies were attempted: (1) increasing the UAS copy number to $14 \times$ UAS, (2) co-injection of a kaloooping construct to allow for continual driving of expression of UAS linked elements and (3) injection of a self-reporting “cis” Gal4/UAS construct (Appendix E). In brief, none of our attempts were successful in establishing a reliable transient expression assay within the timeframe of our visit. Although, our pilot data suggests that the use of a $14 \times$ UAS sequence in conjunction with a “cis” Gal4/UAS construct might offer some promise in this regard. For now, the Mumm lab is proceeding

with the slower, but more definitive process of generate stable zebrafish lines for all nitroreductase variants listed in Table 6.7. At the time of writing, this work was still in progress.

6.4 Discussion

The overarching goal of the research described in this chapter was to identify superior NfsA variants capable of activating metronidazole, tinidazole, RB6145 and misonidazole for use as tools in targeted cell ablation. A 10-week research trip was conducted at our collaborators laboratory at the Wilmer Eye Institute located at Johns Hopkins University. During this 10-week research trip the current lead metronidazole activating nitroreductase variant, NfsB_Vv F70A/F108Y was characterised in zebrafish. NfsB_Vv F70A/F108Y exhibited an improved ability to induce cell ablation compared to the current “gold standard” NfsB_Ec/metronidazole ablation system. Widespread-targeted ablation of cells expressing *nfsB_Vv* F70A/F108Y was achieved in 6-hours with 0.25 mM metronidazole or in 24-hours with 0.1 mM metronidazole. The latter represents a 100-fold decrease in the concentration of metronidazole required for the NfsB_Ec ablation system. Ongoing work in the Mumm lab involves the generation of transgenic zebrafish lines expressing *nfsB_Vv* F70A/F108Y in a variety of cell types to increase the range of degenerative disease models. This work is showing promising results and they have recently achieved ablation of macrophage cells expressing *nfsB_Vv* F70A/F108Y, a notoriously difficult cell-type to ablate (*personal communication, Assoc. Prof. Jeff Mumm*).

Several NfsA variants were recovered that sensitised *E. coli* 7NT cells to significantly lower concentrations of all nil-bystander prodrugs used in this study. The top metronidazole- and tinidazole-activating NfsA variants were able to sensitise *E. coli* 7NT cells to substantially lower concentrations of metronidazole and tinidazole than wild-type NfsA and NfsB_Vv F70A/F108Y (up to 32-fold), whereas the top RB6145- and misonidazole-activating NfsA variants achieved a smaller but still substantial improvement (up to 3-fold). Clear amino acid preferences were evident among the top NfsA variants for all prodrugs tested. These substitutions included aromatic residues at positions 41 and 219 and an asparagine residue at position 225. These amino acid substitutions were consistent with those seen in NfsA variants recovered from positive selection conditions in Chapter 3 and Chapter 5, where the predicted benefit of these three substitutions is outlined (Section

3.3.5.3). It is predicted that these residues may be involved in substrate stabilisation or opening of the active site pocket, rather than being involved in direct catalysis.

The majority of amino acid substitutions in the top NfsA variants appeared to improve reduction with both 5- and 2-nitroimidazole prodrugs, although appeared to act in favour of 5-nitroimidazole reduction, shown by lower metronidazole IC₅₀ values. Some minor amino acid differences were seen between the top 5- and 2-nitroimidazole activating NfsA variants. Some of these substitutions included K222V and L103V in the top 2-nitroimidazole reducing variants. Without testing the 2-nitroimidazole combo library derived NfsA variants with 5-nitroimidazole substrates, we cannot conclude whether the L103V substitution is specific for improved 2-nitroimidazole reduction, or whether this substitution improves overall nitroreductase activity. A crystal structure of NfsA with these prodrugs bound could be used to try determine the mechanistic role of each substitution and provide insights into the reduction of 5-nitroimidazole versus 2-nitroimidazole substrates. This has been attempted several times in the Ackerley lab with no success as acetate appears to bind preferentially over a nitroaromatic substrate (Sharrock, 2018). An analysis of the individual substitutions present in the top NfsA variants might prove an alternative means of gaining insight into differences between 5- and 2-nitroimidazole activation.

Unfortunately, all of the NfsA variants tested appeared substantially more active when evaluated in *E. coli* than in HEK-293 cells, despite good expression of most of these *nfsA* variants in HEK-293 cells. The lower apparent activity observed in HEK-293 cells is likely a consequence of performing directed evolution in the *E. coli* intracellular environment, optimising evolved variants specifically for this cell environment. This phenomenon was described in Chapter 3 (Section 3.4), predicting why nitroreductase activity observed *in vivo* (*E. coli* 7NT) did not translate to improved activity *in vitro* (purified enzyme kinetics). The intracellular environment is very different between *E. coli* and HEK-293 cells, and key competing metabolites present in *E. coli* might be less represented in HEK-293. This may also explain why NfsB_Vv F70A/F108Y exhibited a more consistent activity profile in *E. coli*, HEK-293 cells and zebrafish. NfsB_Vv F70A/F108Y was rationally engineered for improved substrate catalysis, rather than generic evolution for improved performance in a particular environment, therefore activity would be more likely to be retained in different cellular environments. In the future, a detailed analysis of enzyme kinetics with these

nitroaromatic prodrugs and competing metabolites could shed more light on this. This work is ongoing, and it remains to be seen how the top evolved NfsA variants will perform in zebrafish.

One variant, NfsA 40_34, may be of particular interest for nitroreductase mediated cell ablation system due to its ability to reduce both metronidazole and misonidazole. This NfsA variant could be used as a dual-purpose nitroreductase, allowing for the ablation of multiple or difficult to ablate cell types. One prodrug could be administered, mediating cell ablation, followed by a second prodrug, ablating any remaining stubborn cells. Metronidazole (PC = 0.96) and misonidazole (PC = 0.43) have different partition constants. The partition constant denotes the permeability of a membrane to a particular compound and can provide an indication of the bioavailability (Adams *et al.*, 1979; Brown and Workman, 1980). This difference may be useful in a dual-ablation system, using two prodrugs that can penetrate to different tissue depths, allowing for complete ablation of stubborn cell types.

Despite the evolution of NfsA not looking promising at this stage, advances have still been made. We have defined the toxic range of RB6145, setting a benchmark for future experiments with this compound. We have also obtained pilot data for a transient cell ablation assay involving the use of a $14 \times$ UAS sequence in conjunction with a “cis” Gal4/UAS construct. This will allow for a more high-throughput characterisation of nitroreductases in zebrafish. This directed evolution campaign has also provided insight into how we could adapt our current system to evolve better nitroreductases in the future.

Chapter 7 Directed evolution of NfsA for

Multiplex Targeted Cell Ablation

7.1 Introduction

Chapter 6 described the recovery and characterisation of NfsA variants that exhibited superior activation of the prodrugs metronidazole, tinidazole, RB6145 and misonidazole which could be used as tools to enable ablation of a specific cell population. We wanted to expand on this system and develop a multiplex cell ablation system to enable the independent ablation of two different cell types. Researchers in this field would benefit immensely from this system, enabling them to delineate interactions between two distinct cell populations for investigation of specialised immune cell sub-types, or during processes such as cell regeneration and development. The research presented in this chapter sought to identify NfsA variants which could have utility in this system.

7.1.1 Multiplex Cell Ablation

For a multiplex cell ablation system to be functional, two different nitroreductases would be required, each possessing non-overlapping specificities for different nil-bystander prodrugs. Each nitroreductase would be placed under the control of a different tissue-specific promoter, sensitising each cell type to a different nil bystander prodrug (Figure 7.1). Additionally, each nitroreductase can be co-expressed with a different fluorescent reporter to ensure non-overlapping prodrug specificities and monitor ablation.

Figure 7.1: Nitroreductase mediated multiplex cell ablation. A transgenic zebrafish is generated expressing pairs of nitroreductases with non-overlapping prodrug specificities expressed under different tissue-specific promoters. Nitroreductases are co-expressed with a fluorescent reporter. Upon administering prodrug A, only cells expressing NTR A are ablated, while administering prodrug B ablates cells expressing NTR B.

A multiplex ablation system would enable researchers to gain a deeper understanding of the regenerative processes which can occur in human degenerative diseases. One such example is type I diabetes, an autoimmune condition where T-cells and their cytokine products destroy the insulin producing β -cells in the pancreas (Rabinovitch and Suarez-Pinzon, 1998). There is a lot of interest in β -cell regeneration as a diabetic therapy. It has been shown that following extreme β -cell loss, adult pancreatic α -cells are able to undergo transdifferentiation to β -cells, however this process is not well understood (Chung *et al.*, 2010; Thorel *et al.*, 2010). Current diabetic models do not allow for the independent ablation of the α - and β -cells of the pancreas. A multiplex cell ablation system would allow for the independent, temporally controlled ablation of α - and/or β -cells, allowing researchers to gain a better understanding of α -cell transdifferentiation. This system would prove beneficial when studying complex degenerative diseases where multiple cell types are affected. For example, it could be used to model other complications which arise during diabetes, such as retinopathy, allowing for the ablation of both pancreatic β -cells and retinal cells. This system would also provide an immediate benefit to our collaborators in the Mumm lab who are trying to investigate the role resident macrophages play during retinal regeneration. Currently, they can only achieve simultaneous ablation of the macrophages

and rod cells and would benefit immensely from a multiplex cell ablation system allowing for the independent ablation of these two cell types.

7.1.2 Previous multiplex cell ablation successes in the Ackerley lab group

Figure 7.2: Chemical structures of nil-bystander nitroaromatic prodrugs.

Previous members of the Ackerley group have identified and characterised several candidate nitroreductase prodrug pairings for a multiplex cell ablation system. A lot of this work has involved the use of the Ackerley group “core” nitroreductase library consisting of 58 wild-type nitroreductase enzymes (Section 2.7; Prosser *et al.*, 2013; Rich, 2017). Nitroreductases were identified from this library (and if necessary engineered) which exhibited desired prodrug specificities. Previous PhD student, Abigail Sharrock identified pairs of nitroreductases which exhibited non-overlapping specificities with metronidazole and RB6145 (Sharrock, 2018). These nitroreductases were found to be active in both *E. coli* and HEK-293 cells, however it is yet to be seen whether these nitroreductases will be tolerated in zebrafish and whether RB6145 will prove an effective prodrug for targeted cell ablation applications.

An alternative approach was taken by previous Masters student Rory Little who aimed to identify nitroreductases that exhibited non-overlapping specificities with metronidazole and tinidazole (Little, 2015). Both metronidazole and tinidazole are commercially available and low-cost, unlike RB6145, which had to be custom-synthesised for this study. Prodrugs that are low-cost and readily available would be of benefit to this system, allowing more researchers to use this system and in a more high-throughput fashion. However, unlike metronidazole and RB6145, metronidazole and tinidazole share a 5-nitroimidazole scaffold (Figure 7.2), making it more difficult to identify nitroreductases which discriminate between these two activities. Through screening the core nitroreductase library, Rory

identified NfsB_Vv as a nitroreductase that exhibited a substantial preference for metronidazole over tinidazole. Seeking a complementary nitroreductase partner, he found NfsA innately had a slight preference for tinidazole over metronidazole, and used the 7SM *nfsA* mutagenesis library previously described in Section 3.1.2, to identify NfsA variants that exhibited further improved tinidazole selectivity. Through the use of a dual positive-negative selection involving metronidazole and niclosamide, the NfsA variant 5_B9 was isolated which exhibited enhanced tinidazole selectivity in *E. coli*, albeit, not sufficient for use in a multiplex cell ablation system. We believed the tinidazole selectivity of NfsA could be improved further, which led to the focus of the current chapter.

7.2 Aims

- Generate an error-prone PCR library of *nfsA* 5_B9 to identify NfsA variants exhibiting improved tinidazole selectivity.
- Use the Marsden library as an alternative source of NfsA variants to identify additional NfsA variants exhibiting improved tinidazole selectivity.
- Characterise NfsA variants exhibiting enhanced tinidazole selectivity in HEK-293 cells to validate their use in a eukaryote cell ablation system.
- Identify and characterise alternative nitroreductase/prodrug pairings exhibiting non-overlapping specificities with metronidazole and nitrofurazone/nitrofurantoin.

7.3 Results

7.3.1 Improving the tinidazole selectivity of NfsA in *E. coli*

7.3.1.1 Error-prone PCR library creation

The NfsA variant 5_B9 previously identified by Masters student Rory Little was used as a template in an error-prone PCR reaction. A standard PCR protocol was used to amplify *nfsA* 5_B9 and 5 ng of the resulting product was used as a template in an error-prone PCR reaction to yield a medium to high error rate (Section 2.5.2). The resulting DNA was cloned into the pUCX vector and used to transform the screening strain *E. coli* 7NT. Ten randomly-selected variants from the library were sequenced, finding an average of 1.7 amino acid changes per NfsA variant. A final library size of ~560,000 *nfsA* variants was created.

7.3.1.2 Recovery and characterisation of 5_B9 derived NfsA variants exhibiting improved tinidazole selectivity in *E. coli*

An *E. coli* 7NT strain expressing *nfsA* 5_B9 was plated on solid media containing different concentrations of niclosamide and metronidazole, to determine the highest concentration the strain could grow on. *E. coli* 7NT cells expressing *nfsA* 5_B9 were found to grow on agar containing a maximum of either 0.4 μ M niclosamide or 2 mM metronidazole, however lower concentrations were required when these compounds were used in combination. We hypothesised that plating the error-prone PCR library over higher concentrations of niclosamide would enrich for NfsA variants exhibiting improved tinidazole reduction, while simultaneously plating on higher concentrations of metronidazole would identify NfsA variants exhibiting decreased metronidazole reduction. The *nfsA* 5_B9 error-prone PCR library was plated over varying concentrations of 0.3-0.5 μ M niclosamide in combination with 200-10,000 μ M metronidazole. A total of 267 colonies which grew on the selective media were picked into the inner 60-wells of 96-well plates alongside *E. coli* 7NT strains expressing wild-type *nfsA*, *nfsA* 5_B9 and *pUCX_empty*. Strains overexpressing *nfsA* variants were challenged with 150 μ M tinidazole and 1500 μ M metronidazole in growth inhibition assays.

The ten NfsA variants that exhibited the greatest tinidazole selectivity when expressed in *E. coli* 7NT were taken forward for DNA sequencing and more sensitive IC₅₀ growth assays (following validation by re-transformation of *E. coli* 7NT cells and re-testing to ensure any changes in activity were due to mutations introduced into the *nfsA* gene and not acquired chromosomal mutations). Unfortunately, a large portion of the *nfsA* variants sequenced were found to be the starting template, *nfsA* 5_B9. Due to this, sequence data and IC₅₀ values of only the top four recovered variants is shown below (Table 7.1). *E. coli* 7NT strains expressing wild-type *nfsA*, *nfsA* 5_B9 and the lead partner nitroreductase *nfsB_Vv* were included as controls.

Variant	TNZ IC ₅₀ (μM)	MTZ IC ₅₀ (μM)	TNZ Selectively (MTZ/TNZ IC ₅₀)	Amino acid changes
NfsA 5_8	230 ± 40	5330 ± 700	23	D19V, Q100L*
NfsA 6_27	630 ± 230	10230 ± 1200	16	V70I, G187R*
NfsA 4_50	74 ± 20	1080 ± 140	15	E223D*
NfsA 5_B9	180 ± 80	2400 ± 450	13	S41Y/F42C/F83L/K222H/ S224C/R225G/F227I
NfsA 6_21	560 ± 170	7600 ± 630	14	V70I*
Wild-type NfsA	28 ± 4	51 ± 10	2	
NfsB_Vv	250 ± 60	22 ± 2	0.1	

*in addition to the seven NfsA 5_B9 amino acid substitutions.

Table 7.1: Tinidazole and metronidazole IC₅₀ values and sequences of the top four nfsA 5_B9 derived nfsA variants expressed in *E. coli* 7NT. *E. coli* cultures expressing nfsA variants were challenged for 4 hours across serial dilutions of tinidazole (0-1000 μM) or metronidazole (0-10,000 μM) including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represents the average of four biological repeats ± SD. The amino acid substitutions present in nfsA 5_B9 are listed and the additional substitutions introduced during the error-prone PCR.

Three of the recovered NfsA variants exhibited marginally improved tinidazole selectivity over NfsA 5_B9. In general, most NfsA variants had a decreased ability to reduce metronidazole, however this was accompanied by a decrease in tinidazole reduction (NfsA 5_8, 6_27 and 6_21). One variant, NfsA 4_50, exhibited improved tinidazole reduction compared to NfsA 5_B9, however also exhibited improved metronidazole reduction. Although these variants exhibited substantially improved tinidazole selectivity compared to wild-type NfsA, they would not be compatible with the partner nitroreductase (NfsB_Vv) owing to overlapping substrate concentration ranges. This meant that the concentration of tinidazole required to induce cell ablation would also induce cell ablation in cells expressing nfsB_Vv. Although not useful with the current partner enzyme, NfsB_Vv, these evolved NfsA variants may have potential in combination with an evolved form of NfsB_Vv exhibiting an improved metronidazole selectivity.

Figure 7.3: Amino acid substitutions in NfsA 5_B9 and NfsA 5_B9 derived variants. Amino acid substitutions are shown as sticks in blue for NfsA 5_B9 and in green for 5_B9 derived library variants. Only one of the active sites of the NfsA homodimer is shown with the FMN molecule shown in yellow. PDB accession number is 1f5v.

Two of the amino acid substitutions, Q100L and E223D, present in NfsA 5_8 and 4_50 respectively, are located in a close proximity to the active site of NfsA and to the NfsA 5_B9 amino acid substitutions. The proximity of these substitutions is consistent with a possible role in modulating tinidazole selectivity through modification of the active site pocket. NfsA 6_21 (V70I) and NfsA 6_27 (V70I and G187R) contain substitutions which are located further from the active site (Figure 7.3). Both NfsA 6_21 and 6_27 are relatively inactive nitroreductases as they are substantially less active with both tinidazole and metronidazole. This suggests the amino acid substitutions in these two variants may be affecting the stability of the enzyme. Distant mutations affecting protein stability have been frequently reported in the literature (Lo *et al.*, 1995; Morley and Kazlauskas, 2005).

7.3.1.3 Recovery and characterisation of NfsA variants from the Marsden library exhibiting improved tinidazole selectivity in *E. coli*

The work described in Section 7.3.1.1 and 7.3.1.2 was conducted prior to the creation of the Marsden *nfsA* mutagenesis library. It was predicted that a larger library targeting a different set of residues to the 7SM library might identify NfsA variants exhibiting improved selectivity for tinidazole. The dual positive-negative selection conducted in Section 7.3.1.2, yielded NfsA variants which had a decreased ability to reduce both tinidazole and metronidazole. To counter this, the Marsden library was plated on higher

concentrations of niclosamide (0.5-1 μ M) and lower concentrations of metronidazole (500-1000 μ M) to recover NfsA variants capable of reducing tinidazole at a level comparable to wild-type NfsA, while still losing activity with metronidazole. A total of 456 colonies which grew on these selection conditions were picked into the inner 60-wells of 96-well plates alongside *E. coli* 7NT strains expressing wild-type *nfsA* and *pUCX_empty*. Strains overexpressing *nfsA* variants were challenged with 150 μ M tinidazole and 1000 μ M metronidazole in growth inhibition assays.

The ten NfsA variants that exhibited the greatest tinidazole selectivity when expressed in *E. coli* 7NT were taken forward for DNA sequencing and more sensitive IC₅₀ growth assays (following validation by re-transformation of *E. coli* 7NT cells and re-testing). Sequence data and IC₅₀ values of the top ten tinidazole selective *nfsA* variants are shown in Table 7.2. *E. coli* 7NT strains expressing wild-type *nfsA* and the lead partner nitroreductase *nfsB_Vv* were included as controls.

Variant	TNZ IC ₅₀ (μ M)	MTZ IC ₅₀ (μ M)	TNZ Selectively (MTZ/TNZ IC ₅₀)	S41	L43	H215	T219	K222	S224	R225	F227
NfsA 5_55	67 $\pm$ 5	950 $\pm$ 330	14	Y	S	F	V	K	G	C	V
NfsA 8_13	50 $\pm$ 3	590 $\pm$ 110	12	Y	S	G	V	G	C	G	V
NfsA 8_39	34 $\pm$ 2	340 $\pm$ 30	10	Y	V	R	P	L	C	C	R
NfsA 5_45	82 $\pm$ 5	820 $\pm$ 220	10	C	V	H	T	K	S	R	F
NfsA 7_48	270 $\pm$ 22	2720 $\pm$ 430	10	Y	H	N	F	G	D	S	S
NfsA 8_22	58 $\pm$ 13	520 $\pm$ 360	9	Y	F	N	P	K	S	R	F
NfsA 7_34	93 $\pm$ 5	820 $\pm$ 110	9	Y	S	F	V	G	Y	S	Y
NfsA 7_55	220 $\pm$ 36	1880 $\pm$ 350	9	Y	S	G	V	R	F	S	L
NfsA 7_30	55 $\pm$ 5	460 $\pm$ 52	8	Y	Y	G	P	V	R	N	C
NfsA 5_46	28 $\pm$ 2	160 $\pm$ 40	6	Y	I	R	P	G	N	C	I
Wild-type NfsA	28 $\pm$ 4	51 $\pm$ 10	2								
NfsB_Vv	250 $\pm$ 60	22 $\pm$ 2	0.1								

Table 7.2: Tinidazole and metronidazole IC₅₀ values and sequences of the top ten Marsden *nfsA* variants expressed in *E. coli* 7NT. *E. coli* cultures expressing *nfsA* variants were challenged for 4 hours across serial dilutions of tinidazole (0-1000 μ M) or metronidazole (0-10,000 μ M) including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represents the average of four biological repeats $\pm$ SD. The amino acid present at each of the eight-targeted positions in the Marsden library is shown. The most prevalent amino acid at each position is shaded in grey if it makes up $\geq$ 50% of all amino acids at any given position.

The NfsA variants recovered from the Marsden library did not exhibit a substantially enhanced tinidazole selectivity compared to NfsA 5_B9 or the 5_B9 derived NfsA variants,

although were improved over wild-type NfsA (Table 7.1 and Table 7.2). The NfsA variants shown in Table 7.2 exhibited improved tinidazole activity over the variants in Table 7.1, however metronidazole activity was also higher than for those variants. This suggests we may have reached the maximum tinidazole selectivity that can be readily obtained, and any further evolution affects tinidazole and metronidazole reduction equally.

	S41	L43	H215	T219	K222	S224	R225	F227
Non-polar	0	30	30	40	60	20	10	40
Polar	10	40	20	50	0	40	80	20
Acidic	0	0	0	0	0	10	0	0
Basic	0	10	30	0	40	10	10	20
Aromatic	90	20	20	10	0	20	0	20

Table 7.3: Percentage of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of the top 10 NfsA variants. The property of the wild-type amino acid at each position is boxed in red.

Clear amino acid preferences were evident at positions 41 and 225 with a preference for aromatic and polar amino acids respectively (Table 7.3). These preferences have been noted throughout this thesis in NfsA variants exhibiting improved reduction with a broad range of substrates. One difference here is the presence of non-polar or polar residues at position 219. Aromatic residues have been frequently observed at this position in NfsA variants exhibiting improved reduction of metronidazole (Table 5.2 and Table 6.1). This suggests the absence of an aromatic residue at this position may contribute to diminished metronidazole reduction, improving tinidazole selectivity. This observation was also made in Chapter 5, in NfsA variants recovered from a dual positive-negative selection including metronidazole or RB6145 (Table 5.5).

7.3.2 Expression and characterisation of *nfsA* variants exhibiting tinidazole selectivity in HEK-293 cells.

7.3.2.1 Assessing fluoroprobe activation of NfsA variants exhibiting tinidazole selectivity in *E. coli* 7NT cells.

The ultimate goal of this research is to express the top *nfsA* variants in zebrafish. Although the NfsA variants shown in Table 7.1 and Table 7.2 did not exhibit the desired level of tinidazole selectivity in *E. coli*, it was nevertheless possible that they might have potential

in combination with an evolved form of the partner enzyme NfsB_Vv in eukaryotic cells. Prior to generating transfected human cell lines, the abilities of *E. coli* 7NT strains expressing the top tinidazole selective *nfsA* variants to activate the nitro-quenched fluorophore FSL61 were evaluated as per Section 2.8.3. This was necessary to establish whether their functional expression in HEK-293 cells could be monitored using FSL61 or whether evolution had impaired the ability of each NfsA variant to activate this fluorophore.

Figure 7.4: Activation of FSL61 by *E. coli* 7NT strains expressing the top *nfsA* variants exhibiting tinidazole selectivity. Fluorescence of *E. coli* 7NT strains expressing *nfsA* variants, wild-type *nfsA* (positive control) or *pUCX_empty* (negative control) was determined following a 1 hour incubation with 20 μ M FSL61. Fluorescence was quantified on a fluorescent plate reader (ex/em 350/460 nm) and is reported as relative fluorescence units (RFU). Data representative of two biological repeats $\pm$ SD.

All *E. coli* 7NT strains exhibited some level of fluorescence, greater than *E. coli* 7NT cells expressing *pUCX_empty*. However, the *E. coli* 7NT strains expressing *nfsA* 5_B9, 5_8, 6_21, 6_27 and 5_55 exhibited substantially lower fluorescence than the *E. coli* 7NT strain expressing wild-type *nfsA*, suggesting evolution has greatly impaired the ability of these variants to activate FSL61 (Figure 7.4). *E. coli* 7NT strains expressing these variants were also tested with the fluorophores FSL41 and FSL76 to determine if greater activity could be observed with an alternative fluorophore. All *E. coli* strains tested exhibited a substantially decreased ability to activate FSL41 and FSL76 compared to the *E. coli* 7NT strain expressing wild-type *nfsA* (data not shown), even lower than was seen with FSL61. FSL61 was selected as the fluorophore to monitor functionality and nitroreductase expression in HEK-293 cells.

7.3.2.2 Generation of HEK-293 cell lines transiently expressing *nfsA* variants

HEK-293 cell lines transiently expressing the top four tinidazole selective *nfsA* variants from both the *nfsA* 5_8 error-prone PCR and Marsden libraries were generated to gain an initial indication of whether expression would be tolerated. F279-V5 constructs were generated containing the genes encoding *nfsA* 5_8, 6_27, 4_50, 6_21, 8_39, 8_13, 5_55 and 5_45. A F279-V5 construct was also generated containing the gene encoding *nfsA* 5_B9, as this variant had not previously been tested in a human cell line. Forty-eight hours following transfection, HEK-293 cell lines were tested for transient nitroreductase expression using the nitro-quenched fluoroprobe FSL61 (Section 2.9.5.1). HEK-293 cells transiently expressing *nfsA* 4_50, 8_39, 8_13, 5_55 and 5_45 exhibited varying levels of fluorescence, while HEK-293 cells transiently expressing *nfsA* 5_B9, 5_8, 6_21 and 6_27 did not exhibit any detectable fluorescence (images not shown). This mirrored the poor activation of the FSL61 fluoroprobe by these NfsA variants seen in *E. coli* 7NT (Figure 7.4).

7.3.2.3 Generation of HEK-293 stably expressing *nfsA* variants

Stably expressing HEK-293 cell lines were generated for all *nfsA* variants described in Section 7.3.2.2. To assess the level of nitroreductase expression, stable cell lines were tested with the nitro-quenched fluoroprobe FSL61 and assessed by both fluorescence microscopy and flow cytometry. Western blots were also generated as a method to assess nitroreductase expression independent of the ability to activate FSL61.

Figure 7.5: Visualisation of HEK-293 cell lines stably expressing *nfsA* variants using FSL61. HEK-293 cell lines stably expressing *nfsA* variants were incubated with 50 μ M FSL61 fluoroprobe for 1 hour before visualisation on a fluorescence microscope using a DAPI filter. Cells lines were compared to wild-type HEK-293 cells to determine the background level of cell auto-fluorescence. Exposure times were kept consistent between different samples.

Extremely faint patches of fluorescence were evident in the HEK-293 cell lines expressing *nfsA* 5_45, 5_55, 8_13 and 8_39 (Figure 7.5), but at a substantially lower level than was evident in the transiently expressing cell lines. This suggests expression of these nitroreductases is not tolerated long-term and has shut down during the passaging steps required to generate the stable cell lines. No fluorescence was evident in HEK-293 cell lines expressing *nfsA* 5_B9, 4_50, 5_8, 6_21 and 6_27 (Figure 7.5). This is not surprising as most of these cells lines exhibited no fluorescence when transiently expressing the nitroreductase and exhibited substantially less activity with the fluoroprobe when expressed in *E. coli* 7NT (Figure 7.4).

The flow cytometry protocol optimised in Chapter 6 (Section 6.3.5.3.1) was used to quantify *nfsA* expression to assess if any faint FSL61 activation could be detected using a more sensitive method. HEK-293 cells expressing wild-type *nfsA* and wild-type HEK-293 cells were included as controls.

Figure 7.6: Flow cytometry of HEK-293 cells stably expressing the top tinidazole selective *nfsA* variants. Cell lines were incubated with 5 μ M FSL61 for 1 hour. Following the incubation, the cells were analysed by flow cytometry and compared to a non-transgenic population (red) to determine the background level of cell auto-fluorescence. The mean fluorescence for each cell line is presented. Image generated using the FlowJo software.

Consistent with the fluorescence microscopy results (Figure 7.5), a small fluorescent shift was evident in the HEK-293 cell lines expressing *nfsA* 5_45, 5_55, 8_13 and 8_39, while no fluorescent shift was observed in HEK-293 cell lines expressing *nfsA* 5_B9, 4_50, 5_8, 6_21 and 6_27 (Figure 7.6). As an additional means to assess the level of *nfsA* expression in HEK-293 cell lines, western blots were generated to determine whether low nitroreductase expression levels were contributing to the poor FSL61 activity (Figure 7.7). All of the HEK-293 cell lines tested exhibited varying levels of *nfsA* expression, although all substantially lower than HEK-293 cells expressing wild-type *nfsA* (Figure 7.7). *NfsA* 5_B9 and 5_B9 derived variants exhibited moderate expression levels in HEK-293 cells, suggesting the lack of fluorescence seen in Figure 7.5 and Figure 7.6 was likely due to poor FSL61 activation by these variants as well as diminished expression relative to wild-type *nfsA*.

Figure 7.7: Western blot depicting *nfsA* variant expression in HEK-293 cells. HEK-293 cells lines expressing *nfsA* variants with enhanced tinidazole selectivity recovered from (A) the 5_B9 error prone PCR library and (B) the Marsden library. HEK-293 cell samples were lysed with RIPA buffer and probed for NfsA using a NfsA polyclonal primary antibody raised in sheep and a donkey anti-sheep IgG HRP conjugate secondary antibody. The expression level of alpha-tubulin in samples was used as a loading control and was detected via immunofluorescence. Note: Image A used a new secondary donkey anti-sheep IgG HRP conjugate antibody which was purchased after the image B was created.

7.3.2.4 Characterisation of HEK-293 stably expressing *nfsA* variants exhibiting enhanced tinidazole selectivity.

HEK-293 cell lines stably expressing *nfsA* variants were challenged with tinidazole and metronidazole and tested in IC₅₀ assays to determine if any nitroreductase activity was detectable *in vivo* (Table 7.4). HEK-293 cells expressing wild-type *nfsA* and the partner enzyme *nfsB*_Vv were included as controls.

Variant	TNZ IC ₅₀ (μM)	Variant	MTZ IC ₅₀ (μM)
NfsA 5_B9	>1000	NfsA 5_B9	>5000
NfsA 4_50	>1000	NfsA 4_50	>5000
NfsA 5_8	>1000	NfsA 5_8	>5000
NfsA 6_21	>1000	NfsA 6_21	>5000
NfsA 6_27	>1000	NfsA 6_27	>5000
NfsA 8_13	>1000	NfsA 8_13	>5000
NfsA 8_39	>1000	NfsA 8_39	>5000
NfsA 5_45	220 ± 32	NfsA 5_45	980 ± 160
NfsA 5_55	620 ± 160	NfsA 5_55	2590 ± 530
Wild-type NfsA	46 ± 12	Wild-type NfsA	51 ± 13
NfsB_Vv	560 ± 43	NfsB_Vv	32 ± 11
Wild-type HEK-293	>1000	Wild-type HEK-293	>5000

Table 7.4: Tinidazole and metronidazole IC₅₀ values of HEK-293 cell lines stably expressing *nfsA* variants. Replicates of HEK-293 cells expressing *nfsA* variants were grown in 96-well plates and challenged with varying concentrations of tinidazole (0-1000 μM) or metronidazole (0-5000 μM) for 48 hours. MTS reagent was added to the wells and incubated for an additional 2-3 hours. The absorbance of unchallenged wells were measured at 490 nm and were compared to that of challenged wells to determine the percentage viability at each prodrug concentration. The IC₅₀ value was calculated using GraphPad 8.0 software. Data represents the average of three biological replicates ± SD.

Most of the HEK-293 cell lines expressing *nfsA* variants exhibited no activity with tinidazole or metronidazole. This is likely due to the poor expression of these *nfsA* variants in HEK-293 cells (Figure 7.7). HEK-293 cell lines expressing *nfsA* 5_45 and 5_55 exhibited some activity with both tinidazole and metronidazole with selectivity towards tinidazole. Unfortunately, the tinidazole IC₅₀ value of HEK-293 cells expressing either *nfsA* 5_45 or 5_55 would not be low enough to work with the partner nitroreductase, *nfsB_Vv*, as the concentration of tinidazole required to induce cell ablation would also induce cell ablation in cells expressing *nfsB_Vv* (Table 7.4). This low activity is likely the result of poor expression of both *nfsA* 5_45 and 5_55 in HEK-293 cells (Figure 7.7).

7.3.3 Investigating alternative multiplex cell ablation prodrug combinations in *E. coli*

To identify nitroreductases exhibiting a greater prodrug selectivity, we turned to the use of structurally distinct prodrugs which we anticipated would yield better results. We chose to focus on nitrofurazone, nitrofurantoin and metronidazole due to their widespread availability and low-cost. The Ackerley group core nitroreductase library was screened against nitrofurazone, nitrofurantoin and metronidazole to identify any wild-type nitroreductases which exhibited preferential activation of metronidazole or

nitrofurazone/nitrofurantoin while excluding the alternate prodrug. Most nitroreductases exhibited a preference for nitrofurazone and nitrofurantoin over metronidazole, with NfsA_Ls exhibiting the greatest nitrofurazone and nitrofurantoin selectivity. However, no nitroreductases in the core library exhibited the opposing prodrug selectivity. NfsA_Ec exhibited the greatest nitrofurazone/nitrofurantoin “selectivity”, as it was able to reduce nitrofurazone, nitrofurantoin and metronidazole at a comparable level, compared to other nitroreductases in the core library. The IC₅₀ values of *E. coli* 7NT cells expressing *nfsA_Ls* or *nfsA_Ec* is shown in Table 7.5. Based on these results we selected NfsA_Ec as a starting point to identify nitroreductase variants exhibiting preferential metronidazole activation over nitrofurazone or nitrofurantoin.

	NFZ IC ₅₀ (μM)	NFN IC ₅₀ (μM)	MTZ IC ₅₀ (μM)
NfsA_Ls	26 ± 8	21 ± 2	9390 ± 2200
NfsA_Ec	41 ± 8	25 ± 7	61 ± 13

Table 7.5: Nitrofurazone, nitrofurantoin and metronidazole IC₅₀ values of *E. coli* 7NT strains expressing *nfsA_Ls* or *nfsA_Ec*. *E. coli* cultures expressing *nfsA_Ls* or *nfsA_Ec* were challenged for 4 hours across serial dilutions of nitrofurazone (0-1000 μM), nitrofurantoin (0-500 μM) or metronidazole (0-10,000 μM) including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ of cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represents the average of four biological repeats ± SD.

7.3.3.1 Identification and characterisation of NfsA variants exhibiting preferential reduction of metronidazole over nitrofurazone or nitrofurantoin

A dual positive-negative selection was used to identify NfsA variants exhibiting preferential reduction of metronidazole over nitrofurazone or nitrofurantoin. As in the previous dual positive-negative selections, niclosamide was included to ensure some level of generic nitroreductase activity was retained, in combination with nitrofurazone or nitrofurantoin to identify NfsA variants which had lost activity with either of these prodrugs. Following empirical testing, the Marsden library was plated on agar containing various combinations of 0.3-0.5 μM niclosamide and 10-16 μM nitrofurazone or 15-16 μM nitrofurantoin (concentrations slightly higher than *E. coli* 7NT cells expressing wild-type *nfsA* could grow on). Colonies which grew in these selection conditions were picked into the inner 60-wells of 96-well plates alongside *E. coli* 7NT strains expressing wild-type *nfsA* and *pUCX_empty*. A total of 798 colonies were picked from each selection (1596 colonies

total). Strains overexpressing *nfsA* variants were challenged with 75 μ M metronidazole and 40 μ M nitrofurazone or 35 μ M nitrofurantoin in growth inhibition assays (reflecting that the prodrug forms of nitrofurazone and nitrofurantoin are inherently more toxic than metronidazole, and hence lower concentrations of these compounds are required for cell ablation).

The ten *NfsA* variants that exhibited the greatest activity with metronidazole while excluding nitrofurazone or nitrofurantoin, when expressed in *E. coli* 7NT, were taken forward for DNA sequencing and more sensitive IC₅₀ growth assays (following validation by re-transformation of *E. coli* 7NT cells and re-testing). Sequence data and IC₅₀ values for these 20 *nfsA* variants is shown in Table 7.6. *E. coli* 7NT strains expressing *pUCX_empty*, wild-type *nfsA_Ec* and the lead partner nitroreductase *nfsA_Ls* were included as controls.

Variant	NFZ IC ₅₀ (μM)	MTZ IC ₅₀ (μM)	MTZ Selectivity (NFZ IC ₅₀ / MTZ IC ₅₀)
NfsA 12_2	26 ± 4	17 ± 3	1.5
NfsA 3_4	44 ± 4	96 ± 15	0.46
NfsA 5_8	37 ± 6	206 ± 43	0.18
NfsA 6_40	24 ± 5	140 ± 25	0.17
NfsA 2_54	63 ± 9	390 ± 140	0.16
NfsA 6_16	170 ± 46	1470 ± 490	0.12
NfsA 3_51	120 ± 46	1100 ± 95	0.11
NfsA 2_11	230 ± 54	2220 ± 510	0.10
NfsA 1_4	190 ± 65	2370 ± 480	0.08
NfsA 5_48	39 ± 12	1800 ± 760	0.08
NfsA_Ec	41 ± 8	61 ± 14	0.68
NfsA_Ls	26 ± 8	9390 ± 2200	0.003
pUCX_empty	390 ± 130	>10000	

S41	L43	H215	T219	K222	S224	R225	F227	Others
C	L	F	Y	P	V	F	D	-
R	V	N	F	A	L	Y	H	-
Y	C	D	V	G	R	V	C	-
R	F	L	F	R	V	Y	H	-
Y	V	G	F	A	V	C	I	G65C
F	L	L	I	L	G	N	F	-
R	I	I	F	V	G	Y	H	Q67K
Y	C	C	V	V	N	I	Y	G65C
C	C	L	Y	M	C	G	G	G65C
C	C	R	F	L	L	H	G	-

Variant	NFN IC ₅₀ (μM)	MTZ IC ₅₀ (μM)	MTZ Selectivity (NFZ IC ₅₀ / MTZ IC ₅₀)
NfsA 20_3	23 ± 3	6 ± 1	3.8
NfsA 16_13	32 ± 4	20 ± 4	1.6
NfsA 18_58	25 ± 3	19 ± 3	1.3
NfsA 18_45	34 ± 7	45 ± 6	0.8
NfsA 22_28	63 ± 9	390 ± 140	0.6
NfsA_Ec	25 ± 7	61 ± 14	0.4
NfsA_Ls	21 ± 2	9390 ± 2200	0.002
pUCX_empty	66 ± 16	>10000	

S41	L43	H215	T219	K222	S224	R225	F227	Others
G	F	N	I	L	G	F	D	-
N	L	H	L	A	I	R	D	-
G	F	L	N	L	G	V	H	-
R	L	V	F	V	R	G	R	-
H	F	S	Y	P	H	V	L	-

Table 7.6: Nitrofurazone, nitrofurantoin and metronidazole IC₅₀ values and sequences of the top nfsA variants expressed in *E. coli* 7NT. *E. coli* cultures expressing nfsA variants were challenged for 4 hours across serial dilutions of nitrofurazone (0-1000 μM), nitrofurantoin (0-500 μM) or metronidazole (0-10,000 μM) including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represent the average of four biological repeats ± SD. The amino acid present at each of the eight-targeted positions in the Marsden library is shown together with any additional non-targeted substitutions identified. The most prevalent amino acid at each position is shaded in grey if it makes up ≥ 50% of all amino acids at any given position.

Although some NfsA variants were recovered exhibiting preferential metronidazole activation, these variants still had a high level of activity with nitrofurazone or nitrofurantoin, and exhibited an overlapping substrate concentration range with the partner nitroreductase, NfsA_Ls (Table 7.6). Any NfsA variants which did exhibit decreased activity with nitrofurazone or nitrofurantoin also exhibited decreased activity with

metronidazole (Table 7.6). It was concluded that these NfsA variants could not be applied in a multiplex cell ablation system and they were discontinued.

7.3.3.2 Identification and characterisation of NfsA variants exhibiting nitrofurazone selectivity in *E. coli*

As a final attempt, we sought to identify NfsA variants exhibiting preferential nitrofurazone activation, while excluding metronidazole. Although several wild-type nitroreductases naturally exhibit this activity profile, we wanted to investigate the amenability of NfsA for engineering to modify its innate preferences, and confirm that the dual positive-negative selection is a valid approach to identify nitroreductases with biotechnological applications. In Chapter 5, a dual positive-negative selection was conducted using 1 μ M niclosamide and 500 μ M metronidazole. NfsA variants recovered from this selection appeared to have an increased ability to reduce nitrofurazone, but a decreased ability to reduce metronidazole (Figure 5.3). The Marsden library was plated on the same selection conditions. A total of 228 colonies which grew under this selection condition were picked into the inner 60 wells of 96-well plates alongside *E. coli* 7NT cells expressing wild-type *nfsA* or *pUCX_empty*. Strains overexpressing *nfsA* variants were challenged with 30 μ M nitrofurazone and 1000 μ M metronidazole in growth inhibition assays.

The top ten NfsA variants that exhibited that greatest nitrofurazone selectivity were taken forward for DNA sequencing and more sensitive IC₅₀ growth assays (following validation by re-transformation of *E. coli* 7NT cells and re-testing). Sequence data and IC₅₀ values of *E. coli* 7NT cells expressing the top ten nitrofurazone selective *nfsA* variants are shown in Table 7.7.

Variant	MTZ IC ₅₀ (μM)	NFZ IC ₅₀ (μM)	NFZ Selectivity (MTZ IC ₅₀ /NFZ IC ₅₀)	S41	L43	H215	T219	K222	S224	R225	F227
NfsA 4_14	4290 ± 280	35 ± 4	123	Y	H	G	I	V	R	D	V
NfsA 4_3	3280 ± 120	36 ± 8	91	F	S	C	A	A	R	S	L
NfsA 1_18	3550 ± 240	41 ± 5	87	C	C	V	H	R	R	Y	H
NfsA 4_23	3430 ± 400	40 ± 13	86	Y	Y	V	A	G	Y	V	Y
NfsA 1_51	3080 ± 270	36 ± 9	86	C	F	H	T	K	S	R	F
NfsA 4_4	2660 ± 320	32 ± 6	83	F	S	S	S	T	R	Y	F
NfsA 4_43	3390 ± 960	44 ± 10	77	Y	H	V	A	A	S	V	F
NfsA 1_55	2880 ± 540	39 ± 5	74	Y	G	F	L	G	F	S	C
NfsA 1_11	2370 ± 200	33 ± 1	72	C	R	H	V	R	I	D	D
NfsA 1_14	5080 ± 980	80 ± 10	64	F	C	R	V	Q	G	R	I
Wild-type NfsA	61 ± 9	45 ± 5	1.4								
pUCX_empty	>10,000	390 ± 130									

Table 7.7: Metronidazole and nitrofurazone IC₅₀ values and sequences of the top nitrofurazone selective nfsA variants expressed in *E. coli* 7NT. *E. coli* cultures expressing nfsA variants were challenged for 4 hours across serial dilutions of metronidazole (0-10,000 μM) and nitrofurazone (0-400 μM) including an unchallenged control. Growth inhibition was determined by measuring the OD₆₀₀ cultures pre- and post- the 4-hour challenge time. Percentage growth relative to the unchallenged control was calculated for each drug concentration. GraphPad 8.0 software was used to calculate the IC₅₀ value. Data represents the average of four biological repeats ± SD. The amino acid present at each of the eight-targeted positions in the Marsden library is shown and any additional mutations. No amino acid at any position made up ≥ 50% of all amino acids.

All NfsA variants exhibited substantial selectivity for nitrofurazone. In most cases, the ability to reduce nitrofurazone was improved over wild-type NfsA or retained at a similar level, while activity with metronidazole was substantially reduced (39-83 times higher IC₅₀ value; Table 7.7). NfsA 4_14 exhibited the greatest nitrofurazone selectivity, with a selectivity ratio of 123. This variant sensitised *E. coli* 7NT cells to a 1.3-fold lower concentration of nitrofurazone, but a 70-fold higher concentration of metronidazole, compared to wild-type NfsA.

	S41	L43	H215	T219	K222	S224	R225	F227
Non-polar	0	10	40	70	50	20	20	30
Polar	30	40	20	20	20	20	20	10
Acidic	0	0	0	0	0	0	20	10
Basic	0	30	30	10	30	40	20	10
Aromatic	70	20	10	0	0	20	20	40

Table 7.8: Percentage of the five amino acid side-chain categories at the eight-targeted positions in the Marsden library of the top ten nitrofurazone selective NfsA variants. The property of the wild-type amino acid at each position is boxed in red.

Despite the small sequence sample size ($n = 10$) some clear sequence preferences were evident. Like in previous chapters, there was a clear preference for an aromatic residue at position 41 (Table 7.8). This substitution at position 41 may be involved in improving the overall generic activity of NfsA, as it is evident in NfsA variants exhibiting improved reduction of a range of substrates. In contrast to other chapters, there is no preference for an aromatic residue at position 219, with 70% of NfsA variants containing a non-polar residue. There is also no preference for polar residues at position 225, with an equal preference for all residue types, including basic amino acids, which has not been seen in previous chapters.

7.3.4 Expression and characterisation of *nfsA* variants exhibiting nitrofurazone selectivity in HEK-293 cells.

7.3.4.1 Assessing fluoroprobe activation of NfsA variants exhibiting preferential nitrofurazone activation in *E. coli* 7NT cells.

F279-V5 constructs containing the genes encoding the top four NfsA variants exhibiting nitrofurazone selectivity were created. Prior to generating transfected human cell lines, the abilities of *E. coli* 7NT strains expressing the top nitrofurazone selective *nfsA* variants to activate the nitro-quenched fluoroprobe FSL61 were evaluated as per Section 2.8.3. This was necessary to establish whether their functional expression could be monitored using FSL61 or whether evolution had impaired the ability of each NfsA variant to activate this fluoroprobe. In brief, all *E. coli* 7NT strains expressing *nfsA* variants exhibited substantially lower fluorescence than the *E. coli* 7NT strain expressing wild-type *nfsA* (Figure 7.8). Based on results previous obtained (Section 7.3.2), it was considered FSL61 activation by these variants may be harder to detect in HEK-293 cells, as seen in HEK-293 cells expressing some of the tinidazole selective *nfsA* variants.

Figure 7.8: Activation of FSL61 by *E. coli* 7NT strains expressing the top *nfsA* variants exhibiting nitrofurazone selectivity. Fluorescence of *E. coli* 7NT strains expressing *nfsA* variants, wild-type *nfsA* (positive control) or pUCX_empty (negative control) was determined following a 1 hour incubation with 20 μ M FSL61. Fluorescence was quantified on a fluorescent plate reader (ex/em 350/460 nm) and is reported as relative fluorescence units (RFU). Data representative of two technical repeats $\pm$ SD.

7.3.4.2 Generation of HEK-293 cell lines transiently expressing *nfsA* variants

HEK-293 cell lines were generated, transiently expressing *nfsA* 1_18, 4_14, 4_23 and 4_3 as per Section 2.9.4. Forty-eight hours following transfection, the HEK-293 cell lines were tested for transient nitroreductase expression using the nitro-quenched fluorophore FSL61 (Section 2.9.5.1). HEK-293 cells transiently expressing *nfsA* 4_14, 4_3 and 4_23 exhibited weak fluorescence, while HEK-293 cells expressing *nfsA* 1_18 exhibited no detectable fluorescence (images not shown). This weak fluorescence was likely due to poor activation of the FSL61 fluorophore by the NfsA variants as seen in bacteria (Figure 7.8).

7.3.4.3 Generation of HEK-293 stably expressing *nfsA* variants

HEK-293 cell lines were generated stably expressing the *nfsA* variants described in Section 7.3.4.2. To assess the level of nitroreductase expression, stable cell lines were tested with the nitro-quenched fluorophore FSL61 and assessed by both fluorescence microscopy and flow cytometry. Western blots were also generated as a method to assess nitroreductase expression independent of the ability to activate the fluorophore FSL61.

Figure 7.9: Visualisation of HEK-293 cell lines stably expressing *nfsA* variants using the nitro-quenched fluorophore FSL61. HEK-293 cell lines stably expressing *nfsA* variants were incubated with 50 μ M FSL61 fluoroprobe for 1 hour before visualisation on a fluorescence microscope using a DAPI filter. Cell lines were compared to wild-type HEK-293 cells to determine the background level of cell auto-fluorescence. Exposure times were kept consistent between different samples.

Faint fluorescence was evident in the HEK-293 cell lines expressing *nfsA* 4_14, 4_3 and 4_23, but at a substantially lower level than HEK-293 cells expressing wild-type *nfsA* (Figure 7.9). No fluorescence was evident in HEK-293 cells expressing *nfsA* 1_18, consistent with that seen in the transient fluorescence test. Flow cytometry was used to assess if any faint FSL61 activation could be detected using a more sensitive method. HEK-293 cell lines expressing *nfsA* variants were incubated with 5 μ M FSL61 for 1 hour, followed by analysis on the flow cytometer (Figure 7.10). Wild-type HEK-293 cells and HEK-293 cells expressing wild-type *nfsA* were included as controls.

Figure 7.10: Flow cytometry of HEK-293 cells stably expressing the top nitrofurazone selective *nfsA* variants. Cell lines were incubated with 5 μ M FSL61 for 1 hour. Following the incubation, the cells were analysed by flow cytometry and compared to a non-transgenic population (red) to determine the background level of cell auto-fluorescence. The mean fluorescence for each cell line is presented. Image generated using the FlowJo software.

A fluorescent shift was evident in the HEK-293 cell lines expressing *nfsA* 4_14, 4_3 and 4_23, while no fluorescent shift was observed in HEK-293 cell lines expressing *nfsA* 1_18 (Figure 7.10). To confirm expression, independent of fluoroprobe activation, western blots were generated (Figure 7.11). Varying levels of expression were evident in the different HEK-293 cells lines, consistent with the expression levels inferred by flow cytometry (Figure 7.10).

Figure 7.11: Western blot depicting *nfsA* variant expression in HEK-293 cells. HEK-293 cell samples were lysed with RIPA buffer and probed for NfsA using a NfsA polyclonal primary antibody raised in sheep and a donkey anti-sheep IgG HRP conjugate secondary antibody. The expression level of alpha-tubulin in samples was used as a loading control and detected via immunofluorescence.

7.3.4.4 Characterisation of HEK-293 cells stably expressing *nfsA* variants exhibiting enhanced nitrofurazone selectivity

HEK-293 cell lines stably expressing *nfsA* variants exhibiting enhanced nitrofurazone selectivity were challenged with nitrofurazone and metronidazole in IC₅₀ assays to determine if any nitroreductase activity was detectable *in vivo*. Wild-type HEK-293 cells were included as a control to determine the toxicity of nitrofurazone in human cells. HEK-293 cells expressing wild-type *nfsA* were also included as a control.

Variant	NFZ IC ₅₀ (μM)	MTZ IC ₅₀ (μM)	NFZ Selectivity (MTZ IC ₅₀ /NFZ IC ₅₀)
NfsA 4_23	3.7 ± 1	3860 ± 1470	1043
NfsA 4_14	2.6 ± 0.7	1400 ± 330	538
NfsA 4_3	14 ± 4	5070 ± 1480	362
NfsA 1_18	65 ± 7	8600 ± 1940	132
Wild-type NfsA	1.7 ± 0.1	60 ± 15	35
Wild-type HEK-293	80 ± 0.7	7250 ± 450	90

Table 7.9: Nitrofurazone and metronidazole IC₅₀ values of HEK-293 cell lines stably expressing *nfsA* variants. Replicates of HEK-293 cells expressing *nfsA* variants were grown in 96-well plates and challenged with varying concentrations of nitrofurazone (0-250 μM) or metronidazole (0-10,000 μM) for 48 hours. MTS reagent was added to the wells and incubated for an additional 2-3 hours. The absorbance of unchallenged wells were measured at 490 nm and were compared to that of challenged wells to determine the percentage viability at each prodrug concentration. The IC₅₀ value was calculated using GraphPad 8.0 software. Data represents the average of four biological replicates ± SD.

Three out of the four NfsA variants exhibited improved nitrofurazone selectivity over HEK-293 cells expressing wild-type *nfsA*, with activity levels consistent with the expression levels observed in Figure 7.11. HEK29-3 cells expressing *nfsA* 4_23 and 4_14 performed the best, with substantially higher metronidazole IC₅₀ values than HEK-293 cells expressing *wild-type nfsA* (23 to 64-fold higher), while retaining a comparable nitrofurazone IC₅₀ value (Table 7.9). HEK-293 cells expressing *nfsA* 1_18 did not exhibit activity with nitrofurazone or metronidazole. These IC₅₀ results combined with previous data indicate this variant is not tolerated in HEK-293 cells.

7.4 Discussion

The research reported in this chapter describes several attempts to identify NfsA variants for use in a multiplex cell ablation system. For this system to function, pairs of

nitroreductases are required with non-overlapping prodrug specificities. Several wild-type nitroreductases were identified which exhibited an innate selectivity with a particular prodrug. We wanted to identify and characterise NfsA variants that could be paired with these wild-type nitroreductases.

Initially, work was continued from previous Masters student Rory Little, who identified NfsA 5_B9, which exhibited a partial preference for tinidazole over metronidazole. An error-prone PCR library of *nfsA* 5_B9 was generated in an attempt to improve tinidazole selectivity. Only one candidate, NfsA 5_8, exhibited a substantial improvement in tinidazole selectivity, and most of the NfsA variants exhibited a decreased ability to reduce both tinidazole and metronidazole. In addition, the Marsden library was also screened using a dual positive-negative selection. Despite this being a much larger source of NfsA variants, all variants recovered exhibited a comparable tinidazole selectivity to NfsA 5_B9 and 5_B9 derived variants. Any desirable activity that was observed in *E. coli* did not translate into HEK-293 cell lines.

These results suggest we have reached a threshold as to how readily we can distinguish activities with these two compounds. Both metronidazole and tinidazole share a 5-nitroimidazole scaffold, therefore reduction of these compounds is likely promoted by many of the same residues, making it hard to abolish metronidazole reduction without impacting tinidazole reduction. The sequence data presented in Chapter 6 supports this theory. In Chapter 6, NfsA variants exhibiting improved metronidazole reduction (Table 6.1) had strikingly similar amino acid preferences to NfsA variants exhibiting improved tinidazole reduction (Table 6.2). These same amino acid substitutions (aromatic residues at S41 and polar residues at R225) were evident in NfsA variants selected from the Marsden library exhibiting improved tinidazole reduction (Table 7.3). This suggests why substantial selectivity could not readily be achieved, as these substitutions have been previously shown to also be important for metronidazole reduction. Position 219 was the only position which exhibited different amino acid substitutions to those seen in generically improved NfsA variants, with tinidazole selective NfsA variants showing a preference for non-polar and polar residues, compared to the frequently observed aromatic residues. The tinidazole selectivity that was observed in Marsden library derived NfsA variants may be attributed to this substitution.

We then turned to the use of more structurally distinct prodrug pairings, identifying NfsA variants that exhibited selectivity with either nitrofurazone, nitrofurantoin or metronidazole while excluding the other prodrug. Plating the Marsden library on a dual positive-negative selection was unsuccessful in recovering any NfsA variants that exhibited metronidazole selectivity while excluding nitrofurazone or nitrofurantoin (Table 7.6). In contrast, many NfsA variants were recovered that exhibited selectivity with nitrofurazone while excluding metronidazole (Table 7.7). This activity was also found to translate into HEK-293 cells, with three out of the four NfsA variants tested exhibiting a high selectivity for nitrofurazone over metronidazole.

There are likely several reasons why more NfsA variants were recovered exhibiting improved selectivity with nitrofurazone, compared to other prodrugs. Nitrofurazone has a much higher $E^{\prime}_7$ value (-257 mV) compared to metronidazole (-486 mV), indicating nitrofurazone is more readily reduced than metronidazole. This suggests it may be difficult to identify or evolve a nitroreductase which is unable to reduce the highly electron accepting nitrofurazone, while still reducing metronidazole, which is less willing to accept electrons. This may explain why NfsA variants were easily recovered that exhibited nitrofurazone selectivity and not metronidazole selectivity, despite screening ~7-fold more colonies selected under metronidazole selective conditions. Sequence information also provided some insight into how nitrofurazone selectivity could evolve. NfsA variants exhibiting nitrofurazone selectivity frequently only had one of the three amino acid preferences commonly seen in evolved NfsA variants throughout this study. This was a preference for an aromatic residue at position 41, however a preference for aromatic residues at position 219 or polar residues at position 225 was not evident. At position 219, there was a preference for non-polar residues, a preference which was also seen in tinidazole selective NfsA variants. This supports the theory that aromatic residues at this position are important for the reduction of metronidazole. At position 225 there was no preference for any amino acid. In Chapter 4, substitutions at position 225 were shown to play a crucial role in the reduction of several substrates including chloramphenicol, metronidazole, CB1954 and 2,4-DNT but not nitrofurazone. This suggests substitutions at this position might also be used to exclude metronidazole, while not affecting the ability to reduce nitrofurazone.

Although we were able to successfully evolve NfsA variants exhibiting nitrofurazone selectivity, at present we see no application for these variants in isolation, and we will need to develop a nitroreductase exhibiting the opposing prodrug selectivity for multiplex applications. Currently no wild-type nitroreductase in the Ackerley group core collection has been found to exhibit this selectivity, nor were we able to evolve NfsA to exhibit the opposing selectivity. We may need to look beyond the scope of the current core nitroreductase library and identify new nitroreductases which exhibit different prodrug selectivity's. This is an ongoing goal in the Ackerley group, with several members using metagenomic discovery to identify new nitroreductases from uncultivated bacteria.

Chapter 8 Key findings, conclusions and future directions

8.1 Research motivation

The research described in this thesis involved the engineering of NfsA for both a fundamental application, to improve our understanding of directed evolution, and a biomedical application, to identify NfsA variants for use in a targeted cell ablation system in zebrafish or other animal models.

Conventional directed evolution techniques involve the successive accumulation of single amino acid changes, providing incremental improvements in enzyme fitness. However, directed evolution rarely yields enzymes that are as active with alternative substrate(s) as they originally were with the natural substrate(s). We wanted to address whether our best-perceived practice of directed evolution is the best. We investigated whether NfsA variants discovered from a simultaneous site-directed mutagenesis approach could have also been discovered using a conventional stepwise mutagenesis approach. We also wanted to investigate how epistasis and other phenomena such as metabolic inference can affect the directed evolution of an enzyme, as evolved enzymes often do not perform as anticipated when introduced into an existing metabolic system.

We next sought to leverage the simultaneous site-directed mutagenesis library we had developed for our fundamental study for an applied study to identify nitroreductases for use in a targeted cell ablation system in zebrafish. Existing nitroreductase cell ablation systems in zebrafish are limited by the relatively poor activation of metronidazole by NfsB_Ec. In order to achieve sufficient cell ablation, high concentrations of metronidazole are required for long exposure times, causing confounding off-target effects, or even death. Although a triple mutant (NfsB_Ec T41Q/N71S/F142T) was found to exhibit significant improvements in *E. coli*, only modest improvements were observed in zebrafish. To improve on this, we sought to identify nitroreductases exhibiting superior prodrug activation of metronidazole and other nil-bystander prodrugs such as tinidazole, RB6145

and misonidazole. We also sought to develop systems for the independent ablation of multiple cell types, to enable researchers to gain a deeper understanding of regenerative and developmental processes occurring between two distinct cell types.

8.2 Summary of key findings

8.2.1 Identifying NfsA variants exhibiting superior chloramphenicol detoxification

A simultaneous site-directed mutagenesis library of *nfsA*, targeting eight active site residues, generated a predicted 252 million unique *nfsA* variants. This library was plated over a stringent selection containing a high concentration of chloramphenicol. Several NfsA variants were recovered which were able to protect *E. coli* 7NT cells to an almost 10-fold higher concentration of chloramphenicol than wild-type NfsA. The top two NfsA variants, NfsA 36_37 and 20_39, were used in further experiments to investigate how new enzyme functions evolve *in vivo*, using chloramphenicol IC₅₀ values as a measure of enzyme fitness.

8.2.2 Using NfsA 36_37 and 20_39 to improve our understanding of directed evolution

To understand whether a conventional stepwise mutagenesis approach could have evolved NfsA 36_37 and 20_39, all possible intermediates of these enzyme variants were created. The evolutionary trajectories of these two enzyme variants were recreated and it was found that a conventional stepwise evolutionary approach could have plausibly yielded both enzymes. Despite this, most evolutionary trajectories were selectively inaccessible due to deleterious or neutral combinations of substitutions, unlikely to be selected for using a conventional stepwise mutagenesis approach. Moreover, traditional approaches such as iterative site-saturation mutagenesis, would have ignored seven of the eight positions we chose. This is because only one substitution, R225V/D, provided an increase in enzyme fitness when introduced on its own. The remaining substitutions exhibited sign epistasis, only conditionally improving enzyme fitness depending on other specific substitutions also being introduced. Some substitutions like S41Y, T219Y and R225V/D, provided a fitness benefit in most allelic backgrounds. In contrast, other mutations such as H215N had a predominately negative effect on enzyme fitness, and only a provided a positive effect when introduced in combination with four other substitutions (S41Y, T219Y, R225D and F227H). Many of these substitutions may have been overlooked when taking an approach

like iterative saturation mutagenesis, as individually, they appeared neutral or even deleterious to enzyme fitness.

Another phenomenon commonly overlooked is the effect that metabolic interference can exert during directed evolution. Metabolic interference is a common occurrence in directed evolution and occurs when an evolved enzyme has an off-target effect with a metabolite(s) in an existing metabolic network. If this metabolite is part of an essential metabolic pathway it can have lethal consequences on a cell. Metabolic interference can also result in a decrease in the fitness benefit of an evolved enzyme, through competition of the active site with an off-target metabolite(s). We sought to use NfsA to model the off-target effects of directed evolution. We used a panel of nitroaromatic compounds to represent other metabolites that could exist in a metabolic network. We recovered NfsA variants from a variety of positive and dual positive-negative selections and challenged these variants against the panel of nitroaromatic substrates. NfsA variants recovered from a positive selection exhibited improved reduction of the positive selection compounds, chloramphenicol and niclosamide, but also improved reduction of a panel of structurally distinct nitroaromatic compounds (metronidazole, RB6145, CB1954, nitrofurazone and 2,4-DNT). However, the NfsA variants did not exhibit an equal improvement with all compounds, with substantial differences evident between the reduction of metronidazole and RB6145. A dual positive-negative selection was also conducted using the positive selection compound niclosamide in combination with either metronidazole or RB6145. This was used to simulate a scenario where metronidazole or RB6145 represented a metabolite within an essential metabolic pathway, while niclosamide represented the compound for which improved activity was desired. NfsA variants were recovered exhibiting decreased reduction of the “essential metabolite” (metronidazole or RB6145), while retaining niclosamide detoxifying abilities. However, the evolutionary constraints that were imposed did come at a cost in terms of niclosamide detoxification, as the NfsA variants that were recovered had a decreased median capacity to reduce niclosamide, relative to NfsA variants recovered from the exclusively positive selection.

8.2.3 Evolution of NfsA variants for improved targeted cell ablation in zebrafish

NfsB_Vv F70A/F108Y is the current lead candidate for metronidazole induced targeted cell ablation. This nitroreductase variant was rationally designed by previous members of the Ackerley group and was tested in zebrafish during this study as part of a collaborative

project. Robust ablation of cells expressing *nfsB_Vv F70A/F108Y*, was achieved at a 100-fold lower concentration of metronidazole than is currently required for *NfsB_Ec*. We also sought to identify *NfsA* variants that had achieved further gains in efficacy with metronidazole or alternative nil-bystander prodrugs (tinidazole, RB6145 and misonidazole). The Marsden library and the 2-nitroimidazole *nfsA* libraries were screened, enabling recovery of *NfsA* variants exhibiting improved reduction of all nil-bystander prodrugs tested. All *NfsA* variants appeared substantially more active when evaluated in *E. coli* than in HEK-293 cells. Despite this, several variants such as *NfsA 40_34*, may be of particular interest owing to its ability to reduce both metronidazole and misonidazole. It is possible that this variant could have utility as a dual-purpose nitroreductase, allowing for the ablation of stubborn cell types through co-administration of two different prodrugs. The generation of stable transgenic zebrafish lines expressing these evolved *nfsA* variants is currently ongoing in the Mumm lab, and functional testing will follow thereafter.

8.2.4 Evolution of nitroreductases to enable a multiplex targeted cell ablation system in zebrafish.

To enable spatially and temporally controlled targeted cell ablation of two separate cell populations a multiplex cell ablation system was proposed, using pairs of nitroreductases exhibiting non-overlapping drug specificities. *NfsB_Vv* had been previously identified as a nitroreductase exhibiting a preference for metronidazole over tinidazole. We sought to identify a partner nitroreductase, exhibiting the opposing prodrug selectivity. An error-prone PCR library of the previously identified variant *nfsA 5_B9*, was generated and was subjected to a dual positive-negative selection alongside the Marsden library to recover *NfsA* variants exhibiting preferential tinidazole activation over metronidazole. Several *NfsA* variants were recovered, however did not exhibit sufficient tinidazole selectivity to be used in combination with *NfsB_Vv*. This was likely due to prohibitive structural similarities between metronidazole and tinidazole. An alternative approach was taken, identifying nitroreductases that exhibited opposing specificities with the more structurally distinct compounds metronidazole and nitrofurazone. A dual positive-negative selection was used to recover *NfsA* variants from the Marsden library that exhibited high levels of activity with one prodrug, while excluding the other. This approach was successful in recovering *NfsA* variants exhibiting comparable nitrofurazone reduction compared to wild-type *NfsA*, but up to 70-fold diminished metronidazole activity. For some variants, this activity translated into HEK-293 cells, which were used as an indicator of tolerability and

functionality in zebrafish, however the functional expression of other variants was not able to be achieved in HEK-293. Moreover, a partner nitroreductase will need to be identified exhibiting the opposing prodrug selectivity for these NfsA variants to ultimately have utility in a multiplex cell ablation system.

8.2.5 Insights into the contributions of the eight-targeted residues in the Marsden library

Through the recovery of many distinct NfsA variants for a range of purposes, we have gained insights into which residues are crucial for the reduction of different substrates. An aromatic residue at position 41 was frequently observed in NfsA variants exhibiting improved reduction of several nitroaromatic substrates. Based on the crystal structure of NfsA, it has been hypothesised that a tyrosine substitution at position 41 enables planar stabilisation and stacking of aromatic substrates between the FMN isoalloxazine and tyrosine rings (Copp *et al.*, 2017). This explains why this substitution would improve the reduction of multiple substrates. Mutating position 225 to a polar or non-polar residue also appeared to be crucial for the improved reduction of most substrates tested in this study. Mutating this position to encode a smaller amino acid is predicted to enlarge the active site pocket, accommodating larger substrates. At position 219, a tyrosine substitution was highly prevalent in NfsA variants exhibiting improved reduction of metronidazole and RB6145. Position 219 was identified as a key position in NfsA variants evolved for multiplex cell ablation applications. A polar or non-polar amino acid introduced at this position was found to decrease metronidazole reduction without having a substantially negative impact on activity with other substrates such as tinidazole, nitrofurazone and niclosamide.

8.3 Critical evaluation of the methods used in this research

8.3.1 Selection of the eight residues to target in NfsA

Two preliminary *nfsA* mutagenesis libraries (8SM and 7SM libraries), were used to determine the final eight residues to target in the Marsden library. Residues which were frequently mutated in each library and yielded more active NfsA variants were chosen. Based on the high prevalence of epistasis observed in this study, the use of two distinct preliminary libraries was likely an imperfect strategy. In Chapter 3, we showed that only one of the substitutions introduced into the top chloramphenicol detoxifying variants conferred an improvement in enzyme fitness when introduced on its own. The other

substitutions only provided a positive fitness effect when introduced with one or more other substitutions, evidence of positive epistasis. There is a high chance we selected positions that could only provide a fitness benefit in conjunction with other positions from the 8SM or 7SM libraries. This may explain why substitutions at position K222 did not appear to exhibit a positive effect on enzyme fitness.

8.3.2 The use of the pUCX plasmid, library construction and IPTG induction

Mutations were found in the backbone of the pUCX plasmid, in the DNA purchased from Lab Genius. Several mutations were found in the *lacI* region that increased leaky nitroreductase expression. The use of no IPTG in the selection media and subsequent screening with low levels of IPTG biased the recovery of variants toward those containing these mutations. Induction with a high level of IPTG eliminated the selective advantage these variants had, allowing for equal expression of the *nfsA* variants during growth assays. In the future, the Marsden library could be remade using only DNA sourced from GenScript, eliminating *nfsA* variants bearing mutations in the backbone. An alternative expression plasmid could also be considered if the library was remade. As shown in Chapter 3 (Figure 3.3), the pUCX plasmid had leaky expression, and induction with IPTG resulted in high levels of enzyme expression. A high level of expression may not be necessary, and overexpression of an enzyme can be stressful or toxic to a cell, particularly if it interferes with host cell metabolism (Rosano and Ceccarelli, 2014). Slower growth of a recombinant strain compared to a strain expressing empty plasmid is evidence of this, a phenomenon that was observed in some *E. coli* 7NT strains bearing NfsA variants in this study. Lower expression levels may also aid in the selection of NfsA variants exhibiting a higher k_{cat} , as when less enzyme is present, a faster substrate turnover would be required to protect the cell against toxic compounds like chloramphenicol. This would need to be balanced against the lower level of protection conferred to the cell, with NfsA-mediated protection being considerably lower than that conferred by the chloramphenicol acetyltransferase selectable marker that is routinely used in molecular biology.

8.3.3 *In vivo* (*E. coli* 7NT) selection of NfsA variants

The evolved NfsA variants identified in this study consistently performed worse when tested in kinetic assays or when expressed in an alternative host (i.e., HEK-293 cells), than they had in *E. coli* 7NT. It was postulated that this was a result of recovering NfsA variants using an *in vivo* selection, whereby NfsA variants had lost inhibition by competing

metabolites that are absent in kinetic assays and may be present at lower levels in an alternative host cell. Supporting evidence for this theory was provided by the observation that the top five chloramphenicol detoxifying variants exhibited a substantially lowered capacity to reduce *p*-benzoquinone, a likely competing metabolite. Nevertheless, as the primary aim of Chapter 3 was to identify improved activity *in vivo*, this was not a substantial disadvantage. Further evidence for the was provided by the consistent activity profile of NfsB_Vv F70A/F108Y when expressed in *E.coli*, HEK-293 cells and zebrafish, this enzyme having been rationally engineered rather than identified using an *in vivo* selection. Understanding the key metabolic differences between *E. coli*, HEK-293 and zebrafish cells might guide future identification of promising candidates for zebrafish cell ablation applications.

8.3.4 Dual positive-negative selection

In many cases, the use of a dual positive-negative selection recovered generally less-active NfsA variants, rather than NfsA variants exhibiting prodrug selectivity. This may be due to the pressure a dual positive-negative selection imposes on a cell. This type of selection may cause a bottleneck effect, whereby the requirement to exclude one compound while reducing another, results in the elimination of most NfsA variants. A sequential selection may prove more successful, and identify NfsA variants exhibiting enhanced prodrug selectivity. This would involve plating a library on one selection compound to identify a smaller population which is sequentially plated on a second selection compound. This type of selection would impose less stress on the cells than a dual positive-negative selection, enabling healthier cells to be recovered from the selection. Fluorescence-activated cell sorting (FACS) could also be used as another approach (see Section 8.4.2).

8.4 Future directions

8.4.1 Additional evolutionary trajectories of top chloramphenicol detoxifying NfsA variants.

Evolutionary trajectories were generated for the top two chloramphenicol detoxifying NfsA variants (NfsA_36_37 and 20_39). These two variants exhibited almost identical evolutionary trajectories, with the order of the first four substitutions occurring at the same position, despite amino acid differences at two of the positions (R225 and F227). An additional evolutionary trajectory could be recreated for one or more NfsA variants to

assess whether this trajectory pattern is a population wide phenomenon. This experiment would be of particular interest if a NfsA variant was selected that did not exhibit all of the amino acid preferences commonly seen in the top chloramphenicol detoxifying NfsA variants. This could also help to understand the extent of amino acid redundancy evident at several of these positions.

8.4.2 Targeted cell ablation in zebrafish

The generation of transgenic zebrafish lines expressing the best *nfsA* variants identified in this study is currently being pursued by members of the Mumm lab. Preliminary expression tests will be localised to neuronal cells, enabling broad expression, however we hope that several of these *nfsA* variants will ultimately be expressed in many different cells, enabling targeted ablation of a multitude of cell types. Recently, our collaborators have successfully achieved robust cell ablation of macrophages expressing *nfsB_Vv F70A/F108Y*, a cell type which is notoriously difficult to ablate. Although currently limited to metronidazole-induced cell ablation, we hope to expand the utility of this system to other prodrugs such as tinidazole, RB6145 and misonidazole using the NfsA variants recovered in this study. These prodrugs may enable the ablation of more stubborn cell types either used in isolation or in combination with one another. A longer-term goal of this study is to expand this ablation system into other species using pan-species transgenic expression vectors. This would enable regenerative biology to be explored in other model organisms such as hydra, planaria, newt, frog and fish (Looso, 2014), as well as mice and rats.

8.4.3 Alternative selection to identify NfsA variants for multiplex cell ablation

A FACS-based experiment could be used to recover a nitroreductase variant exhibiting a preference for metronidazole over nitrofurazone, to be paired with the evolved NfsA variants presented in this study. In this approach, a *nfsA* mutagenesis library (or other nitroreductase) would need to be cloned into a SOS reporter strain that has a fluorescent reporter gene under control of a DNA damage inducible promoter, allowing for prodrug induced DNA damage to be measured via fluorescence. The library would then be challenged with the prodrug that activity is desired with (i.e., metronidazole). Variants exhibiting activity with the prodrug would then elicit a fluorescent response, allowing the cells to be sorted using FACS. This pool of variants would then be counter-screened using a plate-based assay or a solid media selection to isolate variants which do not exhibit activity with the other compound (i.e., nitrofurazone). This method would have the

advantage of providing a direct selection for metronidazole activity, unlike niclosamide-based selection, which is an indirect selection that may provide an undesirable activity bottleneck. SOS-GFP FACS has been successfully used in the Ackerley group to recover nitroreductase variants highly active with a prodrug of interest (Copp *et al.*, 2014b; Copp *et al.*, 2017). However, it is likely that replacement of the GFP reporter gene with a recent generation red reporter would prove even more effective, as the longer-wavelength emission would be further away from the *E. coli* auto-fluorescence spectrum (Surre *et al.*, 2018). This method may hold promise for recovering nitroreductase variants exhibiting a desired prodrug selectivity.

8.5 Concluding remarks

This thesis demonstrated the amenability of NfsA for use in a range of fundamental and biotechnological applications. We demonstrated that evolution of two NfsA variants exhibiting superior chloramphenicol detoxification required a series of complex epistatic interactions in order to scale the fitness landscape. Using NfsA as a model, we also demonstrated the high frequency of adventitious and off-target interactions which can occur as a result of directed evolution. Finally, we recovered many NfsA variants exhibiting substantially improved prodrug activation *in vivo*, for use in nitroreductase mediated targeted cell ablation. We hope the use of these alternative nitroreductase prodrug pairings will enable more efficient targeted cell ablation in zebrafish with fewer off-target downstream effects, allowing more progress to be made in the regenerative biology field.

References

- Ackerley, D. F., Gonzalez, C. F., Keyhan, M., Blake, R., 2nd and Matin, A. (2004). Mechanism of chromate reduction by the *Escherichia coli* protein, NfsA, and the role of different chromate reductases in minimizing oxidative stress during chromate reduction. *Environ Microbiol.* 6(8), 851-860.
- Adams, G. E., Clarke, E. D., Flockhart, I. R., Jacobs, R. S., Sehmi, D. S., Stratford, I. J., Wardman, P., Watts, M. E., Parrick, J., Wallace, R. G. and Smithen, C. E. (1979). Structure-activity relationships in the development of hypoxic cell radiosensitizers. I. Sensitization efficiency. *Int J Radiat Biol Relat Stud Phys Chem Med.* 35(2), 133-150.
- Aharoni, A., Gaidukov, L., Khersonsky, O., Mc, Q. G. S., Roodveldt, C. and Tawfik, D. S. (2005a). The 'evolvability' of promiscuous protein functions. *Nat Genet.* 37(1), 73-76.
- Aharoni, A., Griffiths, A. D. and Tawfik, D. S. (2005b). High-throughput screens and selections of enzyme-encoding genes. *Curr Opin Chem Biol.* 9(2), 210-216.
- Akagawa, H., Okanishi, M. and Umezawa, H. (1975). A plasmid involved in chloramphenicol production in *Streptomyces venezuelae*: Evidence from genetic mapping. *J Gen Microbiol.* 90(2), 336-346.
- Akitake, C. M., Macurak, M., Halpern, M. E. and Goll, M. G. (2011). Transgenerational analysis of transcriptional silencing in zebrafish. *Dev Biol.* 352(2), 191-201.
- Akiva, E., Copp, J. N., Tokuriki, N. and Babbitt, P. C. (2017). Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc Natl Acad Sci USA.* 114(45), e9549.
- Ariga, J., Walker, S. L. and Mumm, J. S. (2010). Multicolor time-lapse imaging of transgenic zebrafish: Visualizing retinal stem cells activated by targeted neuronal cell ablation. *J Vis Exp.* (43), e2093.
- Arnold, F. H. (1998). Design by directed evolution. *Acc Chem Res.* 31(3), 125-131.
- Asnis, R. E. (1957). The reduction of Furacin by cell-free extracts of Furacin-resistant and parent-susceptible strains of *Escherichia coli*. *Arch Biochem Biophys.* 66(1), 208-216.
- Bantscheff, M., Scholten, A. and Heck, A. J. (2009). Revealing promiscuous drug-target interactions by chemical proteomics. *Drug Discov Today.* 14(21-22), 1021-1029.
- Barak, Y., Thorne, S. H., Ackerley, D. F., Lynch, S. V., Contag, C. H. and Matin, A. (2006). New enzyme for reductive cancer chemotherapy, YieF, and its improvement by directed evolution. *Mol Cancer Ther.* 5(1), 97-103.
- Barriault, D., Plante, M. M. and Sylvestre, M. (2002). Family shuffling of a targeted bphA region to engineer biphenyl dioxygenase. *J Bacteriol.* 184(14), 3794-3800.
- Bloom, J. D. and Arnold, F. H. (2009). In the light of directed evolution: Pathways of adaptive protein evolution. *Proc Natl Acad Sci USA.* 106, 9995-10000.

- Bornscheuer, U. T., Huisman, G. W., Kazlauskas, R. J., Lutz, S., Moore, J. C. and Robins, K. (2012). Engineering the third wave of biocatalysis. *Nature*. 485(7397), 185-194.
- Bremner, J. C. M. (1993). Assessing the bioreductive effectiveness of the nitroimidazole RSU1069 and its prodrug RB6145: With particular reference to *in vivo* methods of evaluation. *Cancer Metastasis Rev.* 12(2), 177-193.
- Bridgewater, J. A., Springer, C. J., Knox, R. J., Minton, N. P., Michael, N. P. and Collins, M. K. (1995). Expression of the bacterial nitroreductase enzyme in mammalian cells renders them selectively sensitive to killing by the prodrug CB1954. *Eur J Cancer*. 31a(13-14), 2362-2370.
- Brown, J. M. and Giaccia, A. J. (1998). The unique physiology of solid tumors: opportunities (and problems) for cancer therapy. *Cancer Res*. 58(7), 1408-1416.
- Brown, J. M. and Workman, P. (1980). Partition coefficient as a guide to the development of radiosensitizers which are less toxic than misonidazole. *Radiat Res*. 82(1), 171-190.
- Bryant, C. and DeLuca, M. (1991). Purification and characterization of an oxygen-insensitive NAD(P)H nitroreductase from *Enterobacter cloacae*. *J Biol Chem*. 266(7), 4119-4125.
- Bryant, D. W., McCalla, D. R., Leeksa, M. and Laneuville, P. (1981). Type I nitroreductases of *Escherichia coli*. *Can J Microbiol*. 27(1), 81-86.
- Cadwell, R. C. and Joyce, G. F. (1992). Randomization of genes by PCR mutagenesis. *PCR Methods Appl.* 2(1), 28-33.
- Chaignon, P., Cortial, S., Ventura, A. P., Lopes, P., Halgand, F., Laprevote, O. and Ouazzani, J. (2006). Purification and identification of a *Bacillus* nitroreductase: Potential use in 3,5-DNBTF biosensing system. *Enzyme Microb Tech.* 39(7), 1499-1506.
- Chan-Hyams, J. V. E., Copp, J. N., Smaill, J. B., Patterson, A. V. and Ackerley, D. F. (2018). Evaluating the abilities of diverse nitroaromatic prodrug metabolites to exit a model Gram negative vector for bacterial-directed enzyme-prodrug therapy. *Biochem Pharmacol.* 158, 192-200.
- Chaplin, D. J., Durand, R. E., Stratford, I. J. and Jenkins, T. C. (1986). The radiosensitizing and toxic effects of RSU1069 on hypoxic cells in a murine tumor. *Int J Radiat Oncol Biol Phys.* 12(7), 1091-1095.
- Chen, C. F., Chu, C. Y., Chen, T. H., Lee, S. J., Shen, C. N. and Hsiao, C. D. (2011). Establishment of a transgenic zebrafish line for superficial skin ablation and functional validation of apoptosis modulators *in vivo*. *PLoS One*. 6(5), e20654.
- Chen, K. Q. and Arnold, F. H. (1993). Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Proc Natl Acad Sci USA*. 90(12), 5618-5622.
- Chen, R. D. (2001). Enzyme engineering: rational redesign versus directed evolution. *Trends Biotechnol.* 19(1), 13-14.

- Chen, W., Bruhlmann, F., Richins, R. D. and Mulchandani, A. (1999). Engineering of improved microbes and enzymes for bioremediation. *Curr Opin Biotechnol.* 10(2), 137-141.
- Chica, R. A., Doucet, N. and Pelletier, J. N. (2005). Semi-rational approaches to engineering enzyme activity: Combining the benefits of directed evolution and rational design. *Curr Opin Biotechnol.* 16(4), 378-384.
- Chu, Y., Senghaas, N., Koster, R. W., Wurst, W. and Kuhn, R. (2008). Novel caspase-suicide proteins for tamoxifen-inducible apoptosis. *Genesis.* 46(10), 530-536.
- Chung, C. H., Hao, E., Piran, R., Keinan, E. and Levine, F. (2010). Pancreatic beta-cell neogenesis by direct conversion from mature alpha-cells. *Stem Cells.* 28(9), 1630-1638.
- Chung, H. W. and Tu, S. C. (2012). Structure-function relationship of *Vibrio harveyi* NADPH-flavin oxidoreductase FRP: Essential residues Lys167 and Arg15 for NADPH binding. *Biochemistry.* 51(24), 4880-4887.
- Cirino, P. C., Mayer, K. M. and Umeno, D. (2003). Generating mutant libraries using error-prone PCR. Arnold, F. H. and Georgiou, G. (eds). Directed Evolution Library Creation. Methods in Molecular Biology, vol 231, 3-9. Humana Press.
- Cobb, L. M., Connors, T. A., Elson, L. A., Khan, A. H., Mitchley, B. C., Ross, W. C. and Whisson, M. E. (1969). 2,4-dinitro-5-ethyleneiminobenzamide (CB1954): A potent and selective inhibitor of the growth of the Walker carcinoma 256. *Biochem Pharmacol.* 18(6), 1519-1527.
- Condon, S. (2013). The characterisation and application of bacterial nitroreductase enzymes. Masters thesis. School of Biological Sciences, Victoria University of Wellington, Wellington.
- Copley, S. D. (2017). Shining a light on enzyme promiscuity. *Curr Opin Struct Biol.* 47, 167-175.
- Copp, J. N., Hanson-Manful, P., Ackerley, D. F. and Patrick, W. M. (2014a). Error-prone PCR and effective generation of gene variant libraries for directed evolution. Gillam, E., Copp, J. N. and Ackerley, D. F. (eds). Directed Evolution Library Creation. Methods in Molecular Biology, vol 1179, 3-22. Springer.
- Copp, J. N., Mowday, A. M., Williams, E. M., Guise, C. P., Ashoorzadeh, A., Sharrock, A. V., Flanagan, J. U., Smaill, J. B., Patterson, A. V. and Ackerley, D. F. (2017). Engineering a multifunctional nitroreductase for improved activation of prodrugs and PET probes for cancer gene therapy. *Cell Chem Biol.* 24(3), 391-403.
- Copp, J. N., Williams, E. M., Rich, M. H., Patterson, A. V., Smaill, J. B. and Ackerley, D. F. (2014b). Toward a high-throughput screening platform for directed evolution of enzymes that activate genotoxic prodrugs. *Protein Eng Des Sel.* 27(10), 399-403.
- Cortial, S., Chaignon, P., Sargent, D., Dezard, S. and Ouazzani, J. (2012). Dehydrogenation, oxidative denitration and ring contraction of N,N-dimethyl-5-nitrouacil by a *Bacillus* nitroreductase Nfr-A1. *J Mol Catal B: Enzymatic.* 76, 1-8.

- Crofts, T. S., Sontha, P., King, A. O., Wang, B., Biddy, B. A., Zanolli, N., Gaumnitz, J. and Dantas, G. (2019). Discovery and characterization of a nitroreductase capable of conferring bacterial resistance to chloramphenicol. *Cell Chem Biol.* 26(4), 559-570.
- Curado, S., Anderson, R. M., Jungblut, B., Mumm, J., Schroeter, E. and Stainier, D. Y. (2007). Conditional targeted cell ablation in zebrafish: a new tool for regeneration studies. *Dev Dyn.* 236(4), 1025-1035.
- Dalby, P. A. (2011). Strategy and success for the directed evolution of enzymes. *Curr Opin Struct Biol.* 21(4), 473-480.
- Deist, B. R., Rausch, M. A., Fernandez-Luna, M. T., Adang, M. J. and Bonning, B. C. (2014). Bt toxin modification for enhanced efficacy. *Toxins (Basel).* 6(10), 3005-3027.
- Denny, W. A. (2002). Nitroreductase-based GDEPT. *Curr Pharm Des.* 8(15), 1349-1361.
- Dent, J. G., Schnell, S. R. and Guest, D. (1981). Metabolism of 2,4-dinitrotoluene by rat hepatic microsomes and cecal flora. *Adv Exp Med Biol.* 136, 431-436.
- Dische, S., Saunders, M. I., Flockhart, I. R., Lee, M. E. and Anderson, P. (1979). Misonidazole-a drug for trial in radiotherapy and oncology. *Int J Radiat Oncol Biol Phys.* 5(6), 851-860.
- Distel, M., Wullimann, M. F. and Koster, R. W. (2009). Optimized Gal4 genetics for permanent gene expression mapping in zebrafish. *Proc Natl Acad Sci USA.* 106(32), 13365-13370.
- Dooley, K. and Zon, L. I. (2000). Zebrafish: a model system for the study of human disease. *Curr Opin Genet Dev.* 10(3), 252-256.
- Driever, W., SolnicaKrezel, L., Schier, A. F., Neuhauss, S. C. F., Malicki, J., Stemple, D. L., Stainier, D. Y. R., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J. and Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development.* 123, 37-46.
- Drummond, I. A., Majumdar, A., Hentschel, H., Elger, M., Solnica-Krezel, L., Schier, A. F., Neuhauss, S. C., Stemple, D. L., Zwartkruis, F., Rangini, Z., Driever, W. and Fishman, M. C. (1998). Early development of the zebrafish pronephros and analysis of mutations affecting pronephric function. *Development.* 125(23), 4655-4667.
- Edwards, D. I. (1980). Mechanisms of selective toxicity of metronidazole and other nitroimidazole drugs. *Br J Vener Dis.* 56(5), 285-290.
- Elshami, A. A., Saavedra, A., Zhang, H., Kucharczuk, J. C., Spray, D. C., Fishman, G. I., Amin, K. M., Kaiser, L. R. and Albelda, S. M. (1996). Gap junctions play a role in the 'bystander effect' of the herpes simplex virus thymidine kinase ganciclovir system in vitro. *Gene Ther.* 3(1), 85-92.
- Engqvist, M. K. M. and Rabe, K. S. (2019). Applications of protein engineering and directed evolution in plant research. *Plant Physiol.* 179(3), 907-917.

- Fernandes, A. M., Fero, K., Arrenberg, A. B., Bergeron, S. A., Driever, W. and Burgess, H. A. (2012). Deep brain photoreceptors control light-seeking behavior in zebrafish larvae. *Curr Biol.* 22(21), 2042-2047.
- Festa, F., Steel, J., Bian, X. F. and Labaer, J. (2013). High-throughput cloning and expression library creation for functional proteomics. *Proteomics.* 13(9), 1381-1399.
- Fielden, E. M., Adams, G. E., Cole, S., Naylor, M. A., O'Neill, P., Stephens, M. A. and Stratford, I. J. (1992). Assessment of a range of novel nitro-aromatic radiosensitizers and bioreductive drugs. *Int J Radiat Oncol Biol Phys.* 22(4), 707-711.
- Fishman, M. C. and Olson, E. N. (1997). Parsing the heart: Genetic modules for organ assembly. *Cell.* 91(2), 153-156.
- Frayha, G. J., Smyth, J. D., Gobert, J. G. and Savel, J. (1997). The mechanisms of action of antiprotozoal and anthelmintic drugs in man. *Gen Pharmacol.* 28(2), 273-299.
- Freeman, C. D., Klutman, N. E. and Lamp, K. C. (1997). Metronidazole: A therapeutic review and update. *Drugs.* 54(5), 679-708.
- Freeman, S. M., Abboud, C. N., Whartenby, K. A., Packman, C. H., Koeplin, D. S., Moolten, F. L. and Abraham, G. N. (1993). The bystander effect: Tumor regression when a fraction of the tumor mass is genetically-modified. *Cancer Res.* 53(21), 5274-5283.
- Friedlos, F., Denny, W. A., Palmer, B. D. and Springer, C. J. (1997). Mustard prodrugs for activation by *Escherichia coli* nitroreductase in gene-directed enzyme prodrug therapy. *J Med Chem.* 40(8), 1270-1275.
- Fu, Y., Buryanovskyy, L. and Zhang, Z. (2005). Crystal structure of quinone reductase 2 in complex with cancer prodrug CB1954. *Biochem Biophys Res Commun.* 336(1), 332-338.
- Fung, H. B. and Doan, T. L. (2005). Tinidazole: A nitroimidazole antiprotozoal agent. *Clin Ther.* 27(12), 1859-1884.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. and Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. Walker, J. M. (eds). The proteomics protocols handbook, 571-607. Humana Press.
- Gemberling, M., Bailey, T. J., Hyde, D. R. and Poss, K. D. (2013). The zebrafish as a model for complex tissue regeneration. *Trends Genet.* 29(11), 611-620.
- Georgiou, G. and DeWitt, N. (1999). Enzyme beauty. *Nat Biotechnol.* 17(12), 1161-1162.
- Gocke, D., Walter, L., Gauchenova, E., Kolter, G., Knoll, M., Berthold, C. L., Schneider, G., Pleiss, J., Muller, M. and Pohl, M. (2008). Rational protein design of ThDP-dependent enzymes-engineering stereoselectivity. *ChemBiochem.* 9(3), 406-412.
- Goldman, P., Koch, R. L., Yeung, T. C., Chrystal, E. J., Beaulieu, B. B., Jr., McLafferty, M. A. and Sudlow, G. (1986). Comparing the reduction of nitroimidazoles in bacteria and mammalian tissues and relating it to biological activity. *Biochem Pharmacol.* 35(1), 43-51.

- Goll, M. G., Anderson, R., Stainier, D. Y., Spradling, A. C. and Halpern, M. E. (2009). Transcriptional silencing and reactivation in transgenic zebrafish. *Genetics*. 182(3), 747-755.
- Gondry, M., Lautru, S., Fusai, G., Meunier, G., Menez, A. and Genet, R. (2001). Cyclic dipeptide oxidase from *Streptomyces noursei*: Isolation, purification and partial characterization of a novel, amino acyl alpha,beta-dehydrogenase. *Eur J Biochem*. 268(6), 1712-1721.
- Gong, P., Siciliano, S. D., Greer, C. W., Paquet, L., Hawari, J. and Sunahara, G. I. (1999). Effects and bioavailability of 2,4,6-trinitrotoluene in spiked and field-contaminated soils to indigenous microorganisms. *Environ Toxicol Chem*. 18(12), 2681-2688.
- Gonzalez, C. F., Ackerley, D. F., Lynch, S. V. and Matin, A. (2005). ChrR, a soluble quinone reductase of *Pseudomonas putida* that defends against H₂O₂. *J Biol Chem*. 280(24), 22590-22595.
- Goodwin, A., Kersulyte, D., Sisson, G., van Zanten, S. J. O. V., Berg, D. E. and Hoffman, P. S. (1998). Metronidazole resistance in *Helicobacter pylori* is due to null mutations in a gene (rdxA) that encodes an oxygen-insensitive NADPH nitroreductase. *Mol Microbiol*. 28(2), 383-393.
- Gooyit, M. and Janda, K. D. (2016). Reprofiled anthelmintics abate hypervirulent stationary-phase *Clostridium difficile*. *Sci Rep*. 6, 33642.
- Gray, C., Loynes, C. A., Whyte, M. K. B., Crossman, D. C., Renshaw, S. A. and Chico, T. J. A. (2011). Simultaneous intravital imaging of macrophage and neutrophil behaviour during inflammation using a novel transgenic zebrafish. *Thromb Haemost*. 105(5), 811-819.
- Green, K. D., Fosso, M. Y., Mayhoub, A. S., Garneau-Tsodikova, S. (2019). Investigating the promiscuity of the chloramphenicol nitroreductase from *Haemophilus influenzae* towards the reduction of 4-nitrobenzene derivatives. *Bioorg Med Chem Lett*. 29(9), 1127-1132.
- Green, L. K., La Flamme, A. C. and Ackerley, D. F. (2014). *Pseudomonas aeruginosa* MdaB and WrbA are water-soluble two-electron quinone oxidoreductases with the potential to defend against oxidative stress. *J Microbiol*. 52(9), 771-777.
- Grove, J. I., Lovering, A. L., Guise, C., Race, P. R., Wrighton, C. J., White, S. A., Hyde, E. I. and Searle, P. F. (2003). Generation of *Escherichia coli* nitroreductase mutants conferring improved cell sensitization to the prodrug CB1954. *Cancer Res*. 63(17), 5532-5537.
- Guise, C. P., Grove, J. I., Hyde, E. I. and Searle, P. F. (2007). Direct positive selection for improved nitroreductase variants using SOS triggering of bacteriophage lambda lytic cycle. *Gene Ther*. 14(8), 690-698.
- Gulick, A. M. and Fahl, W. E. (1995). Forced evolution of glutathione S-transferase to create a more efficient drug detoxication enzyme. *Proc Natl Acad Sci USA*. 92(18), 8140-8144.

- Gwenin, V. V., Poornima, P., Halliwell, J., Ball, P., Robinson, G. and Gwenin, C. D. (2015). Identification of novel nitroreductases from *Bacillus cereus* and their interaction with the CB1954 prodrug. *Biochem Pharmacol.* 98(3), 392-402.
- Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., van Eeden, F. J., Jiang, Y. J., Heisenberg, C. P., Kelsh, R. N., Furutani-Seiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C. and Nusslein-Volhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development.* 123, 1-36.
- Halpern, M. E., Rhee, J., Goll, M. G., Akitake, C. M., Parsons, M. and Leach, S. D. (2008). Gal4/UAS transgenic tools and their application to zebrafish. *Zebrafish.* 5(2), 97-110.
- Hamel, W., Magnelli, L., Chiarugi, V. P. and Israel, M. A. (1996). Herpes simplex virus thymidine kinase/ganciclovir-mediated apoptotic death of bystander cells. *Cancer Res.* 56(12), 2697-2702.
- Hannink, N., Rosser, S. J., French, C. E., Basran, A., Murray, J. A., Nicklin, S. and Bruce, N. C. (2001). Phytodetoxification of TNT by transgenic plants expressing a bacterial nitroreductase. *Nat Biotechnol.* 19(12), 1168-1172.
- Hansen, E. H., Moller, B. L., Kock, G. R., Bunner, C. M., Kristensen, C., Jensen, O. R., Okkels, F. T., Olsen, C. E., Motawia, M. S. and Hansen, J. (2009). *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). *Appl Environ Microbiol.* 75(9), 2765-2774.
- Harel, M., Aharoni, A., Gaidukov, L., Brumshtein, B., Khersonsky, O., Meged, R., Dvir, H., Ravelli, R. B. G., McCarthy, A., Toker, L., Silman, I., Sussman, J. L. and Tawfik, D. S. (2004). Structure and evolution of the serum paraoxonase family of detoxifying and anti-atherosclerotic enzymes. *Nat Struct Mol Biol.* 11(12), 412-419.
- Harris, J. A., Cheng, A. G., Cunningham, L. L., MacDonald, G., Raible, D. W. and Rubel, E. W. (2003). Neomycin-induced hair cell death and rapid regeneration in the lateral line of zebrafish (*Danio rerio*). *J Assoc Res Otolaryngol.* 4(2), 219-234.
- Hawari, J., Beaudet, S., Halasz, A., Thiboutot, S. and Ampleman, G. (2000). Microbial degradation of explosives: Biotransformation versus mineralization. *Appl Microbiol Biotechnol.* 54(5), 605-618.
- Heiss, G. and Knackmuss, H. J. (2002). Bioelimination of trinitroaromatic compounds: Immobilization versus mineralization. *Curr Opin Microbiol.* 5(3), 282-287.
- Hellinga, H. W. (1997). Rational protein design: combining theory and experiment. *Proc Natl Acad Sci USA.* 94(19), 10015-10017.
- Helsby, N. A., Ferry, D. M., Patterson, A. V., Pullen, S. M. and Wilson, W. R. (2004). 2-Amino metabolites are key mediators of CB1954 and SN23862 bystander effects in nitroreductase GDEPT. *Br J Cancer.* 90(5), 1084-1092.
- Holmes, R. K. (2000). Biology and molecular epidemiology of diphtheria toxin and the tox gene. *J Infect Dis.* 181, S156-S167.

Hoogenraad, M., van der Linden, J. B., Smith, A. A., Hughes, B., Derrick, A. M., Harris, L. J., Higginson, P. D. and Pettman, A. J. (2004). Accelerated process development of pharmaceuticals: Selective catalytic hydrogenations of nitro compounds containing other functionalities. *Org Process Res Dev.* 8(3), 469-476.

Horvat, C. N. (2012). Development and applications of nitroreductase activated masked fluorophores. PhD thesis. School of Biological Sciences, Victoria University of Wellington, Wellington.

Hou, F., Miyakawa, T., Kitamura, N., Takeuchi, M., Park, S. B., Kishino, S., Ogawa, J. and Tanokura, M. (2015). Structure and reaction mechanism of a novel enone reductase. *FEBS Journal.* 282(8), 1526-1537.

Hu, S. Y., Lin, P. Y., Liao, C. H., Gong, H. Y., Lin, G. H., Kawakami, K. and Wu, J. L. (2010). Nitroreductase-mediated gonadal dysgenesis for infertility control of genetically modified zebrafish. *Mar Biotechnol (NY).* 12(5), 569-578.

Huang, H., Pandya, C., Liu, C., Al-Obaidi, N. F., Wang, M., Zheng, L., Keating, S. T., Aono, M., Love, J. D., Evans, B., Seidel, R. D., Hillerich, B. S., Garforth, S. J., Almo, S. C., Mariano, P. S., Dunaway-Mariano, D., Allen, K. N. and Farelli, J. D. (2015). Panoramic view of a superfamily of phosphatases through substrate profiling. *Proc Natl Acad Sci USA.* 112(16), e1974-e1983.

Imai, A. and Gloyna, E. F. (1990). Effects of pH and oxidation-state of chromium on the behavior of chromium in the activated-sludge process. *Water Res.* 24(9), 1143-1150.

Iqbal, T. and Byrd-Jacobs, C. (2010). Rapid degeneration and regeneration of the zebrafish olfactory epithelium after triton X-100 application. *Chem Senses.* 35(5), 351-361.

Jaberipour, M., Vass, S. O., Guise, C. P., Grove, J. I., Knox, R. J., Hu, L., Hyde, E. I. and Searle, P. F. (2010). Testing double mutants of the enzyme nitroreductase for enhanced cell sensitisation to prodrugs: effects of combining beneficial single mutations. *Biochem Pharmacol.* 79(2), 102-111.

Jacob, F. and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.* 3, 318-356.

Jenkins, T. C., Naylor, M. A., O'Neill, P., Threadgill, M. D., Cole, S., Stratford, I. J., Adams, G. E., Fielden, E. M., Suto, M. J. and Stier, M. A. (1990). Synthesis and evaluation of α -[[2-(haloethyl)amino]methyl]-2-nitro-1*H*-imidazole-1-ethanols as prodrugs of α -[(1-aziridinyl)methyl]-2-nitro-1*H*-imidazole-1-ethanol (RSU-1069) and its analogues which are radiosensitizers and bioreductively activated cytotoxins. *J. Med. Chem.* 33, 2603-2610.

Jenkins, T. F., Leggett, D. C., Miyares, P. H., Walsh, M. E., Ranney, T. A., Cragin, J. H. and George, V. (2001). Chemical signatures of TNT-filled land mines. *Talanta.* 54(3), 501-513.

Jensen, R. A. (1976). Enzyme recruitment in evolution of new function. *Annu Rev Microbiol.* 30, 409-425.

Ju, K. S. and Parales, R. E. (2010). Nitroaromatic compounds, from synthesis to biodegradation. *Microbiol Mol Biol Rev.* 74(2), 250-272.

- Kapoor, S., Rafiq, A. and Sharma, S. (2017). Protein engineering and its applications in food industry. *Crit Rev Food Sci Nutr.* 57(11), 2321-2329.
- Khan, A. H. and Ross, W. C. (1969). Tumour-growth inhibitory nitrophenylaziridines and related compounds: Structure-activity relationships. *Chem Biol Interact.* 1(1), 27-47.
- Kim, H. Y. and Song, H. G. (2005). Purification and characterization of NAD(P)H-dependent nitroreductase I from *Klebsiella* sp. C1 and enzymatic transformation of 2,4,6-trinitrotoluene. *Appl Microbiol Biotechnol.* 68(6), 766-773.
- Kim, J. and Copley, S. D. (2012). Inhibitory cross-talk upon introduction of a new metabolic pathway into an existing metabolic network. *Proc Natl Acad Sci USA.* 109(42), e2856-e2864.
- Kim, M. S. and Lei, X. G. (2008). Enhancing thermostability of *Escherichia coli* phytase AppA2 by error-prone PCR. *Appl Microbiol Biotechnol.* 79(1), 69-75.
- Knight, R. C., Rowley, D. A., Skolimowski, I. and Edwards, D. I. (1979). Mechanism of action of nitroimidazole antimicrobial and antitumour radiosensitizing drugs. Effects of reduced misonidazole on DNA. *Int J Radiat Biol Relat Stud Phys Chem Med.* 36(4), 367-377.
- Knox, R. J., Friedlos, F. and Boland, M. P. (1993). The bioactivation of CB 1954 and its use as a prodrug in antibody-directed enzyme prodrug therapy (ADEPT). *Cancer Metast Rev.* 12(2), 195-212.
- Knox, R. J., Friedlos, F., Marchbank, T. and Roberts, J. J. (1991). Bioactivation of CB 1954: reaction of the active 4-hydroxylamino derivative with thioesters to form the ultimate DNA-DNA interstrand cross-linking species. *Biochem Pharmacol.* 42(9), 1691-1697.
- Kobori, T., Sasaki, H., Lee, W. C., Zenno, S., Saigo, K., Murphy, M. E. P. and Tanokura, M. (2001). Structure and site-directed mutagenesis of a flavoprotein from *Escherichia coli* that reduces nitrocompounds: alteration of pyridine nucleotide binding by a single amino acid substitution. *J Biol Chem.* 276(4), 2816-2823.
- Koder, R. L., Haynes, C. A., Rodgers, M. E., Rodgers, D. W. and Miller, A. F. (2002). Flavin thermodynamics explain the oxygen insensitivity of enteric nitroreductases. *Biochemistry.* 41(48), 14197-14205.
- Koder, R. L. and Miller, A. F. (1998). Steady-state kinetic mechanism, stereospecificity, substrate and inhibitor specificity of *Enterobacter cloacae* nitroreductase. *Biochimica et Biophysica Acta.* 1387(1-2), 395-405.
- Koga, A., Suzuki, M., Inagaki, H., Bessho, Y. and Hori, H. (1996). Transposable element in fish. *Nature.* 383(6595), 30.
- Kumamaru, T., Suenaga, H., Mitsuoka, M., Watanabe, T. and Furukawa, K. (1998). Enhanced degradation of polychlorinated biphenyls by directed evolution of biphenyl dioxygenase. *Nat Biotechnol.* 16(7), 663-666.
- Kurumata, M., Takahashi, M., Sakamoto, A., Ramos, J. L., Nepovim, A., Vanek, T., Hirata, T. and Morikawa, H. (2005). Tolerance to, and uptake and degradation of 2,4,6-

trinitrotoluene (TNT) are enhanced by the expression of a bacterial nitroreductase gene in *Arabidopsis thaliana*. *Z Naturforsch C*. 60(3-4), 272-278.

Kutty, R. and Bennett, G. N. (2005). Biochemical characterization of trinitrotoluene transforming oxygen-insensitive nitroreductases from *Clostridium acetobutylicum* ATCC 824. *Arch Microbiol*. 184(3), 158-167.

Kvitek, D. J. and Sherlock, G. (2011). Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet*. 7(4), e1002056.

Kwak, Y. H., Lee, D. S. and Kim, H. B. (2003). *Vibrio harveyi* nitroreductase is also a chromate reductase. *Appl Environ Microbiol*. 69(8), 4390-4395.

Langenau, D. M., Ferrando, A. A., Traver, D., Kutok, J. L., Hezel, J. P., Kanki, J. P., Zon, L. I., Look, A. T. and Trede, N. S. (2004). *In vivo* tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc Natl Acad Sci USA*. 101(19), 7369-7374.

Leemhuis, H., Kelly, R. M. and Dijkhuizen, L. (2009). Directed evolution of enzymes: library screening strategies. *IUBMB Life*. 61(3), 222-228.

Lehouritis, P., Springer, C. and Tangney, M. (2013). Bacterial-directed enzyme prodrug therapy. *J Control Release*. 170(1), 120-131.

Lewis, M. (2005). The lac repressor. *C R Biol*. 328(6), 521-548.

Li, W. J., Xu, S. T., Zhang, B., Zhu, Y. L., Hua, Y., Kong, X., Sun, L. H. and Hong, J. (2017). Directed evolution to improve the catalytic efficiency of urate oxidase from *Bacillus subtilis*. *PLoS One*. 12(5), e0177877.

Li, X. L., Montgomery, J., Cheng, W., Noh, J. H., Hyde, D. R. and Li, L. (2012). Pineal photoreceptor cells are required for maintaining the circadian rhythms of behavioral visual sensitivity in zebrafish. *PLoS One*. 7(7), e40508.

Liochev, S. I., Hausladen, A. and Fridovich, I. (1999). Nitroreductase A is regulated as a member of the soxRS regulon of *Escherichia coli*. *Proc Natl Acad Sci USA*. 96(7), 3537-3539.

Little, R. F. (2015). Directed evolution and discovery of nitroreductase enzymes for targeted cell ablation. School of Biological Sciences, Victoria University of Wellington, Wellington.

Lo, T. P., Komarpanicucci, S., Sherman, F., Mclendon, G. and Brayer, G. D. (1995). Structural and functional effects of multiple mutations at distal sites in cytochrome-C. *Biochemistry*. 34(15), 5259-5268.

Looso, M. (2014). Opening the genetic toolbox of niche model organisms with high throughput techniques: Novel proteins in regeneration as a case study. *Bioessays*. 36(4), 407-418.

Lutz, S. and Patrick, W. M. (2004). Novel methods for directed evolution of enzymes: quality, not quantity. *Curr Opin Biotechnol*. 15(4), 291-297.

- Macheroux, P., Kappes, B. and Ealick, S. E. (2011). Flavogenomics - a genomic and structural view of flavin-dependent proteins. *FEBS J.* 278(15), 2625-2634.
- Mallet, V. O., Mitchell, C., Guidotti, J. E., Jaffray, P., Fabre, M., Spencer, D., Arnoult, D., Kahn, A. and Gilgenkrantz, H. (2002). Conditional cell ablation by tight control of caspase-3 dimerization in transgenic mice. *Nat Biotechnol.* 20(12), 1234-1239.
- Manina, G., Bellinzoni, M., Pasca, M. R., Neres, J., Milano, A., Ribeiro, A. L., Buroni, S., Skovierova, H., Dianiskova, P., Mikusova, K., Marak, J., Makarov, V., Giganti, D., Haouz, A., Lucarelli, A. P., Degiacomi, G., Piazza, A., Chiarelli, L. R., De Rossi, E., Salina, E., Cole, S. T., Alzari, P. M. and Riccardi, G. (2010). Biological and structural characterization of the *Mycobacterium smegmatis* nitroreductase NfnB, and its role in benzothiazinone resistance. *Mol Microbiol.* 77(5), 1172-1185.
- Mathias, J. R., Zhang, Z., Saxena, M. T. and Mumm, J. S. (2014). Enhanced cell-specific ablation in zebrafish using a triple mutant of *Escherichia coli* nitroreductase. *Zebrafish.* 11(2), 85-97.
- Matin, A. C., Barak, Y., Lynch, S. V., Ackerley, D. F., Thorne, S. H. and Contag, C. H. (2010). Nitroreductase Enzymes. US Patent 7,754,470.
- McCalla, D. R., Kaiser, C. and Green, M. H. (1978). Genetics of nitrofurazone resistance in *Escherichia coli*. *J Bacteriol.* 133(1), 10-16.
- Medico, E., Gambarotta, G., Gentile, A., Comoglio, P. M. and Soriano, P. (2001). A gene trap vector system for identifying transcriptionally responsive genes. *Nat Biotechnol.* 19(6), 579-582.
- Mildvan, A. S. (2004). Inverse thinking about double mutants of enzymes. *Biochemistry.* 43(46), 14517-14520.
- Miller, A. F., Park, J. T., Ferguson, K. L., Pitsawong, W. and Bommarius, A. S. (2018). Informing efforts to develop nitroreductase for amine production. *Molecules.* 23(2), e211.
- Morley, K. L. and Kazlauskas, R. J. (2005). Improving enzyme properties: When are closer mutations better? *Trends Biotechnol.* 23(5), 231-237.
- Moss, J. B., Koustubhan, P., Greenman, M., Parsons, M. J., Walter, I. and Moss, L. G. (2009). Regeneration of the pancreas in adult zebrafish. *Diabetes.* 58(8), 1844-1851.
- Mukherjee, A. and Rokita, S. E. (2015). Single amino acid switch between a flavin-dependent dehalogenase and nitroreductase. *J Am Chem Soc.* 137(49), 15342-15345.
- Nakamura, S. (1955). Structure of azomycin, a new antibiotic. *Pharm Bull.* 3(5), 379-383.
- Nam, H., Lewis, N. E., Lerman, J. A., Lee, D. H., Chang, R. L., Kim, D. and Palsson, B. O. (2012). Network context and selection in the evolution to enzyme specificity. *Science.* 337(6098), 1101-1104.
- O'Brien, P. J. and Herschlag, D. (1999). Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol.* 6(4), R91-R105.

- Oldfield, E. H., Ram, Z., Culver, K. W., Blaese, R. M. and Devroom, H. L. (1993). Gene therapy for the treatment of brain tumors using intra tumoral transduction with the thymidine kinase gene and intravenous ganciclovir. *Hum Gene Ther.* 4(1), 39-69.
- Olekhovich, I. N., Goodwin, A. and Hoffman, P. S. (2009). Characterization of the NAD(P)H oxidase and metronidazole reductase activities of the RdxA nitroreductase of *Helicobacter pylori*. *FEBS J.* 276(12), 3354-3364.
- Palmer, D. H., Mautner, V., Mirza, D., Oliff, S., Gerritsen, W., van der Sijp, J. R., Hubscher, S., Reynolds, G., Bonney, S., Rajaratnam, R., Hull, D., Horne, M., Ellis, J., Mountain, A., Hill, S., Harris, P. A., Searle, P. F., Young, L. S., James, N. D. and Kerr, D. J. (2004). Virus-directed enzyme prodrug therapy: intratumoral administration of a replication-deficient adenovirus encoding nitroreductase to patients with resectable liver cancer. *J Clin Oncol.* 22(9), 1546-1552.
- Pang, S. C., Wang, H. P., Zhu, Z. Y. and Sun, Y. H. (2015). Transcriptional activity and DNA methylation dynamics of the Gal4/UAS system in zebrafish. *Mar Biotechnol (NY).* 17(5), 593-603.
- Patrick, W. M., Firth, A. E. and Blackburn, J. M. (2003). User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng.* 16(6), 451-457.
- Paulish-Miller, T. E., Augostini, P., Schuyler, J. A., Smith, W. L., Mordechai, E., Adelson, M. E., Gyax, S. E., Secor, W. E. and Hilbert, D. W. (2014). *Trichomonas vaginalis* metronidazole resistance is associated with single nucleotide polymorphisms in the nitroreductase genes ntr4(Tv) and ntr6(Tv). *Antimicrob Agents Ch.* 58(5), 2938-2943.
- Pavelka, A., Chovancova, E. and Damborsky, J. (2009). HotSpot Wizard: a web server for identification of hot spots in protein engineering. *Nucleic Acids Res.* 37, W376-383.
- Pearson, R. D. and Hewlett, E. L. (1985). Niclosamide therapy for tapeworm infections. *Ann Intern Med.* 102(4), 550-551.
- Peterson, F. J., Mason, R. P., Hovsepian, J. and Holtzman, J. L. (1979). Oxygen-sensitive and oxygen-insensitive nitroreduction by *Escherichia coli* and rat hepatic microsomes. *J Biol Chem.* 254(10), 4009-4014.
- Pfeiffer, B. D., Ngo, T. T., Hibbard, K. L., Murphy, C., Jenett, A., Truman, J. W. and Rubin, G. M. (2010). Refinement of tools for targeted gene expression in *Drosophila*. *Genetics.* 186(2), 735-755.
- Pinto, A. L., Hellinga, H. W. and Caradonna, J. P. (1997). Construction of a catalytically active iron superoxide dismutase by rational protein design. *Proc Natl Acad Sci USA.* 94(11), 5562-5567.
- Pisharath, H., Rhee, J. M., Swanson, M. A., Leach, S. D. and Parsons, M. J. (2007). Targeted ablation of beta cells in the embryonic zebrafish pancreas using *E. coli* nitroreductase. *Mech Dev.* 124(3), 218-229.

Pitera, D. J., Paddon, C. J., Newman, J. D. and Keasling, J. D. (2007). Balancing a heterologous mevalonate pathway for improved isoprenoid production in *Escherichia coli*. *Metab Eng.* 9(2), 193-207.

Pitsawong, W., Hoben, J. P. and Miller, A. F. (2014). Understanding the broad substrate repertoire of nitroreductase based on its kinetic mechanism. *J Biol Chem.* 289(22), 15203-15214.

Poelwijk, F. J., Kiviet, D. J., Weinreich, D. M. and Tans, S. J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. *Nature.* 445(7126), 383-386.

Poelwijk, F. J., Tanase-Nicola, S., Kiviet, D. J. and Tans, S. J. (2011). Reciprocal sign epistasis is a necessary condition for multi-peaked fitness landscapes. *J Theor Biol.* 272(1), 141-144.

Portsmouth, D., Hlavaty, J. and Renner, M. (2007). Suicide genes for cancer therapy. *Mol Aspects Med.* 28(1), 4-41.

Prosser, G. A., Copp, J. N., Mowday, A. M., Guise, C. P., Syddall, S. P., Williams, E. M., Horvat, C. N., Swe, P. M., Ashoorzadeh, A., Denny, W. A., Smaill, J. B., Patterson, A. V. and Ackerley, D. F. (2013). Creation and screening of a multi-family bacterial oxidoreductase library to discover novel nitroreductases that efficiently activate the bio-reductive prodrugs CB1954 and PR-104A. *Biochem Pharmacol.* 85(8), 1091-1103.

Prosser, G. A., Copp, J. N., Syddall, S. P., Williams, E. M., Smaill, J. B., Wilson, W. R., Patterson, A. V. and Ackerley, D. F. (2010). Discovery and evaluation of *Escherichia coli* nitroreductases that activate the anti-cancer prodrug CB1954. *Biochem Pharmacol.* 79(5), 678-687.

Rabinovitch, A. and Suarez-Pinzon, W. L. (1998). Cytokines and their roles in pancreatic islet beta-cell destruction and insulin-dependent diabetes mellitus. *Biochem Pharmacol.* 55(8), 1139-1149.

Race, P. R., Lovering, A. L., Green, R. M., Ossor, A., White, S. A., Searle, P. F., Wrighton, C. J. and Hyde, E. I. (2005). Structural and mechanistic studies of *Escherichia coli* nitroreductase with the antibiotic nitrofurazone. Reversed binding orientations in different redox states of the enzyme. *J Biol Chem.* 280(14), 13256-13264.

Race, P. R., Lovering, A. L., White, S. A., Grove, J. I., Searle, P. F., Wrighton, C. W. and Hyde, E. I. (2007). Kinetic and structural characterisation of *Escherichia coli* nitroreductase mutants showing improved efficacy for the prodrug substrate CB1954. *Journal of Molecular Biology.* 368(2), 481-492.

Rainov, N. G. and Grp, G. I. S. (2000). A phase III clinical evaluation of herpes simplex virus type 1 thymidine kinase and ganciclovir gene therapy as an adjuvant to surgical resection and radiation in adults with previously untreated glioblastoma multiforme. *Hum Gene Ther.* 11(17), 2389-2401.

Rajagopalan, P. T. R., Lutz, S. and Benkovic, S. J. (2002). Coupling interactions of distal residues enhance dihydrofolate reductase catalysis: Mutational effects on hydride transfer rates. *Biochemistry.* 41(42), 12618-12628.

- Rajamuthiah, R., Fuchs, B. B., Conery, A. L., Kim, W., Jayamani, E., Kwon, B., Ausubel, F. M. and Mylonakis, E. (2015). Repurposing salicylanilide anthelmintic drugs to combat drug resistant *Staphylococcus aureus*. *PLoS One*. 10(4), e0124595.
- Ramos, J. L., Gonzalez-Perez, M. M., Caballero, A. and van Dillewijn, P. (2005). Bioremediation of polynitrated aromatic compounds: plants and microbes put up a fight. *Curr Opin Biotechnol*. 16(3), 275-281.
- Rasila, T. S., Pajunen, M. I. and Savilahti, H. (2009). Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Anal Biochem*. 388(1), 71-80.
- Reetz, M. T. and Carballeira, J. D. (2007). Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat Protoc*. 2(4), 891-903.
- Reetz, M. T., Kahakeaw, D. and Lohmer, R. (2008). Addressing the numbers problem in directed evolution. *ChemBiochem*. 9(11), 1797-1804.
- Reetz, M. T., Torre, C., Eipper, A., Lohmer, R., Hermes, M., Brunner, B., Maichele, A., Bocola, M., Arand, M., Cronin, A., Genzel, Y., Archelas, A. and Furstoss, R. (2004). Enhancing the enantioselectivity of an epoxide hydrolase by directed evolution. *Org Lett*. 6(2), 177-180.
- Rich, M. H. (2017). Discovery and directed evolution of nitroreductase enzymes for activation of prodrugs and PET imaging compounds. School of Biological Sciences, Victoria University of Wellington, Wellington.
- Rich, M. H., Sharrock, A. V., Hall, K. R., Ackerley, D. F. and MacKichan, J. K. (2018). Evaluation of NfsA-like nitroreductases from *Neisseria meningitidis* and *Bartonella henselae* for enzyme-prodrug therapy, targeted cell ablation, and dinitrotoluene bioremediation. *Biotechnol Lett*. 40(2), 359-367.
- Riefler, R. G. and Smets, B. F. (2000). Enzymatic reduction of 2,4,6-trinitrotoluene and related nitroarenes: Kinetics linked to one-electron redox potentials. *Environ Sci Technol*. 34(18), 3900-3906.
- Robins, K. J., Hooks, D. O., Rehm, B. H. A. and Ackerley, D. F. (2013). *Escherichia coli* NemA is an efficient chromate reductase that can be biologically immobilized to provide a cell free system for remediation of hexavalent chromium. *PLoS One*. 8(3), e59200.
- Roldan, M., Perez-Reinado, E., Castillo, F. and Moreno-Vivian, C. (2008). Reduction of polynitroaromatic compounds: the bacterial nitroreductases. *FEMS Microbiol Rev*. 32(3), 474-500.
- Rolfs, A., Montor, W. R., Yoon, S. S., Hu, Y., Bhullar, B., Kelley, F., McCarron, S., Jepson, D. A., Shen, B., Taycher, E., Mohr, S. E., Zuo, D., Williamson, J., Mekalanos, J. and Labaer, J. (2008). Production and sequence validation of a complete full length ORF collection for the pathogenic bacterium *Vibrio cholerae*. *Proc Natl Acad Sci USA*. 105(11), 4364-4369.
- Romero, P. A. and Arnold, F. H. (2009). Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol*. 10(12), 866-876.

- Rosano, G. L. and Ceccarelli, E. A. (2014). Recombinant protein expression in *Escherichia coli*: advances and challenges. *Front Microbiol.* 5, 172.
- Rossolini, G. M., Arena, F., & Giani, T. (2017). Mechanisms of antibacterial resistance. Cohen, J., Powderly, W. G. and Opal, S. M. (eds). *Infectious Diseases*. vol 2, 1181–1196. Elsevier Ltd.
- Ryan, A., Kaplan, E., Laurieri, N., Lowe, E. and Sim, E. (2011). Activation of nitrofurazone by azoreductases: multiple activities in one enzyme. *Sci Rep.* 1, 63.
- Sandstrom, A. G., Wikmark, Y., Engstrom, K., Nyhlen, J. and Backvall, J. E. (2012). Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proc Natl Acad Sci USA.* 109(1), 78-83.
- Schmidt-Dannert, C. and Arnold, F. H. (1999). Directed evolution of industrial enzymes. *Trends Biotechnol.* 17(4), 135-136.
- Selzer, T., Albeck, S. and Schreiber, G. (2000). Rational design of faster associating and tighter binding protein complexes. *Nat Struct Biol.* 7(7), 537-541.
- Shao, E., Lin, L., Chen, C., Chen, H., Zhuang, H., Wu, S., Sha, L., Guan, X. and Huang, Z. (2016). Loop replacements with gut-binding peptides in Cry1Ab domain II enhanced toxicity against the brown planthopper, *Nilaparvata lugens* (Stal). *Sci Rep.* 6, 20106.
- Sharrock, A. V. (2018). Engineering nitroreductase enzymes for improved cancer gene therapy and targeted cell ablation. PhD Thesis School of Biological Sciences, Victoria University of Wellington, Wellington.
- Shin, J. H. and Song, H. G. (2009). Nitroreductase II involved in 2,4,6-trinitrotoluene degradation: purification and characterization from *Klebsiella* sp Cl. *J Microbiol.* 47(5), 536-541.
- Siim, B. G., Denny, W. A. and Wilson, W. R. (1997). Nitro reduction as an electronic switch for bioreductive drug activation. *Oncol Res.* 9(6-7), 357-369.
- Smith, A. L., Erwin, A. L., Kline, T., Unrath, W. C. T., Nelson, K., Weber, A. and Howald, W. N. (2007). Chloramphenicol is a substrate for a novel nitroreductase pathway in *Haemophilus influenzae*. *Antimicrob Agents Chemother.* 51(8), 2820-2829.
- Smith, M. A. and Edwards, D. I. (1995). Redox potential and oxygen concentration as factors in the susceptibility of *Helicobacter pylori* to nitroheterocyclic drugs. *J Antimicrob Chemother.* 35(6), 751-764.
- Soballe, B. and Poole, R. K. (1999). Microbial ubiquinones: Multiple roles in respiration, gene regulation and oxidative stress management. *Microbiology.* 145, 1817-1830.
- Stratford, I. J., Williamson, C., Hoe, S. and Adams, G. E. (1981). Radiosensitizing and cytotoxicity studies with CB 1954 (2,4-dinitro-5-aziridinylbenzamide). *Radiat Res.* 88(3), 502-509.

- Su, J., Guise, C. P. and Wilson, W. R. (2013). FSL-61 is a 6-nitroquinolone fluorogenic probe for one-electron reductases in hypoxic cells. *Biochem J.* 452(1), 79-86.
- Surre, J., Saint-Ruf, C., Collin, V., Orenge, S., Ramjeet, M. and Matic, I. (2018). Strong increase in the autofluorescence of cells signals struggle for survival. *Sci Rep.* 8(1), 12088.
- Swe, P. M., Copp, J. N., Green, L. K., Guise, C. P., Mowday, A. M., Smaill, J. B., Patterson, A. V. and Ackerley, D. F. (2012). Targeted mutagenesis of the *Vibrio fischeri* flavin reductase FRase I to improve activation of the anticancer prodrug CB1954. *Biochem Pharmacol.* 84(6), 775-783.
- Taga, M. E., Larsen, N. A., Howard-Jones, A. R., Walsh, C. T. and Walker, G. C. (2007). BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature.* 446(7134), 449-453.
- Tharmalingam, N., Port, J., Castillo, D. and Mylonakis, E. (2018). Repurposing the anthelmintic drug niclosamide to combat *Helicobacter pylori*. *Sci Rep.* 8(1), 3701.
- Thomas, S. R., McTamney, P. M., Adler, J. M., Laronde-Leblanc, N. and Rokita, S. E. (2009). Crystal structure of iodotyrosine deiodinase, a novel flavoprotein responsible for iodide salvage in thyroid glands. *J Biol Chem.* 284(29), 19659-19667.
- Thorel, F., Nepote, V., Avril, I., Kohno, K., Desgraz, R., Chera, S. and Herrera, P. L. (2010). Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature.* 464(7292), 1149-1154.
- Tocher, J. H. and Edwards, D. I. (1994). Evidence for the direct interaction of reduced metronidazole derivatives with DNA bases. *Biochem Pharmacol.* 48(6), 1089-1094.
- Tokuriki, N., Jackson, C. J., Afriat-Jurnou, L., Wyganowski, K. T., Tang, R. M. and Tawfik, D. S. (2012). Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun.* 3, 1257.
- Toscano, M. D., Woycechowsky, K. J. and Hilvert, D. (2007). Minimalist active-site redesign: Teaching old enzymes new tricks. *Angew Chem Int Ed Engl.* 46(18), 3212-3236.
- Urasaki, A., Morvan, G. and Kawakami, K. (2006). Functional dissection of the Tol2 transposable element identified the minimal cis-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics.* 174(2), 639-649.
- Valiauga, B., Miseviciene, L., Rich, M. H., Ackerley, D. F., Sarlauskas, J. and Cenas, N. (2018). Mechanism of two-/four-electron reduction of nitroaromatics by oxygen-insensitive nitroreductases: The role of a non-enzymatic reduction step. *Molecules.* 23(7), e1672.
- Valiauga, B., Williams, E. M., Ackerley, D. F. and Cenas, N. (2017). Reduction of quinones and nitroaromatic compounds by *Escherichia coli* nitroreductase A (NfsA): Characterization of kinetics and substrate specificity. *Arch Biochem Biophys.* 614, 14-22.
- Van Dillewijn, P., Couselo, J. L., Corredoira, E., Delgado, A., Wittich, R. M., Ballester, A. and Ramos, J. L. (2008). Bioremediation of 2,4,6-trinitrotoluene by bacterial nitroreductase expressing transgenic aspen. *Environ Sci Technol.* 42(19), 7405-7410.

- Vass, S. O., Jarrom, D., Wilson, W. R., Hyde, E. I. and Searle, P. F. (2009). *E. coli NfsA*: An alternative nitroreductase for prodrug activation gene therapy in combination with CB1954. *Br J Cancer*. 100(12), 1903-1911.
- Venitt, S. and Levy, L. S. (1974). Mutagenicity of chromates in bacteria and its relevance to chromate carcinogenesis. *Nature*. 250(5466), 493-494.
- Vihtelic, T. S. and Hyde, D. R. (2000). Light-induced rod and cone cell death and regeneration in the adult albino zebrafish (*Danio rerio*) retina. *J Neurobiol*. 44(3), 289-307.
- Voigt, C. A., Kauffman, S. and Wang, Z. G. (2000). Rational evolutionary design: The theory of *in vitro* protein evolution. *Adv Protein Chem*. 55, 79-160.
- Walker, S. L., Ariga, J., Mathias, J. R., Coothankandaswamy, V., Xie, X. Y., Distel, M., Koster, R. W., Parsons, M. J., Bhalla, K. N., Saxena, M. T. and Mumm, J. S. (2012). Automated reporter quantification *in vivo*: High-throughput screening method for reporter-based assays in zebrafish. *PLoS One*. 7(1), e29916.
- Wan, H. Y., Korzh, S., Li, Z., Mudumana, S. P., Korzh, V., Jiang, Y. J., Lin, S. and Gong, Z. Y. (2006). Analyses of pancreas development by generation of gfp transgenic zebrafish using an exocrine pancreas-specific elastaseA gene promoter. *Exp Cell Res*. 312(9), 1526-1539.
- Wang, G. and Maier, R. J. (2004). An NADPH quinone reductase of *Helicobacter pylori* plays an important role in oxidative stress resistance and host colonization. *Infect Immun*. 72(3), 1391-1396.
- Wang, J., Panakova, D., Kikuchi, K., Holdway, J. E., Gemberling, M., Burris, J. S., Singh, S. P., Dickson, A. L., Lin, Y. F., Sabeh, M. K., Werdich, A. A., Yelon, D., Macrae, C. A. and Poss, K. D. (2011). The regenerative capacity of zebrafish reverses cardiac failure caused by genetic cardiomyocyte depletion. *Development*. 138(16), 3421-3430.
- Weinreich, D. M., Delaney, N. F., Depristo, M. A. and Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*. 312(5770), 111-114.
- Weinreich, D. M., Watson, R. A. and Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution*. 59(6), 1165-1174.
- Weinstein, B. M., Schier, A. F., Abdelilah, S., Malicki, J., SolnicaKrezel, L., Stemple, D. L., Stainier, D. Y. R., Zwartkruis, F., Driever, W. and Fishman, M. C. (1996). Hematopoietic mutations in the zebrafish. *Development*. 123, 303-309.
- White, D. T. and Mumm, J. S. (2013). The nitroreductase system of inducible targeted ablation facilitates cell-specific regenerative studies in zebrafish. *Methods*. 62(3), 232-240.
- White, D. T., Sengupta, S., Saxena, M. T., Xu, Q., Hanes, J., Ding, D., Ji, H. and Mumm, J. S. (2017). Immunomodulation-accelerated neuronal regeneration following selective rod photoreceptor cell ablation in the zebrafish retina. *Proc Natl Acad Sci USA*. 114(18), e3719-e3728.

Wiebe, L. I. (1999). Radiohalogenated nitroimidazoles for single-photon scintigraphic imaging of hypoxic tissues. Machulla, H. J. (eds). *Imaging of Hypoxia*. vol 33, 155-176. Springer.

Williams, E. M. (2013). Development of bacterial nitroreductase enzymes for noninvasive imaging in cancer gene therapy. PhD Thesis. School of Biological Sciences, Victoria University of Wellington, Wellington.

Williams, E. M., Little, R. F., Mowday, A. M., Rich, M. H., Chan-Hyams, J. V., Copp, J. N., Smaill, J. B., Patterson, A. V. and Ackerley, D. F. (2015). Nitroreductase gene-directed enzyme prodrug therapy: insights and advances toward clinical utility. *Biochem J.* 471(2), 131-153.

Wingler, L. M. and Cornish, V. W. (2011). Reiterative Recombination for the *in vivo* assembly of libraries of multigene pathways. *Proc Natl Acad Sci USA.* 108(37), 15135-15140.

Wu, D. M., Schneiderman, T., Burgett, J., Gokhale, P., Barthel, L. and Raymond, P. A. (2001). Cones regenerate from retinal stem cells sequestered in the inner nuclear layer of adult goldfish retina. *Invest Ophthalmol Vis Sci.* 42(9), 2115-2124.

Xi, Y., Noble, S. and Ekker, M. (2011). Modeling neurodegeneration in zebrafish. *Curr Neurol Neurosci Rep.* 11(3), 274-282.

Xiao, H., Bao, Z. and Zhao, H. (2015). High throughput screening and selection methods for directed enzyme evolution. *Ind Eng Chem Res.* 54(16), 4011-4020.

Xie, X., Mathias, J. R., Smith, M. A., Walker, S. L., Teng, Y., Distel, M., Koster, R. W., Sirotkin, H. I., Saxena, M. T. and Mumm, J. S. (2012). Silencer-delimited transgenesis: NRSE/RE1 sequences promote neural-specific transgene expression in a NRSF/REST-dependent manner. *BMC Biol.* 10, 93.

Yang, C. T. and Johnson, S. L. (2006). Small molecule-induced ablation and subsequent regeneration of larval zebrafish melanocytes. *Development.* 133(18), 3563-3573.

Yang, Y. Y., Lin, J. P. and Wei, D. Z. (2016). Heterologous overexpression and biochemical characterization of a nitroreductase from *Gluconobacter oxydans* 621H. *Mol Biotechnol.* 58(6), 428-440.

Yu, T. Y., Mok, K. C., Kennedy, K. J., Valton, J., Anderson, K. S., Walker, G. C. and Taga, M. E. (2012). Active site residues critical for flavin binding and 5,6-dimethylbenzimidazole biosynthesis in the flavin destructase enzyme BluB. *Protein Sci.* 21(6), 839-849.

Yunis, A. A. (1988). Chloramphenicol: relation of structure to activity and toxicity. *Annu Rev Pharmacol Toxicol.* 28, 83-100.

Zenno, S., Kobori, T., Tanokura, M. and Saigo, K. (1998). Purification and characterization of NfrA1, a *Bacillus subtilis* nitro/flavin reductase capable of interacting with the bacterial luciferase. *Biosci Biotechnol Biochem.* 62(10), 1978-1987.

Zenno, S., Koike, H., Tanokura, M. and Saigo, K. (1996). Gene cloning, purification, and characterization of NfsB, a minor oxygen-insensitive nitroreductase from *Escherichia coli*,

similar in biochemical properties to FRase I, the major flavin reductase in *Vibrio fischeri*. *J Biochem.* 120(4), 736-744.

Zhang, Q., Lambert, G., Liao, D., Kim, H., Robin, K., Tung, C. K., Pourmand, N. and Austin, R. H. (2011). Acceleration of emergence of bacterial antibiotic resistance in connected microenvironments. *Science.* 333(6050), 1764-1767.

Zhao, X. F., Ellingsen, S. and Fjose, A. (2009). Labelling and targeted ablation of specific bipolar cell types in the zebrafish retina. *BMC Neurosci.* 10(107).

Zhou, W. B. and Hildebrandt, F. (2012). Inducible podocyte injury and proteinuria in transgenic zebrafish. *J Am Soc Nephrol.* 23(6), 1039-1047.

Zhu, H. and Dean, R. A. (1999). A novel method for increasing the transformation efficiency of *Escherichia coli* application for bacterial artificial chromosome library construction. *Nucleic Acids Res.* 27(3), 910-911.

Appendices

Appendix A: Top chloramphenicol detoxifying NfsA variants

A.1 Example of raw IC_{50} data

Appendix A.1: Example raw IC_{50} data of NfsA 36_37 and 20_39 Data represents two technical repeats. The IC_{50} value (the concentration of prodrug at which 50% growth inhibition occurs) was calculated using a dose-response inhibition four-parameter variable slope equation in GraphPad Prism 8.0. IC_{50} values from individual biological repeats were averaged to give the values reported in Table 3.8. This method was applied to other IC_{50} experiments.

A.2 Reduction of chloramphenicol by purified His₆-tagged NfsA variants

Appendix A.2: SDS-PAGE gel of purified His-tagged NfsA variants 4 μg of each purified protein was loaded onto the gel with the appropriate volume of 5 x SDS loading dye. The approximate size of the monomeric form of NfsA including the His-tag is 29 kDa.

Appendix A.3: Michaelis-Menten kinetic plots of the top chloramphenicol detoxifying variants The initial rates of reduction at varying concentrations of chloramphenicol and 250 μM NADPH were monitored at 340 nm using an extinction coefficient of 12,440 $M^{-1}cm^{-1}$ and normalised for protein concentration. Reactions were measured in triplicate. Michaelis-Menten plots were generated in Graphpad 8.0 and apparent K_M and k_{cat} were calculated from these plots and shown in Table 3.9.

Appendix A.4: Michaelis-Menten kinetic plot of the reduction of p-benzoquinone by wild-type NfsA The initial rates of reduction at varying concentrations of p-benzoquinone and 250 μM NADPH were monitored at 340 nm using an extinction coefficient of 6,220 $M^{-1}cm^{-1}$ and normalised for protein concentration. Reactions were measured in triplicate. The Michaelis-Menten plot was generated in Graphpad 8.0 and apparent K_M and k_{cat} were calculated from these plots and shown in Table 3.10.

Appendix B: IC₅₀ values of all NfsA 36_37 and 20_39 intermediates

Variant	S41Y	H215C	T219Y	K222V	S224R	R225V	F227G	CM IC ₅₀ (μM)
127_SHTKSRF	-	-	-	-	-	-	-	4.8 ± 0.7
1_SHTKSRG	-	-	-	-	-	-	+	3.6 ± 1
3_SHTKSVF	-	-	-	-	-	+	-	13 ± 2
5_SHTKRRF	-	-	-	-	+	-	-	2.2 ± 0.2
9_SHTVSRF	-	-	-	+	-	-	-	4.6 ± 0.8
32_SCTKSRF	-	+	-	-	-	-	-	3.7 ± 0.3
118_YHTKSRF	+	-	-	-	-	-	-	5.1 ± 1
119_SHYKSRF	-	-	+	-	-	-	-	4.3 ± 0.7
2_SHTKSRG	-	-	-	-	-	+	+	17 ± 3
4_SHTKRRG	-	-	-	-	+	-	+	2.2 ± 0.3
7_SHTKRVF	-	-	-	-	+	+	-	14 ± 1
8_SHTVSRG	-	-	-	+	-	-	+	7.1 ± 2
11_SHTVSRF	-	-	-	+	-	+	-	13 ± 2
13_SHTVRRF	-	-	-	+	+	-	-	3.1 ± 0.3
16_SHYKSRG	-	-	+	-	-	-	+	4.4 ± 0.9
18_SHYKSRF	-	-	+	-	-	+	-	13 ± 2
20_SHYKRRF	-	-	+	-	+	-	-	3.5 ± 0.5
24_SHYVSRF	-	-	+	+	-	-	-	5.7 ± 1
31_SCTKSRF	-	+	-	-	-	-	+	6.1 ± 1
34_SCTKSVF	-	+	-	-	-	+	-	2.0 ± 0.4
36_SCTKRRF	-	+	-	-	+	-	-	2.5 ± 0.3
40_SCTVSRF	-	+	-	+	-	-	-	3.7 ± 0.6
48_SCYKSRF	-	+	+	-	-	-	-	4.2 ± 0.6
62_YHTKSRG	+	-	-	-	-	-	+	1.6 ± 0.3
64_YHTKSVF	+	-	-	-	-	+	-	20 ± 2
66_YHTKRRF	+	-	-	-	+	-	-	2.8 ± 0.2
70_YHTVSRF	+	-	-	+	-	-	-	6.6 ± 0.7
78_YHYKSRF	+	-	+	-	-	-	-	5.4 ± 0.7
93_YCTKSRF	+	+	-	-	-	-	-	5.3 ± 0.1
6_SHTKRVG	-	-	-	-	+	+	+	10 ± 2
10_SHTVSRG	-	-	-	+	-	+	+	20 ± 1
12_SHTVRRG	-	-	-	+	+	-	+	3.7 ± 0.3
15_SHTVRVF	-	-	-	-	+	+	-	15 ± 0.6
17_SHYKSRG	-	-	+	-	-	+	+	14 ± 3
19_SHYKRRG	-	-	+	-	+	-	+	3.8 ± 0.5
22_SHYKRVF	-	-	+	-	+	+	-	22 ± 4
23_SHYVSRG	-	-	+	+	-	-	+	7.7 ± 0.8
26_SHYVSRF	-	-	+	+	-	+	-	13 ± 3
28_SHYVRRF	-	-	+	+	+	-	-	5.5 ± 1
33_SCTKSRG	-	+	-	-	-	+	+	13 ± 2
35_SCTKRRG	-	+	-	-	+	-	+	2.8 ± 0.5
38_SCTKRVF	-	+	-	-	+	+	-	3.4 ± 0.1
39_SCTVSRG	-	+	-	+	-	-	+	6.4 ± 0.8
42_SCTVSRF	-	+	-	+	-	+	-	1.9 ± 0.2
44_SCTVRRF	-	+	-	+	+	-	-	3.0 ± 0.4
47_SCYKSRG	-	+	+	-	-	-	+	4.3 ± 0.9
50_SCYKSVF	-	+	+	-	-	+	-	4.6 ± 0.8
52_SCYKRRF	-	+	+	-	+	-	-	3.0 ± 0.5
56_SCYVSRF	-	+	+	+	-	-	-	3.6 ± 0.6
63_YHTKSRG	+	-	-	-	-	+	+	24 ± 5
65_YHTKRRG	+	-	-	-	+	-	+	4.2 ± 0.2
68_YHTKRVF	+	-	-	-	+	+	-	23 ± 2
69_YHTVSRG	+	-	-	+	-	-	+	9.7 ± 1
72_YHTVSRF	+	-	-	+	-	+	-	24 ± 3
74_YHTVRRF	+	-	-	+	+	-	-	2.3 ± 0.4
77_YHYKSRG	+	-	+	-	-	-	+	8.3 ± 1
80_YHYKSVF	+	-	+	-	-	+	-	30 ± 4
82_YHYKRRF	+	-	+	-	+	-	-	3.1 ± 0.6
86_YHYVSRF	+	-	+	+	-	-	-	6.1 ± 0.9
92_YCTKSRG	+	+	-	-	-	-	+	9.0 ± 0.9
95_YCTKSVF	+	+	-	-	-	+	-	17 ± 2
97_YCTKRRF	+	+	-	-	+	-	-	3.3 ± 0.3
101_YCTVSRF	+	+	-	+	-	-	-	7.3 ± 0.8
108_YCYKSRF	+	+	+	-	-	-	-	4.5 ± 0.6
14_SHTVRVG	-	-	-	+	+	+	+	20 ± 2

Variant	S41Y	H215C	T219Y	K222V	S224R	R225V	F227G	CM IC ₅₀ (μM)
21_SHYKRVG	-	-	+	-	+	+	+	14 ± 2
25_SHYVSRG	-	-	+	+	-	+	+	20 ± 4
27_SHYVRRG	-	-	+	+	+	-	+	5.1 ± 0.6
30_SHYVRVF	-	-	+	+	+	+	-	19 ± 2
37_SCTKRVG	-	+	-	-	+	+	+	16 ± 3
41_SCTVSRG	-	+	-	+	-	+	+	12 ± 3
43_SCTVRRG	-	+	-	+	+	-	+	3.7 ± 0.1
46_SCTVRVF	-	+	-	+	+	+	-	2.7 ± 0.1
49_SCYKSRG	-	+	+	-	-	+	+	15 ± 2
51_SCYKRRG	-	+	+	-	+	-	+	3.4 ± 0.4
54_SCYKRVF	-	+	+	-	+	+	-	12 ± 2
55_SCYVSRG	-	+	+	+	-	-	+	5.6 ± 0.6
58_SCYVSRF	-	+	+	+	-	+	-	3.8 ± 0.9
60_SCYVRRF	-	+	+	+	+	-	-	3.7 ± 0.7
67_YHTKRVG	+	-	-	-	+	+	+	25 ± 4
71_YHTVSRG	+	-	-	+	-	+	+	18 ± 1
73_YHTVRRG	+	-	-	+	+	-	+	4.5 ± 0.9
76_YHTVRVF	+	-	-	+	+	+	-	27 ± 3
79_YHYKSRG	+	-	+	-	-	+	+	36 ± 6
81_YHYKRRG	+	-	+	-	+	-	+	4.6 ± 0.5
84_YHYKRVF	+	-	+	-	+	+	-	20 ± 4
85_YHYVSRG	+	-	+	+	-	-	+	10 ± 1
88_YHYVSRF	+	-	+	+	-	+	-	27 ± 4
90_YHYVRRF	+	-	+	+	+	-	-	2.6 ± 0.5
94_YCTKSRG	+	+	-	-	-	+	+	29 ± 0.1
96_YCTKRRG	+	+	-	-	+	-	+	3.5 ± 0.4
99_YCTKRVF	+	+	-	-	+	+	-	22 ± 2
100_YCTVSRG	+	+	-	+	-	-	+	11 ± 2
103_YCTVSRF	+	+	-	+	-	+	-	19 ± 2
105_YCTVRRF	+	+	-	+	+	-	-	3.4 ± 0.3
107_YCYKSRG	+	+	+	-	-	-	+	10 ± 2
110_YCYKSVF	+	+	+	-	-	+	-	22 ± 3
112_YCYKRRF	+	+	+	-	+	+	-	2.9 ± 0.5
115_YCYVSRF	+	+	+	+	-	-	-	4.7 ± 1
29_SHYVRVG	-	-	+	+	+	+	+	32 ± 4
45_SCTVRVG	-	+	-	-	+	+	+	11 ± 3
53_SCYKRVG	-	+	+	-	+	+	+	24 ± 3
57_SCYVSRG	-	+	+	+	-	+	+	14 ± 2
59_SCYVRRG	-	+	+	+	+	-	+	4.2 ± 1
61_SCYVRVF	-	+	+	+	+	+	-	13 ± 3
75_YHTVRVG	+	-	-	+	+	+	+	28 ± 3
83_YHYKRVG	+	-	+	-	+	+	+	39 ± 7
87_YHYVSRG	+	-	+	+	-	+	+	33 ± 6
89_YHYVRRG	+	-	+	+	+	-	+	5.6 ± 0.6
91_YHYVRVF	+	-	+	+	+	+	-	44 ± 3
98_YCTKRVG	+	+	-	-	+	+	+	28 ± 2
102_YCTVSRG	+	+	-	+	-	+	+	31 ± 3
104_YCTVRRG	+	+	-	+	+	-	+	4.2 ± 0.8
106_YCTVRVF	+	+	-	+	+	+	-	23 ± 1
111_YCYKRRG	+	+	+	-	+	-	+	5.5 ± 0.5
113_YCYKRVF	+	+	+	-	+	+	-	41 ± 6
114_YCYVSRG	+	+	+	+	-	-	+	23 ± 5
116_YCYVSRF	+	+	+	+	-	+	-	38 ± 3
117_YCYVRRF	+	+	+	+	+	-	-	3.0 ± 0.5
124_YCYKSRG	+	+	+	-	-	+	+	30 ± 3
120_YHYVRVG	-	-	+	+	+	+	+	36 ± 6
121_YCTVRVG	+	+	-	+	+	+	+	29 ± 6
122_YCYKRVG	+	+	+	-	+	+	+	39 ± 4
123_YCYVSRG	+	+	+	+	-	+	+	37 ± 3
125_YCYVRRG	+	+	+	+	+	-	+	6.2 ± 1
126_YCYVRVF	+	+	+	+	+	+	-	42 ± 5
129_SCYVRVG	-	+	+	+	+	+	+	19 ± 3
128_YCYVRVG	+	+	+	+	+	+	+	44 ± 2
pUCX_empty								0.6 ± 1

Appendix B.1: Chloramphenicol IC₅₀ values of the 128 intermediates of NfsA 36_37. Data represents the average of at least four biological repeats ± SD. Each presence (+) or absence (-) of each of the seven substitutions is indicated for each intermediate.

Variant	S41Y	H215C	T219Y	K222V	S224R	R225V	F227G	CM IC ₅₀ (μM)
127_SHTKSRF	-	-	-	-	-	-	-	4.4 ± 0.7
113_YHTKSRF	+	-	-	-	-	-	-	4.8 ± 0.9
114_SNTKSRF	-	+	-	-	-	-	-	4.0 ± 0.6
115_SHYKSRF	-	-	+	-	-	-	-	4.3 ± 0.4
116_SHTSRF	-	-	-	+	-	-	-	4.3 ± 0.7
117_SHTKYRF	-	-	-	-	+	-	-	4.3 ± 0.6
118_SHTKSDF	-	-	-	-	-	+	-	12 ± 2
119_SHTKSRH	-	-	-	-	-	-	+	3.8 ± 0.3
1_SNTKSRH	-	+	-	-	-	-	+	3.8 ± 0.7
3_SNTKSDF	-	+	-	-	-	+	-	1.4 ± 0.3
5_SNTKYRF	-	+	-	-	+	-	-	2.5 ± 0.3
9_SNTRSRF	-	+	-	+	-	-	-	3.7 ± 0.6
17_SNYKSRF	-	+	+	-	-	-	-	3.1 ± 0.4
31_SHTKSDH	-	-	-	-	-	+	+	17 ± 2
32_SHTKYRH	-	-	-	-	+	-	+	4.0 ± 0.6
34_SHTKYDF	-	-	-	-	+	+	-	7.1 ± 1
35_SHTSRH	-	-	-	+	-	-	+	3.7 ± 0.6
37_SHTSRDF	-	-	-	+	-	+	-	9.4 ± 0.6
39_SHTRYRF	-	-	-	+	+	-	-	3.6 ± 0.7
42_SHYKSRH	-	-	+	-	-	-	+	3.5 ± 0.6
44_SHYKSDF	-	-	+	-	-	+	-	13 ± 3
46_SHYKYRF	-	-	+	-	+	-	-	4.1 ± 0.8
50_SHYRSRF	-	-	+	+	-	-	-	3.9 ± 0.8
58_YNTKSRF	+	+	-	-	-	-	-	5.2 ± 0.5
83_YHTKSRH	+	-	-	-	-	-	+	4.0 ± 0.3
85_YHTKSDF	+	-	-	-	-	+	-	18 ± 4
87_YHTKYRF	+	-	-	-	+	-	-	3.1 ± 0.6
91_YHTSRF	+	-	-	+	-	-	-	4.2 ± 0.8
99_YHYKSRF	+	-	+	-	-	-	-	4.7 ± 0.8
2_SNTKSDH	-	+	-	-	-	+	+	4.7 ± 0.7
4_SNTKYRH	-	+	-	-	+	-	+	3.7 ± 0.7
7_SNTKYDF	-	+	-	-	+	+	-	1.3 ± 0.1
8_SNTRSRH	-	+	-	+	-	-	+	3.5 ± 0.9
11_SNTRSDF	-	+	-	+	-	+	-	1.3 ± 0.2
13_SNTRYRF	-	+	-	+	+	-	-	2.0 ± 0.3
16_SNYKSRH	-	+	+	-	-	-	+	2.9 ± 0.4
19_SNYKSDF	-	+	+	-	-	+	-	3.6 ± 0.7
21_SNYKYRF	-	+	+	-	+	-	-	3.2 ± 0.3
25_SNYRSRF	-	+	+	+	-	-	-	2.7 ± 0.3
33_SHTKYDH	-	-	-	-	+	+	+	13 ± 2
36_SHTSRDH	-	-	-	+	-	+	+	18 ± 2
38_SHTRYRH	-	-	-	+	+	-	+	4.3 ± 0.6
41_SHTRYDF	-	-	-	+	+	+	-	6.3 ± 0.7
43_SHYKSDH	-	-	+	-	-	+	+	15 ± 3
45_SHYKYRH	-	-	+	-	+	-	+	4.7 ± 0.4
48_SHYKYDF	-	-	+	-	+	+	-	8.7 ± 0.4
49_SHYRSRH	-	-	+	+	-	-	+	3.5 ± 0.3
52_SHYRSDF	-	-	+	+	-	+	-	12 ± 2
54_SHYRYRF	-	-	+	+	+	-	-	4.4 ± 0.8
57_YNTKSRH	+	+	-	-	-	-	+	4.1 ± 0.5
60_YNTKSDF	+	+	-	-	-	+	-	11 ± 2
62_YNTKYRF	+	+	-	-	+	-	-	3.7 ± 0.7
66_YNTRSRF	+	+	-	+	-	-	-	5.2 ± 0.5
73_YNYKSRF	+	+	+	-	-	-	-	3.6 ± 0.7
84_YHTKSDH	+	-	-	-	-	+	+	24 ± 5
86_YHTKYRH	+	-	-	-	+	-	+	3.0 ± 0.6
89_YHTKYDF	+	-	-	-	+	+	-	16 ± 2
90_YHTSRH	+	-	-	+	-	-	+	4.0 ± 0.8
93_YHTSRDF	+	-	-	+	-	+	-	16 ± 1
95_YHTRYRF	+	-	-	+	+	-	-	2.9 ± 0.9
98_YHYKSRH	+	-	+	-	-	-	+	3.7 ± 0.7
101_YHYKSDF	+	-	+	-	-	+	-	24 ± 6
103_YHYKYRF	+	-	+	-	+	-	-	4.8 ± 0.8
107_YHYRSRF	+	-	+	+	-	-	-	4.7 ± 0.5
6_SNTKYDH	-	+	-	-	+	+	+	2.5 ± 0.4

Variant	S41Y	H215C	T219Y	K222V	S224R	R225V	F227G	CM IC ₅₀ (μM)
10_SNTRSDH	-	+	-	+	-	+	+	3.1 ± 0.6
12_SNTRYRH	-	+	-	+	+	-	+	3.6 ± 0.6
15_SNTRYDF	-	+	-	+	+	+	-	1.0 ± 0.3
18_SNYKSDH	-	+	+	-	-	+	+	12.6 ± 2
20_SNYKYRH	-	+	+	-	+	-	+	3.7 ± 0.3
23_SNYKYDF	-	+	+	-	+	+	-	1.7 ± 0.4
24_SNYRSRH	-	+	+	+	-	-	+	2.8 ± 0.5
27_SNYRSDF	-	+	+	+	-	+	-	3.3 ± 0.6
29_SNYRYRF	-	+	+	+	+	-	-	2.9 ± 0.4
40_SHTRYDH	-	-	-	+	+	+	+	12 ± 1
47_SHYKYDH	-	-	+	-	+	+	+	15 ± 1
51_SHYRSDH	-	-	+	+	-	+	+	21 ± 2
53_SHYRYRH	-	-	+	+	+	-	+	4.5 ± 1
56_SHYRYDF	-	-	+	+	+	+	-	7.5 ± 3
59_YNTKSDH	+	+	-	-	-	+	+	25 ± 4
61_YNTKYRH	+	+	-	-	-	-	+	4.1 ± 0.7
64_YNTKYDF	+	+	-	-	+	+	-	11 ± 0.8
65_YNTRSRH	+	+	-	+	-	-	+	4.4 ± 0.6
68_YNTRSDF	+	+	-	+	-	+	-	10 ± 1
70_YNTRYRF	+	+	-	+	+	-	-	3.9 ± 1
72_YNYKSRH	+	+	+	-	-	-	+	2.7 ± 0.4
75_YNYKSDF	+	+	+	-	-	+	-	28 ± 3
77_YNYKYRF	+	+	+	-	+	-	-	4.4 ± 0.7
80_YNYRSRF	+	+	+	+	-	-	-	4.1 ± 0.8
88_YHTKYDH	+	-	-	-	+	+	+	19 ± 2.3
92_YHTRSDH	+	-	-	+	-	+	+	23 ± 4
94_YHTRYRH	+	-	-	+	+	-	+	3.7 ± 0.9
97_YHTRYDF	+	-	-	+	+	+	-	15 ± 0.7
100_YHYKSDH	+	-	+	-	-	-	+	31 ± 3
102_YHYKYRH	+	-	+	-	+	-	+	4.7 ± 1
105_YHYKYDF	+	-	+	-	+	+	-	19 ± 2
106_YHYRSRH	+	-	+	+	-	-	+	3.6 ± 0.6
109_YHYRSDF	+	-	+	+	-	+	-	21 ± 2
111_YHYRYRF	+	-	+	+	+	-	-	5.1 ± 1
14_SNTRYDH	-	+	-	+	+	+	+	1.9 ± 0.4
22_SNYKYDH	-	+	+	-	+	+	+	8.0 ± 2
26_SNYRSDH	-	+	+	+	-	+	+	9.2 ± 0.8
28_SNYRYRH	-	+	+	+	+	-	+	3.0 ± 0.5
30_SNYRYDF	-	+	+	+	+	+	-	1.5 ± 0.1
55_SHYRYDH	-	-	+	+	+	+	+	15 ± 1
63_YNTKYDH	+	+	-	-	+	+	+	20 ± 2
67_YNTRSDH	+	+	-	+	-	+	+	25 ± 3
69_YNTRYRH	+	+	-	+	+	-	+	4.5 ± 0.9
71_YNTRYDF	+	+	-	+	+	+	-	12 ± 2
74_YNYKSDH	+	+	+	-	-	+	+	42 ± 7
76_YNYKYRH	+	+	+	-	+	-	+	3.5 ± 0.7
78_YNYKYDF	+	+	+	-	+	+	-	25 ± 2.4
79_YNYRSRH	+	+	+	+	+	-	+	2.9 ± 0.6
81_YNYRSDF	+	+	+	+	-	+	-	24 ± 3
82_YNYRYRF	+	+	+	+	+	-	-	3.6 ± 0.4
96_YHTRYDH	+	-	-	+	+	+	+	20 ± 4
104_YHYKYDH	+	-	+	-	+	+	+	32 ± 6
108_YHYRSDH	+	-	+	+	-	+	+	30 ± 2
110_YHYRYRH	+	-	+	+	+	-	+	4.1 ± 0.4
112_YHYRYDF	+	-	+	+	+	+	-	20 ± 2
120_SNYRYDH	-	+	+	+	+	+	+	5.1 ± 1
121_YHYRYDH	+	-	+	+	+	+	+	33 ± 5
122_YNTRYDH	+	+	-	+	+	+	+	17 ± 2
123_YNYKYDH	+	+	+	-	+	+	+	36 ± 6
124_YNYRSDH	+	+	+	+	-	+	+	41 ± 4
125_YNYRYRH	+	+	+	+	+	-	+	3.1 ± 0.2
126_YNYRYDF	+	+	+	+	+	+	-	24 ± 3
128_YNYRYDH	+	+	+	+	+	+	+	42 ± 7
pUCX_empty								0.6 ± 1

Appendix B.2: Chloramphenicol IC₅₀ values of the 128 intermediates of NfsA 20_39. Data represents the average of at least four biological repeats ± SD. Each presence (+) or absence (-) of each of the seven substitutions is indicated for each intermediate.

Appendix C: Growth inhibition assay and sequence data of NfsA variants recovered from various selection conditions

Growth inhibition /Growth (%)	antibiotic								Amino acid makes up >20% of amino acids at this position						
	75 μM metronidazole	18 μM RB6145	250 μM CB1954	27.5 μM nitrofurazone	375 μM 2,4-DNT	4 μM chloramphenicol	4 μM niclosamide								
19_2	100.0	26.8	87.7	89.2	81.8	93.2	92.6	Y	L	D	Y	Q	V	G	H
19_3	100.0	30.4	88.7	81.2	71.2	93.7	96.0	Y	L	D	V	V	H	L	R
19_4	100.0	59.3	94.2	72.4	94.5	92.4	91.0	F	V	I	Y	R	D	N	D
19_5	100.0	42.3	90.8	83.8	87.4	96.3	96.4	Y	V	H	V	V	R	L	D
19_6	100.0	35.0	94.5	68.6	97.1	96.9	96.5	Y	L	S	Y	L	H	C	D
19_7	100.0	45.3	94.4	86.6	93.7	89.8	87.5	Y	I	G	Y	V	H	C	D
19_8	100.0	38.9	74.5	83.4	51.9	94.2	91.7	Y	L	D	V	L	L	S	G
19_9	100.0	36.8	87.8	80.1	66.7	90.8	91.7	Y	S	D	F	K	I	N	F
19_10	100.0	32.3	95.2	75.8	99.3	95.2	94.3	Y	L	G	Y	V	S	C	G
19_11	100.0	41.6	90.5	74.1	74.4	92.4	95.9	Y	I	L	P	L	N	N	H
19_12	100.0	19.9	76.3	78.3	58.7	94.2	93.0	Y	C	D	F	E	I	N	F
19_13	100.0	61.5	94.7	99.9	99.8	87.7	86.1	C	L	I	V	L	V	N	C
19_14	100.0	35.4	89.2	81.6	71.3	91.8	92.3	Y	L	D	F	R	H	C	R
19_15	100.0	26.7	94.7	76.1	87.8	93.9	88.2	Y	L	D	Y	Q	G	N	G
19_16	100.0	45.2	96.2	79.5	100.0	92.8	91.3	F	S	H	F	L	F	N	H
19_17	100.0	42.4	95.2	85.2	100.0	95.2	91.4	Y	L	N	Y	V	R	C	F
19_19	100.0	18.9	92.6	76.3	73.4	98.7	94.9	Y	L	D	F	V	H	C	S
19_20	100.0	23.4	91.0	86.0	94.5	93.6	88.5	Y	L	D	V	G	L	N	G
19_21	99.6	20.1	96.9	78.7	96.9	95.3	90.8	H	L	D	Y	E	H	N	C
19_22	100.0	57.0	98.3	94.7	100.0	96.3	91.6	Y	S	V	F	Q	N	N	C
19_23	100.0	39.8	89.3	73.5	85.8	94.2	93.3	Y	S	H	L	V	H	L	D
19_24	100.0	70.2	95.6	99.0	100.0	96.2	91.9	Y	C	F	F	E	G	H	G
19_25	99.7	33.0	95.4	82.0	100.0	95.9	95.0	F	S	H	F	M	I	N	H
19_26	100.0	31.2	95.0	89.5	99.1	89.3	88.5	Y	L	D	F	M	I	N	Y
19_27	99.5	53.0	95.4	86.8	100.0	94.1	94.3	Y	L	G	Y	V	R	N	I
19_29	100.0	49.7	96.7	82.3	100.0	95.4	94.2	Y	S	H	F	L	V	N	D
19_31	99.8	81.6	94.4	100.0	100.0	98.3	94.0	H	L	S	H	G	V	C	H
19_32	100.0	15.7	93.4	76.1	97.5	95.6	95.7	Y	L	D	F	L	I	N	S
19_33	99.5	38.8	96.2	86.0	100.0	95.4	92.7	Y	V	H	F	L	Y	N	G
19_34	100.0	27.9	89.2	82.4	67.9	89.2	85.8	Y	L	D	F	A	V	C	N
19_35	99.7	48.6	93.1	72.0	77.1	87.3	86.5	Y	S	N	Y	L	H	N	H
19_37	100.0	36.6	96.1	77.4	94.7	89.7	86.3	Y	S	H	L	E	N	L	D
19_38	99.7	80.4	86.7	77.4	68.4	87.5	87.4	H	L	C	Y	A	R	C	C
19_39	100.0	70.5	97.4	97.5	100.0	92.6	93.6	F	L	G	L	M	Y	N	G
19_40	99.4	51.8	91.6	78.8	100.0	94.8	91.9	Y	L	H	I	M	R	L	D
19_41	98.2	63.7	94.5	95.6	98.8	92.5	88.8	Y	I	V	Y	G	V	N	R
19_43	97.8	30.2	96.5	73.6	99.7	96.1	88.5	Y	L	G	Y	E	R	D	H
19_44	99.1	43.7	93.8	73.9	97.9	96.8	87.2	Y	S	H	F	M	L	N	H
19_45	99.4	21.0	83.5	80.2	76.2	96.8	91.4	Y	L	D	I	P	N	L	D
19_48	99.5	45.6	82.3	93.6	73.9	92.5	91.4	Y	C	D	F	K	G	H	D
19_49	98.5	29.6	92.2	95.2	98.2	95.6	96.0	Y	L	D	F	K	I	S	G
19_50	100.0	40.6	96.7	85.6	100.0	91.1	89.5	Y	S	N	F	T	V	N	G
19_51	99.3	53.7	98.6	95.9	100.0	92.9	87.9	F	V	F	F	T	C	N	G
19_52	100.0	63.4	98.7	96.6	100.0	91.3	89.3	H	I	C	Y	V	V	N	H
19_54	99.8	28.3	81.0	72.4	53.7	92.8	85.8	Y	L	D	L	M	L	C	N
19_55	100.0	53.0	99.4	90.6	100.0	97.9	96.3	Y	C	F	F	Q	S	G	N
19_56	99.3	79.7	97.8	97.0	100.0	93.3	88.6	Y	C	G	Y	R	D	N	F
19_53	99.3	27.0	79.0	74.0	69.2	89.5	90.4	Y	L	D	I	T	H	L	D
19_58	99.1	62.6	95.5	84.6	100.0	92.2	91.5	Y	L	H	Y	V	R	L	H
20_2	99.7	60.8	98.0	99.8	100.0	95.5	95.4	H	I	C	Y	V	V	H	H
20_3	98.4	46.0	94.0	94.6	100.0	94.6	94.8	Y	L	D	F	T	R	L	S
20_4	100.0	80.8	98.2	92.3	100.0	90.7	85.8	Y	C	H	F	E	R	N	D
20_5	98.8	51.0	93.3	86.6	100.0	99.9	92.8	Y	C	F	F	Q	H	C	C
20_6	98.4	43.0	91.3	64.2	97.9	92.8	92.3	Y	L	C	Y	V	R	D	H
20_7	98.4	11.8	84.9	64.4	91.8	96.7	89.7	Y	L	D	I	G	N	L	D
20_8	99.2	38.7	94.8	68.2	94.2	96.4	94.7	Y	I	G	Y	V	R	D	H
20_9	98.5	11.1	93.7	66.2	99.6	96.7	95.9	Y	L	D	F	R	V	N	C
pUCX_empty	3.4	1.5	5.5	6.1	7.0	5.8	0.0								
NfsA	49.0	42.4	57.4	42.0	44.7	53.1	72.1								
NfsA	55.7	52.5	59.4	63.3	62.1	39.6	71.5								

Table C.1: Percentage growth inhibition and amino acid substitutions for NfsA variants selected on 45 μ M chloramphenicol. Data represent the average of four biological repeats, standard deviation is omitted for clarity (standard error of the mean less than 5). The amino acid present at each of the substituted positions is shown. An amino acid is shaded if it makes up 20% of all amino acids at any given position.

Growth inhibition /Growth (%)	75 μM metronidazole	18 μM RB6145	250 μM CB1954	27.5 μM nitrofurazone	375 μM 2,4-DNT	4 μM chloramphenicol	4 μM niclosamide	Amino acid makes up >20% of amino acids at this position									
								S41	L43	H215	T219	K222	S224	R225	F227		
R2	21.9	13.4	35.0	68.9	31.3	22.1	70.9	Y	S	S	P	R	F	F	D		
R3	12.4	10.5	30.2	65.2	22.1	12.1	47.3	Y	S	C	V	L	R	F	V		
R4	31.5	7.9	30.1	70.8	28.6	25.3	51.4	Y	C	R	F	A	Y	L	I		
R5	61.4	16.2	50.8	67.0	33.4	35.9	86.4	Y	N	R	N	G	N	Y	V		
R6	9.6	6.9	20.2	56.7	31.4	10.8	62.8	F	F	R	P	G	R	G	S		
R7	17.2	16.5	41.3	73.6	32.2	6.7	26.4	Y	G	S	I	A	R	C	F		
R8	13.2	12.9	30.9	68.8	25.1	28.9	31.0	Y	C	R	I	L	L	S	F		
R9	56.6	34.4	32.7	60.4	20.8	35.3	83.1	Y	S	D	V	M	F	V	D		
R10	61.2	23.5	33.1	71.5	23.2	33.2	86.7	Y	C	G	P	A	C	C	S		
R11	19.7	13.9	33.9	65.9	28.7	10.5	67.1	Y	V	R	P	R	S	C	G		
R12	9.5	2.8	30.3	57.3	25.3	16.4	38.2	Y	N	S	D	R	R	Y	L		
R13	19.9	14.6	41.1	77.5	34.1	11.3	15.0	Y	H	R	H	T	H	V	L		
R14	24.7	27.8	36.4	83.9	34.1	10.4	68.2	Y	C	V	P	M	R	Y	G		
R15	45.3	29.8	33.8	89.4	28.4	11.3	50.0	Y	S	F	F	R	C	V	V		
R16	53.3	15.4	44.5	82.1	37.1	19.3	86.8	Y	S	G	P	R	R	C	D		
R17	32.2	9.4	37.2	77.4	23.8	16.2	28.5	Y	V	R	V	T	C	G	I		
R18	13.4	8.0	36.8	76.0	34.7	19.8	34.3	Y	Y	C	P	R	G	L	F		
R19	51.8	27.2	59.0	87.6	45.0	12.5	69.6	Y	N	R	A	R	H	Y	H		
R23	55.8	26.3	64.1	82.8	59.2	10.0	70.9	F	N	V	N	K	Y	L	G		
R24	47.5	41.1	44.9	77.7	34.3	18.7	81.9	Y	C	C	P	K	G	C	R		
R25	60.2	17.0	47.0	78.9	35.1	21.6	52.1	Y	N	I	L	R	N	G	L		
R28	42.8	18.0	51.1	59.3	33.0	20.2	83.6	Y	N	R	V	T	N	D	G		
R30	48.6	12.2	48.1	71.5	36.5	30.5	93.8	Y	S	R	P	K	G	G	D		
R31	35.6	29.6	55.6	83.9	48.6	7.6	58.9	Y	L	R	P	R	F	H	I		
R32	31.5	33.2	47.2	90.1	39.9	16.6	72.4	Y	L	G	P	R	C	C	L		
R33	66.2	36.1	44.2	89.8	36.4	39.7	86.2	Y	L	R	V	R	S	L	I		
R34	38.9	32.3	44.3	84.2	42.2	21.8	60.2	Y	C	C	D	L	R	H	I		
R35	26.7	22.7	43.3	77.1	36.9	9.3	77.6	Y	I	R	P	P	N	Y	L		
R36	30.2	15.8	32.9	48.6	20.3	39.5	73.3	Y	C	R	H	L	H	C	V		
R37	63.1	39.9	59.8	84.5	45.5	17.5	76.4	Y	S	C	A	K	R	S	F		
R38	54.4	23.4	46.4	89.3	41.6	17.2	80.3	Y	F	R	P	P	C	G	R		
R39	43.4	31.4	39.6	83.3	27.4	19.5	39.4	Y	C	Y	H	G	S	L	I		
R40	32.9	17.7	41.1	77.2	38.0	30.5	68.1	Y	S	C	D	V	N	Y	G		
R41	68.8	29.1	53.0	81.8	38.6	34.6	88.2	Y	N	S	H	L	V	Y	D		
R42	54.7	18.3	50.9	83.4	45.4	26.0	91.6	Y	N	Y	P	A	H	F	D		
R44	17.2	24.7	41.7	80.5	29.0	8.7	17.5	Y	C	R	A	M	Y	F	S		
R45	58.9	43.5	53.9	85.7	45.1	16.1	84.2	Y	L	C	P	R	N	V	I		
R46	25.4	20.4	37.4	44.9	27.3	23.8	41.0	C	S	V	F	E	H	D	C		
R47	16.7	18.9	39.8	83.2	39.9	11.8	54.6	Y	Y	F	P	R	H	S	V		
R48	20.4	14.1	41.0	80.3	31.6	17.0	24.6	Y	H	N	H	A	R	L	V		
R51	6.8	2.9	26.8	77.9	22.7	14.9	9.7	Y	Y	V	F	G	N	V	F		
R52	55.9	22.2	49.3	76.1	43.3	21.4	80.0	Y	D	D	D	D	D	G	S		
R53	44.0	28.4	55.4	83.1	37.2	14.1	57.7	Y	C	C	C	C	C	N	V		
R54	17.3	18.6	41.2	82.0	28.7	17.5	30.7	Y	F	F	F	F	F	S	G		
R55	39.8	13.7	18.4	72.2	14.6	32.4	47.1	Y	C	C	C	C	C	V	I		
R58	58.6	14.5	38.8	79.8	28.1	26.1	89.9	Y	F	R	F	V	R	D	I		
R60	25.9	28.2	29.4	73.9	22.9	16.6	36.1	Y	V	G	Y	R	V	F	I		
R61	28.8	19.7	41.3	75.5	37.1	8.7	50.7	Y	Y	R	P	R	G	G	F		
R62	67.8	24.1	50.5	74.5	40.9	42.5	74.8	Y	S	C	D	A	G	F	H		
R63	26.3	20.3	47.7	77.8	38.3	13.9	78.4	Y	S	R	P	R	F	Y	D		
R65	38.7	23.2	33.1	74.7	27.7	25.2	72.8	Y	C	V	P	E	F	F	D		
R66	1.6	2.5	13.9	68.0	20.3	4.2	9.1	Y	V	L	P	K	C	F	C		
R67	5.1	8.3	37.8	77.5	42.0	1.7	16.8	F	F	R	P	R	V	N	Y		
R68	36.0	27.1	41.3	73.6	36.9	28.6	68.8	F	S	S	P	P	F	N	R		
R69	63.6	21.6	52.5	75.4	42.6	30.2	76.6	F	N	R	F	L	R	S	F		
R70	59.5	37.4	40.4	82.4	29.8	39.3	84.7	Y	S	R	N	P	L	V	S		
R72	46.9	34.6	34.2	75.5	27.9	29.5	57.7	Y	Y	D	Y	G	R	H	F		
pUCX_empty	4.4	2.0	15.8	12.8	9.3	2.4	0.0										
NfsA	41.4	31.5	36.4	39.9	42.1	37.4	76.9										
NfsA	44.4	44.2	42.4	44.7	40.0	36.2	74.5										

Table C.5 Percentage growth inhibition and amino acid substitutions of NfsA variants selected on 1 μ M niclosamide + 40 μ M RB6145. Data represent the average of four biological repeats, standard deviation is omitted for clarity (standard error of the mean less than 5). The amino acid present at each of the substituted positions is shown. An amino acid is shaded if it makes up $\geq 20\%$ of all amino acids at any given position.

Appendix D: Michaelis-Menten plots of wild-type NfsA with nitroaromatic compounds.

Appendix D.1: Michaelis-Menten plots for the reduction of CB1954, nitrofurazone, 2,4-DNT, RB6145 and metronidazole by His₆-tagged NfsA. The initial rate of reduction across 0-800 μM CB1954 and 250 μM NADPH was monitored at 420 nm. The initial rate of reduction across 0-200 μM nitrofurazone, 0-400 μM 2,4-DNT and 0-1200 μM RB6145 at 250 μM NADPH was monitored at 340 nm. The initial rate of reduction across 0-800 μM metronidazole was determined using the NADPH regeneration system and monitoring of metronidazole depletion at 340 nm. Data is the average of three technical replicates.

Appendix E: Development of a transient cell ablation assay in zebrafish

In order for a transient cell ablation assay to work, it would require broad transient expression, capable of detection on a plate reader in order for cell ablation to be quantified. We reasoned that expression using the standard $5 \times$ UAS reporter construct injected into the gmc617 driver line (described in Section 6.3.1) would not be sufficient, as expression often occurs in only a few individual cells. Several methods attempting to broaden the transient expression were trialled.

E.1 Increasing the UAS copy number.

Changing the UAS copy number in the Gal4/UAS expression system can be used to modulate expression of downstream elements (Pfeiffer *et al.*, 2010). An increased UAS copy number can cause problems when generating stably expressing transgenic zebrafish lines. Higher copy number UAS constructs, such as $14 \times$ UAS, are susceptible to trans-generation silencing, as tandem repetitive sequences can attract DNA methylation (Goll *et al.*, 2009; Akitake *et al.*, 2011; Pang *et al.*, 2015). This is why $5 \times$ UAS constructs are traditionally used when generating stably expressing transgenic zebrafish lines. However, due to the short time span of a transient ablation assay (0-6 dpf), we considered it unlikely that genetic silencing would occur. pTOL2 TagYFP:NTR constructs containing a modified $14 \times$ UAS element were injected into eggs from a gmc617 driver line. Brighter expression was evident in the $14 \times$ UAS-injected larvae, however this expression was not widespread enough to be detected in a plate reader based assay.

E.2 Co-injection of a kalooping construct.

In the second method to improve transient expression, a “kalooping” construct was co-injected with the $5 \times$ UAS pTOL2 TagYFP:NTR construct. This construct contained an autoregulatory Gal4-mediated expression loop (Distel *et al.*, 2009). This continually self-maintained a period of transient embryonic Gal4 expression, allowing for prolonged expression of the nitroreductase. The “kalooping” construct ($50 \text{ ng.}\mu\text{L}^{-1}$) was co-injected with the $5 \times$ UAS pTOL2 construct ($20 \text{ ng.}\mu\text{L}^{-1}$) into eggs from a gmc617 driver line.

Figure E.1: Fluorescence microscopy images of a zebrafish larvae 1 dpf co-injected with the *kalooping* construct and the pTOL2 TagYFP:NTR construct.

Bright TagYFP fluorescence (evidence of nitroreductase expression) was seen in the embryos co-injected with the *kalooping* construct for the first two days following injection (Figure E.1). A lot of undesirable yolk expression was evident, however clear mosaic expression was evident in the tail of the zebrafish larvae. This bright fluorescence faded quickly and was almost completely gone by 4-5 dpf. It was also noted that several fish that exhibited bright TagYFP fluorescence died, suggesting this *kalooping* construct or the high levels of *tagYFP:NTR* expression may be toxic to the larvae. This technique was deemed unsuitable for use in transient cell ablation assays as the transient expression was not prolonged enough.

E.3 Injection of a “cis” Gal4/UAS construct

Figure E.2: Schematic of the *Gal4/UAS cis* expression system. A single driver/reporter line contains both the transcription activator (*Gal4*) and the *UAS* element upstream of the *tagYFP:NTR*.

The final attempt to increase transient nitroreductase expression in zebrafish was the use of a “cis” Gal4/UAS construct. This is a self-reporting construct where the Gal4 transcriptional activator is located upstream of UAS element in the same construct (Figure E.2). Two Gal4/UAS cis constructs containing *NfsB_Ec* were trialled. These constructs had drivers allowing for broad expression in the muscle and neuronal cells and contained either

5 × UAS or 14 × UAS binding sites upstream of TagYFP:NfsB_Ec. These two constructs were injected into wild-type *roy orbison* eggs at the single cell stage. These zebrafish were selected as they have reduced iridophore numbers and translucent skin, allowing for better fluorescent imaging. TagYFP fluorescence (indicative of *nfsB_Ec* expression) was monitored from 1-7 dpf.

Figure E.3: Images of zebrafish larvae injected with either a 5 × UAS or 14 × UAS cis pTOL2 tagYFP:nfsB_Ec construct. Images were taken of each zebrafish in both bright field (left column) and under fluorescence (right column). Images are shown from 4-7 dpf once the zebrafish larvae had been released from the chorion.

Zebrafish larvae injected with the 5 × UAS construct showed prolonged TagYFP fluorescence lasting until 7 dpf, however this expression was confined to only a few cells (Figure E.3). Zebrafish larvae injected with the 14 × UAS construct showed widespread bright TagYFP fluorescence which was also prolonged until 7 dpf (Figure E.3). A preliminary plate reader assay was conducted to determine whether this fluorescence was bright enough to be detected on a plate reader.

E.4 Small-scale transient cell ablation assay

Following injection with the 14 × UAS cis construct, three zebrafish embryos were identified exhibiting widespread TagYFP fluorescence. At 4 dpf, two of the larvae were challenged with 5 mM metronidazole for a period of 48 hours, while the remaining larva was given the equivalent amount of DMSO and treated as the control. This is the standard

concentration and exposure time used when ablating cells stably expressing *nfsB_Ec*. The cell ablation assay was carried out as previously described in Section 2.10.6, however TagYFP fluorescence measurements were taken pre- and post- treatment due to the variability in TagYFP fluorescence between the three larvae. Cell ablation was calculated by determining the percentage TagYFP fluorescence of post-treatment fluorescence to pre-treatment fluorescence for each larva.

Figure E.4: Small-scale transient metronidazole cell ablation assay of zebrafish larvae transiently expressing tagYFP:nfsB_Ec. Zebrafish larvae transiently expressing tagYFP:nfsB_Ec were challenged with 5 mM metronidazole (treatment group, n=2) or an equivalent amount of DMSO (control group, n=1). TagYFP fluorescence was measured on the plate reader pre-and post-treatment to determine the percentage decrease in TagYFP fluorescence after treatment with metronidazole.

Both zebrafish larvae challenged with 5 mM metronidazole showed a substantial decrease in TagYFP fluorescence indicating metronidazole mediated cell ablation by NfsB_Ec (Figure E.4). No change in TagYFP fluorescence was observed in the control zebrafish larva treated with DMSO (Figure E.4). Time constraints during our 10-week research trip meant this experiment was only repeated once. This work is ongoing in the Mumm lab, to assess whether this transient cell ablation assay would have utility to assess preliminary nitroreductase expression.