STUDIES ON RIBOSOMAL OXYGENASES

A thesis submitted to the board of the Faculty of Physical Sciences at the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy



Rok Sekirnik

St Catherine's College, Oxford

Hilary Term 2014

Abstract

The 2OG oxygenases comprise a superfamily of ferrous iron dependent dioxygenases with multiple biological roles, including in hypoxia sensing, transcriptional control, and splicing control. It was recently proposed that 2OG oxygenases catalyse the hydroxylation of ribosomal proteins in prokaryotes (ycfD) and in humans (NO66 and MINA53), raising the possibility that 2OG oxygenases also control translation. The work described in this thesis concerned investigations on the biochemical and functional aspects of prokaryotic and mammalian ribosomal protein hydroxylases (ROX) *in vitro* and in cells.

An efficient chromatographic system linked to mass spectrometric analysis (LC-MS) was developed for studying the masses of individual ribosomal proteins (>90% coverage of ribosomal proteome) to ±1 Da accuracy. It was demonstrated that ycfD catalyses the hydroxylation of R81 on L16 in *E. coli*, in a manner dependent on atmospheric oxygen levels. Y*cfD* deletion results in growth phenotype at low temperatures and in minimal medium, and in decreased global translation rates in minimal medium; *ycfD* deletion does not affect translational accuracy and ribosome assembly.

Furthermore, *ycfD*-deletion results in increased sensitivity to the antibiotics chloramphenicol and lincomycin. Consistent with a 2OG-oxygenase mediated mechanism of antibiotic resistance, chloramphenicol sensitivity of the *E. coli* wild-type strain could be increased by inhibiting the activity of ycfD by removing co-factors required for catalytic activity (Fe(II) and O_2), and, at least in part, by using a ycfD inhibitor, IOX1, which inhibits ycfD with IC_{50} of 38 μM *in vitro*. The therapeutic potential of a post-translational modification mediating antibiotic resistance provides an opportunity for medicinal targeting of ribosome-modifying enzymes, for example ycfD, which may be more 'druggable' than the ribosome itself. In co-treatment with an existing antibiotic, such as chloramphenicol, a small molecule inhibitor would achieve a potentiated antibiotic effect.

Structural aspects of ROX hydroxylation were pursued by characterising a thermophilic ROX-substrate complex; a ycfD homologue was identified in the thermophilic bacterium *Rhodothermus marinus* and shown to be a thermophilic 2OG oxygenase $ycfD_{RM}$, acting on R82 of ribosomal protein $L16_{RM}$. The activity of $ycfD_{RM}$ in cells was limited at high growth temperature and oxygen solubility was demonstrated as a likely limiting factor of $ycfD_{RM}$ activity, thus identifiying a potential 2OG oxygenase oxygen sensor in prokaryotes. A crystal structure of $ycfD_{RM}$ in complex with $L16_{RM}$ substrate fragment was determined to 3.0 Å resolution. Structural analyses suggested that $ycfD_{RM}$ contains 30% more hydrophobic interactions and 100% more salt-bridge interactions than $ycfD_{EC}$, suggesting that these interactions are important for thermal stabilisation of $ycfD_{RM}$. The structures reveal key interactions required for binding of ribosomal proteins. Substantial structural changes were observed in the presence of the substrate fragment, which implies induced-fit binding of the $L16_{RM}$ substrate. The work has informed further structural studies on the evolutionarily related human ROX, NO66 and MINA53, for which substrate structures have been obtained since the completion of the work.

The LC-MS analysis of ribosomal proteins was extended to mouse and human cells to demonstrate that the human ROX homologue of ycfD, MINA53, hydroxylates the 60S ribosomal protein rpL27a in cells. It was demonstrated that rpL27a hydroxylation is widespread and found in all mouse organs analysed, as well as in cancer cell lines and in clinical cancer tissues. A partial or complete reduction of rpL27a hydroxylation was observed in a number of clinically identified *MINA53* mutations from the COSMIC database of cancer mutations. Structural analysis suggested that mutations occur more frequently at structurally important regions of MINA53, including the βIV-βV insert in the core fold of MINA53. The identification of inhibiting clinical mutations suggests that rpL27a hydroxylation level could be used as a cancer mark, and in the future for selective inhibition by ribosomal antibiotics. The work presented in this thesis demonstrates that it is possible to selectively inhibit modified ribosomes; an inhibitor of unhydroxylated rpL27a could therefore, at least in principle, be active against the sub-set of tumours with inactivating mutation(s) of *MINA53*, but not normal tissue. Future work should therefore focus on identifying a selective inhibitor of unhydroxylated eukaryotic ribosomes which could be applied for treatment of cancers harbouring deactivating *MINA53* mutations. The same approach could be applied to other ribosome modifications (to rRNA, ribosomal proteins, and ribosome-associate factors) that are different in cancer compared to normal cells.

Contents

Abstract **II**

Contents **III**

Acknowledgements **VII**

Abbreviations **VIII**

1. Introduction: contents **1**

1.1. Oxygen, iron and their biological uses 3
1.1.1. Iron 3
1.1.2. Iron-containing oxygenases 4
1.2. The discovery of 2OG- and Fe(II)-dependent hydroxylations 5
1.2.1. Small molecule hydroxylases and a common fold 5
1.2.2. Double stranded beta helix fold – a common structural feature of 2OG oxygenases 6
1.2.3. Structural characteristics of the DSBH 7
1.2.4. The 2OG-binding residue defines the 2OG oxygenase subfamily 8
1.2.5. The catalytic cycle of 2OG oxygenases 9
1.3. The versatility of 2OG oxygenase biological substrates 11
1.3.1. 2OG oxygenases are hypoxia sensors in animals 13
1.3.2. Prolyl hydroxylase domain (PHD) enzymes 13
1.3.3. Factor inhibiting HIF (FIH) asparaginyl hydroxylase 14
1.4. Oxygen sensing in prokaryotes 16
1.4.1. Prokaryotic homologue of FIH – a prokaryotic oxygen sensor? 18
1.4.2. *E. coli* ycfD is an arginyl hydroxylase homologous to NO66 and MINA53 18
1.4.3. Ribosomal hydroxylases – a new biological function of 2OG oxygenases? 20
1.5. The ribosome and the process of translation 21
1.5.1. Initiation 24
1.5.2. Elongation 24
1.5.3. Decoding 25
1.5.4. Termination 27
1.6. Post-translational modifications on the ribosome 28
1.6.1. rRNA modifications 28
1.6.2. Post-translational modifications of ribosomal proteins 29
1.6.2.1. *N*-Terminal methionine removal 31
1.6.2.2. Methylation 31
1.6.2.3. Acetylation 34
1.6.2.4. β-Methylthiolation 36
1.6.2.5. Phosphorylation 36
1.6.2.6. Isopeptide bond formation 37
1.6.2.7. Ribosomal protein oxidation/hydroxylation 37
1.6.3. Modifications of ribosome-associated factors 39
1.6.3.1. EF-Tu modifications 39
1.6.3.2. Release factor (RF) glutamyl methylation 39
1.6.3.3. EF-P lysyl hydroxylation 40
1.7. Conclusions 42
1.8. Aims of the work described in this thesis 43
1.9. References 44

2.	**Mass spectrometric and functional studies on ycfD: contents**	**53**
2.1.	Prior mass spectrometric approaches to studying ribosomal proteins	54
2.1.1.	Aims	55
2.2.	Development LC-MS method for the study of ribosomal proteins	55
2.3.	*In vivo* hydroxylation of L16 is ycfD-dependent	66
2.4.	Effect of O2 and environmental conditions on ribosomal protein modifications	71
2.5.	YcfD as a putative oxygen sensor	77
2.6.	Phenotypic studies on the effect of *ycfD* deletion	78
2.6.1.	*YcfD* deletion affects growth rate at 20°C and in minimal media	79
2.6.2.	*YcfD* deletion affects growth in minimal media	80
2.6.3.	*YcfD* does not affect ribosomal assembly	81
2.6.4.	*YcfD* does not affect translational accuracy	83
2.6.5.	*YcfD* affects global translation rate	84
2.6.6.	*YcfD* deletion effects on growth using single carbon sources	86
2.7.	YcfD and TCA cycle metabolites	89
2.8.	Conclusions	92
2.9.	References	93
3.1.	**YcfD and ribosomal antibiotics: contents**	**97**
3.1.1.	Aims	98
3.2.	The effect of *ycfD* deletion on antibiotic sensitivity	99
3.2.1.	Sensitivity to antibiotics affecting the translation mechanism	100
3.2.2.	Effect of chloramphenicol analogues on growth of *ycfD*-deleted strain	103
3.3.	Increasing antibiotic sensitivity of BW25113 by inhibiting ycfD	104
3.3.1.	Iron chelator DFO increases susceptibility of BW25113 to chloramphenicol	105
3.3.2.	Co(II) treatment increases the susceptibility of BW25113 to chloramphenicol	105
3.3.3.	Chloramphenicol sensitivity of BW25113 increases in hypoxia	106
3.4.	Isogenic model system of ycfD activity	108
3.5.	YcfD inhibitor screening	110
3.5.1.	pH optimisation for ycfD activity	111
3.5.2.	Optimising kinetic parameters for ycfD activity screening	112
3.5.3.	In vitro MALDI-MS Inhibitor Screen	113
3.5.4.	Differential scanning fluorimetry (DSF) screen of ycfD stabilisers	115
3.5.5.	Determination of IC50 values for IOX1 analogues against ycfD	118
3.5.6.	IOX1 treatment increases the antibiotic potency of chloramphenicol	120
3.6.	Conclusions	123
3.7.	References	126
4.1.	**Biochemical characterisation of a thermophilic 2OG oxygenase ycfD$_{RM}$: contents**	**130**
4.1.1.	Aims	130
4.2.	Bioinformatic search for ycfD homologues	131
4.2.1.	Thermophilic bacterium *Rhodothermus marinus* contains a ycfD homologue	132
4.3.	Ribosomal analysis of *Rhodothermus marinus*	133
4.3.1.	Large scale growth	133
4.3.2.	Lysis of *Rhodothermus marinus* cells	134
4.3.3.	LC-MS analysis of *Rhodothermus marinus* ribosomal proteins	135
4.3.4.	L16$_{RM}$ is hydroxylated on R82	139
4.4.	Cloning of ycfD$_{RM}$ from *Rhodothermus marinus* gDNA	141
4.4.1.	Regional order neural network analysis of ycfD$_{RM}$	141
4.4.2.	Codon usage and GC-content of ycfD$_{RM}$	142
4.4.3.	Optimisation of PCR conditions for ycfD$_{RM}$ amplification from gDNA	143
4.5.	Expression and purification of ycfD$_{RM}$	144
4.6.	Activity studies of recombinant ycfD$_{RM}$	147
4.6.1.	YcfD$_{RM}$ hydroxylates L16$_{RM}$ fragment peptide *in vitro*	147

4.6.2.	MS/MS analysis confirms hydroxylation occurs on R82	148
4.6.3.	$L16_{RM}$ hydroxylation by $ycfD_{RM}$ is Fe(II) and 2OG-dependent	149
4.7.	Kinetic studies on recombinant $ycfD_{RM}$	150
4.8.	Investigations of the $L16_{RM}$ binding mode to $ycfD_{RM}$	153
4.8.1.	$YcfD_{RM}$ catalysis of cyclic L16 peptide	153
4.9.	YcfD substrate selectivity	156
4.9.1.	Identification of L16 residues conferring $ycfD_{RM}$ substrate selectivity	158
4.9.2.	Alanine scanning study confirms E-2 as the key residue for $ycfD_{RM}$ catalysis	159
4.9.3.	$L16_{EC}$ A79E variant improves hydroxylation kinetics with $ycfD_{EC}$	161
4.9.4.	Studying $ycfD_{RM}$ substrate selectivity in vivo	162
4.10.	Biochemical comparison of $ycfD_{EC}$ and $ycfD_{RM}$: biocatalytic potential	164
4.10.1.	Sequence similarity	165
4.10.2.	Enzymatic activity and stability	166
4.10.3.	Circular dichroism studies	169
4.10.4.	Temperature dependence of activity	172
4.10.5.	Cofactor thermostability studies	173
4.11.	Oxygen dependence	174
4.12.	Conclusions	175
4.13.	References	177
5.	**Structural studies on $ycfD_{RM}$: contents**	**181**
5.1.	Aims	182
5.2.	Stabilisation and inhibition studies on $ycfD_{RM}$	183
5.2.1.	The DSF thermal stabilisation assay identifies $ycfD_{RM}$ stabilisers	184
5.2.2.	TCA cycle intermediate inhibition of $ycfD_{RM}$	185
5.2.3.	TCA cycle intermediate inhibition of $ycfD_{EC}$	187
5.3.	Crystallization trials with $ycfD_{RM}$	188
5.3.1.	Anion Exchange Chromatography	188
5.3.2.	Crystallisation of $ycfD_{RM}$.Mn(II).NOG.$L16_{RM}$ complex	191
5.3.3.	Optimisation of the JCSG_A11 crystallisation condition	191
5.3.4.	Crystallisation of a $ycfD_{RM}$.Mn(II).FG2216 complex	193
5.3.5.	Data collection and processing	194
5.4.	Crystal structure of $ycfD_{RM}$	195
5.4.1.	Overall fold	195
5.4.2.	Topology analysis of $ycfD_{RM}$	197
5.4.3.	Salt bridge analysis	198
5.4.4.	Cation-π interactions	199
5.4.5.	Hydrophobic interactions	199
5.5.	The DSBH and substrate binding of $ycfD_{RM}$	199
5.5.1.	DSBH strand βI	200
5.5.2.	DSBH strands βII-βVIII	202
5.6.	Comparison of VioC and $ycfD_{RM}$ substrate complex structures	206
5.7.	NOG binding	208
5.7.1.	Substitution of the basic 2OG-binding residue on βVIII	210
5.8.	Induced fit upon L16 Binding	212
5.8.1.	Comparison of $ycfD_{RM}$ and $ycfD_{EC}$ substrate-binding residue conformations	213
5.9.	A hydrophobic pocket accommodates the substrate residue V81	215
5.10.	The dimerisation domain	215
5.10.1.	Testing the effect of salt-bridge interactions in the ycfD dimerisation domain	217
5.10.2.	Disruption of dimerisation leads to loss of catalytic activity	218
5.11.	The C-terminal (‘winged-helix’) domain	221
5.12.	Towards the structure of a ycfD.L16 protein-protein complex	223
5.13.	Conclusions	228
5.14.	References	229

6. Biochemical characterisation of MINA53 activity in cells: contents 231

6.1. MINA53 and cancer 232
6.1.1. Localisation 233
6.1.2. MINA53 levels in cancer indicate a potential tumour suppressor role 233
6.1.3. A loss-of-function cellular model of MINA53 234
6.1.4. MINA53 and autophagy 235
6.1.5. Biochemical studies on MINA53 reveal rpL27a as a potential substrate 236
6.1.6. Aims 239
6.2. Isolation of ribosomal proteins from mouse livers 239
6.3. Identification of rpL27a in the UPLC chromatogram 240
6.4. MINA53 knock-down reduces rpL27A hydroxylation 243
6.5. Oxygen-dependence of rpL27a hydroxylation 246
6.6. Analysis of rpL27a hydroxylation in tumour samples 249
6.7. Effect of somatic mutations of MINA53 on hydroxylation of rpL27a 250
6.8. *In vitro* testing of potential non-ribosomal MINA53 substrates 263
6.9. Conclusions 265
6.10. References 267

7. Conclusions: contents 269

7.1. Aims of the work described in this thesis 270
7.2. Mass spectrometric and functional studies on ycfD 270
7.3. YcfD mediates resistance to peptidyl transferase ribosomal antibiotics 271
7.4. Biochemical characterisation of a thermophilic 2OG oxygenase $ycfD_{RM}$ 271
7.5. Structural studies on $ycfD_{RM}$ 272
7.6. Biochemical characterisation of MINA53 activity in cells 273
7.7. Future work 274

8. Materials and methods: contents 275

8.1. Reagents for molecular biology 276
8.2. Growth media 277
8.3. Polyacrylamide-Gel Electrophoresis (PAGE) 278
8.4. Standard procedures 280
8.5. Molecular biology procedures 281
8.6. Cloning procedures 282
8.7. Microbiological techniques 289
8.8. Protein production 291
8.9. Protein purification 293
8.10. $YcfD_{RM}$ crystallisation 295
8.11. Biophysical techniques 296
8.11.1. Differential scanning fluorimetry 296
8.12. Ribosome Purification 298
8.13. Functional studies on $ycfD_{EC}$ 301
8.14. Liquid chromatography-mass spectrometric studies of ribosomal proteins 305
8.15. LC-MS/MS protein analysis 308
8.16. Peptide synthesis 311
8.17. Enzyme activity studies 312

9. Appendix 317

Acknowledgements

As I was starting my DPhil, wise people told me it was a marathon, not a sprint. They were right. It was a marathon over many years, cities, laboratories, topics and ideas, run together with a number of team-members, afflicted, as myself, by an unreasonable, almost self-destructive passion for science. I realised that, although in theory the most rational of all human pursuits, science can engage motives and forces that are far removed from the *ratio*.

At the end of the journey, I would like to thank my supervisor, Prof Christopher Schofield FRS, who has guided me all throughout in a kind and generous manner, always supporting my rebellious pursuits of the work I have not been asked to do, and forgiving me for not pursuing the work I have been asked to do. His inspired ideas, rightly called visionary by some, both frustrated and enlightened me, and inevitably turned out to be right.

I am also tremendously grateful for the co-supervision I received from Prof. Sir Peter Ratcliffe FRS. His focused, knowledgeable and innovative approach to science has been inspiring, and meetings with him encouraging.

Next, I would like to thank Angie, who has joined me at the final miles of the race, but provided an enormous amount of energy to push through and cross the line. If only every DPhil student would have the privilege of enjoying the selfless support and encouragement which transformed the most strenuous of academic activities into fun and enjoyment.

The friends whom I met along the way have been there for me at all times. I am deeply indebted to John, Wei Shen and Louise for everything they have done for me. Fellow lab members, all 50 of them, have made the experience enjoyable and practical. In particular, I would like to acknowledge the enjoyment that discussions and musings with Mike, Christoph, Armin, Hanna, Jurgen and Sarah have provided. To Adam and Wendy I am very grateful for always providing everything, and more, than I ever needed. To Manso, who enabled me to receive the first 'Highlighter of the Year award' for all the guidance that he provided in exercising tidiness and organisation. Working with Dr Mat Coleman has been utterly enjoyable, educational, and very fruitful.

I was fortunate to also have made friends outside of the lab. To Tony, Darja, Blaz and Tom, and many others who have kindly and patiently supported me in times when most would give up.

This thesis was begun, and nearly finished, at Quarr Abbey. To the monks of Quarr, I feel deeply indebted for providing a place of peace and love, which inspired a great deal of work described here.

Finally, my parents, especially my mother and father, have been extraordinarily patient and generous to have let their only son escape into the world for nearly a decade. Alas, the prodigal son returns, thankful for the patience, wisdom and love shared along the way.

Abbreviations

2D-LC	2-Dimensional liquid chromatographic
2-HG	2-Hydroxyglutarate
2OG	2-Oxoglutarate
2TY	2-Tryptone/yeast extract
a.u.	Absorbance units
A_{280}	Absorbance at 280 nm
Ac	Acetyl
APS	Ammonium persulfate
ARNT	Aryl hydrocarbon nuclear translocator
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BEH	Ethylene bridged hybrid
Bis-Tris	2,2-Bis-(hydroxyethyl)-iminotris(hydroxymethyl)methane
β-ME	β-Mercaptoethanol
BSA	Bovine serum albumin
CAS	Clavaminic acid synthase
CBP	CREB binding protein
CD	Circular dichroism
CHCA	α-cyano-4-hydroxycinnamic acid
$CODD_{556-574}$	C-terminal oxygen-dependent degradation domain
COSMIC	Catalogue of somatic mutations in cancer
CREB	c-AMP-response element binding protein
Da	Dalton (unit of mass)
DAOCS	Deacetoxycephalosporin C synthase
dd	Double distilled
DFO	Desferrioxamine
DFT	Density functional theory
DLR	Dual luciferase reporter
DMOG	Dimethyl *N*-oxalylglycine
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside 5'-phopshate
DSBH	Double-stranded beta helix
DSF	Differential scanning fluorimetry
DTT	Dithiothreol
E. coli	*Escherichia coli* bacteria
EDTA	Ethylenediaminetetraacetic acid
Ef-G	Elogation factor G
Ef-Tu	Elongation factor Tu
ESI	Electrospray ionization
ESI-FT-ICR-MS	Electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometer
EV	Empty vector
FAD	Flavin adenine dinucleotide
FIH	Factor inhibiting HIF

FPLC	Fast protein liquid chromatography
gDNA	Genomic DNA
GFP	Green fluorescent protein
GTP	Guanosine-5'-triphosphate
h	Hour(s)
H	Histone protein
HA	Human influenza hemagglutinin
HEK	Human embryonic kidney
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HIF	Hypoxia-inducible factor
HPLC	High-performance liquid chromatography
IC_{50}	Half maximal inhibitory concentration
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
JW1114_AxA	ycfD KO strain JW1114 with vector pRSet5D containing catalytically inactive YcfD variant GFP-ycfD H125A_H127A
JW1114_EV	YcfD KO strain JW1114 with vector pRSet5D containing GFP
JW1114_WT	YcfD KO strain JW1114 with vector pRSet5D containing WT GFP-ycfD
K_m^{app}	Apparent Michaelis-Menten constant
KO	Knock-out
LC-MS	Liquid chromatography–mass spectrometry
LC-MS/MS	Liquid chromatography–tandem mass spectrometry
MALDI-MS	Matrix-assisted laser desorption ionisation-mass spectrometry
MALDI-TOF-MS	Matrix-assisted laser desorption ionisation-time-of-flight-mass spectrometry
Me	Methyl
MEF	Mouse embryonic fibroblast
MIC	Minimum inhibitory concentration
min	Minute(s)
MINA53	Myc-induced nuclear antigen with a molecular mass of 53 kDa
MPD	2-Methyl-2,4-pentanediol
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MTS	Mitochondrial activity assay
MW	Molecular weight
NAD+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide reduced form
Ni-TED	Nickel- tris(carboxymethyl)ethylene diamine
NML	*N*-terminal methionine loss
NO66	Nucleolar protein 66
NOYD	N-oxalyl-D-tyrosine
NTM	*N*-terminal methionine
O/N	Overnight
OGFOD1	2-oxoglutarate and iron-dependent oxygenase domain containing 1
OH	Hydroxyl
P3H	Prolyl-3-hydroxylase

P4H	Prolyl-4-hydroxylase
PAGE	Polyacrylamide gel electrophoresis
PAS	PER-ARNT-SIM
PHD	Prolyl hydroxylase domain enzyme
PHD	Plant homeodomain
PHD2	Prolyl hydroxylase domain-containing protein 2
PHF8	PHD finger protein 8
Pro	Proline
PTC	Peptidyl transferase centre
PTM	Post translational modification
PTM	Post-translational modification
pVHL	Von Hippel-Lindau tumour suppressor protein
Ras	Rat sarcoma
RESP	Restrained electrostatic potential
RNA	Ribonucleic acid
ROS	Reactive oxygen species
ROX	Ribosomal oxygenase
RP	Reverse-phase
rRNA	Ribosomal RNA
s	Second
s.d.	Standard deviation
s.e.m.	Standard error of the mean
SCX	Strong cation exchange
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
shRNA	Short hairpin RNA
t	Time
T	Temperature
TCA	Tricarboxylic acid
TED	Tris(carboxymethyl)ethylene diamine
TEMED	Tetramethylethylenediamine
TFA	Trifluoroacetic acid
T_m	Denaturation ('melting') temperature
TOF	Time-of-flight
Tris	Tris(hydroxymethyl)aminomethane
tRNA	Transfer RNA
UPLC	Ultra-performance liquid chromatography
UV	Ultraviolet
UV-vis	UV-visible spectroscopy
WT	Wild-type

Standard abbreviations for nucleotides, proteinogenic amino acids, SI units/prefixes and the chemical elements are used throughout.

1. Introduction

Contents

1.1. Oxygen, iron and their biological uses 3
1.1.1. Iron 3
1.1.2. Iron-containing oxygenases 4
1.2. The discovery of 2OG- and Fe(II)-dependent hydroxylations 5
1.2.1. Small molecule hydroxylases and a common fold 5
1.2.2. Double stranded beta helix fold – a common structural feature of 2OG oxygenases . 6
1.2.3. Structural characteristics of the DSBH 7
1.2.4. The 2OG-binding residue defines the 2OG oxygenase subfamily 8
1.2.5. The catalytic cycle of 2OG oxygenases 9
1.3. The versatility of 2OG oxygenase biological substrates 11
1.3.1. 2OG oxygenases are hypoxia sensors in animals 13
1.3.2. Prolyl hydroxylase domain (PHD) enzymes 13
1.3.3. Factor inhibiting HIF (FIH) asparaginyl hydroxylase 14
1.4. Oxygen sensing in prokaryotes 16
1.4.1. Prokaryotic homologue of FIH – a prokaryotic oxygen sensor? 18
1.4.2. *E. coli ycfD* is an arginyl hydroxylase homologous to NO66 and MINA53 18
1.4.3. Ribosomal hydroxylases – a new biological function of 2OG oxygenases? 20
1.5. The ribosome and the process of translation 21
1.5.1. Initiation 24

1.5.2. Elongation 24

1.5.3. Decoding 25

1.5.4. Termination 27

1.6. Post-translational modifications on the ribosome 28

1.6.1. rRNA modifications 28

1.6.2. Post-translational modifications of ribosomal proteins 29

1.6.2.1. *N*-Terminal methionine removal 31

1.6.2.2. Methylation 31

1.6.2.3. Acetylation 34

1.6.2.4. β-Methylthiolation 36

1.6.2.5. Phosphorylation 36

1.6.2.6. Isopeptide bond formation 37

1.6.2.7. Ribosomal protein oxidation/hydroxylation 37

1.6.3. Modifications of ribosome-associated factors 39

1.6.3.1. EF-Tu modifications 39

1.6.3.2. Release factor (RF) glutamyl methylation 39

1.6.3.3. EF-P lysyl hydroxylation 40

1.8. Aims of the work described in this thesis 43

1.9. References 44

1.1. Oxygen, iron and their biological uses

It is likely a characteristic of all aerobic organisms that they contain oxygen-dependent modifications of their biochemical components, including small molecules, proteins, lipids, and nucleic acids. Oxygen can act as a signalling modification; its small size and hydrophilic character enable it to function as the smallest modification possible which can convert a hydrophobic methylene unit into a polar derivative, which is ionisable (pKa 13-14), and has the potential to act as both a hydrogen bond donor and acceptor (1). It introduces a stereoelectronic *gauche* effect, thereby influencing the ring conformation of biomolecules (2,3). Atmospheric oxygen exists in the triplet electronic state, whereas most biomolecules exist in the singlet state. Hydroxylation is thus a spin-forbidden process; this kinetic obstacle can be overcome by a paramagnetic transition-metal catalyst, which promotes an exchange of electrons between the oxygen molecule and the catalyst by exchange of electrons via a set of spin-conserved steps (4). In nature, transition-metal containing dioxygenases catalyse the incorporation of both oxygen atoms into the products; the second atom can be removed by a co-substrate in order for the catalyst to be regenerated. Monooxygenases incorporate one oxygen atom into the product, and the other oxygen atom into water (*ibid.*).

1.1.1. Iron

Iron is highly abundant in the Earth's crust (5%). Its bioinorganic chemistry is versatile in terms of its oxidation states, its redox potential, coordination number, spin states, ligand types and ligand equilibria (5). Iron can adopt multiple oxidation states because of its position in the middle of the first transition metal series in the periodic table, resulting in partially filled 3d orbitals. Known inorganic examples of iron oxidation states range from Fe(-II) (e.g. $Fe(CO)_4^{2-}$) to Fe(VI) (e.g. FeO_4^{2-}). Biological examples include Fe(II) (ferrous, d^6 configuration), Fe(III) (ferric, d^5), and Fe(IV) (ferryl, d^4) (5). The redox chemistry of iron depends on the pH of the solution and the nature of the ligand. The ligand charge, field effects, stereochemistry, inductive effects and hard/soft characteristics influence the iron redox potential (5). Ligand field stabilisation energy (LFSE) is maximised in low-spin d^6, leading to stabilisation of octahedral low-spin Fe(II) by strong field ligands. Conversely, high-

spin tetrahedral complexes have weaker LFSE and high-spin d^5 (Fe(III)) is relatively more stable due to maximisation of spin exchange energy (5).

1.1.2. Iron-containing oxygenases

The capacity of iron to cycle through a wide range of oxidation states, and the requirement of the catalysis of oxygen-dependent modifications to undergo a spin-forbidden process, represents a complementary combination of chemical reagents capable of carrying out catalysis of oxidative modifications with signalling and metabolic importance. Iron-dependent oxygenases can broadly be classified into two families, depending on whether iron is chelated by a heme cofactor which is bound to a protein (heme oxygenases) or by protein residues directly (non-heme dependent oxygenases). Metal-dependent oxidases or oxygenases comprise diiron-using enzymes (e.g. methane monooxygenase, ribonucleotide reductase), and mono-iron using oxygenases, including Fe(III) (e.g. lipooxygenase, intradiol cleaving catechol dioxygenase) and Fe(II)-dependent oxygenases (extradiol cleaving catechol oxygenases, tyrosine/phenylalanine hydroxylases, benzene/naphthalene dioxygenases and the 2-oxoglutarate (2OG)-dependent oxygenases), of which the latter are the largest known group and catalyse possibly the most diverse set of oxidative modifications of any enzymatic family, including desaturation, ring closure, epimerization and chlorination (Figure 1.1) (6-10).

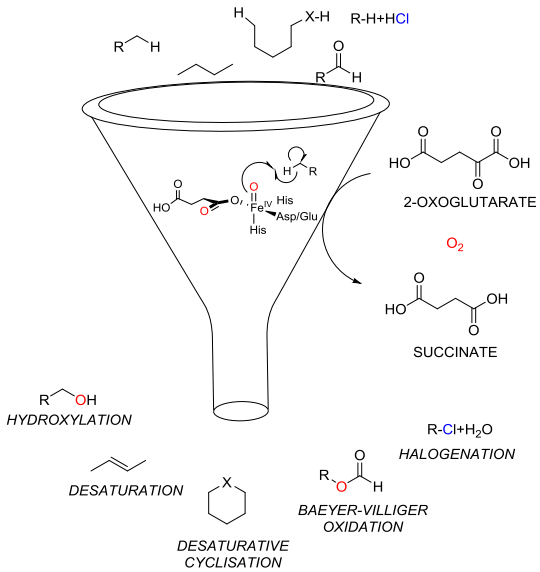


Figure 1.1: 2OG oxygenases catalyse the widest range of chemical transformations, including hydroxylation of unactivated C-H bonds, desaturations, ring closures, epimerizations and chlorinations. Adapted from Clifton *et al.* (7).

1.2. The discovery of 2OG- and Fe(II)-dependent hydroxylations

In 1753, the navy physician James Lind suggested that ingestion of citrus fruit may relieve symptoms of scurvy (11). By the 1930s, the lack of ascorbic acid (vitamin C) was established as the cause of scurvy (12) and it was later understood that the connection between the formation of the collagen triple helix, post-translational hydroxylation on the repeating Pro-Gly motif, and the requirement for ascorbate converge on an enzyme from the family of 2OG oxygenases (13). Hydroxylation of prolyl residues in pro-collagen is Fe(II)- and 2OG-dependent and catalysed by collagen prolyl hydroxylases (Figure 1.2) (14), which also require ascorbate for catalytic activity (15). This pioneering discovery provided a demonstration that 2OG- and Fe(II)-dependent enzymes are of significant physiological importance, and drove further academic interest in the field.

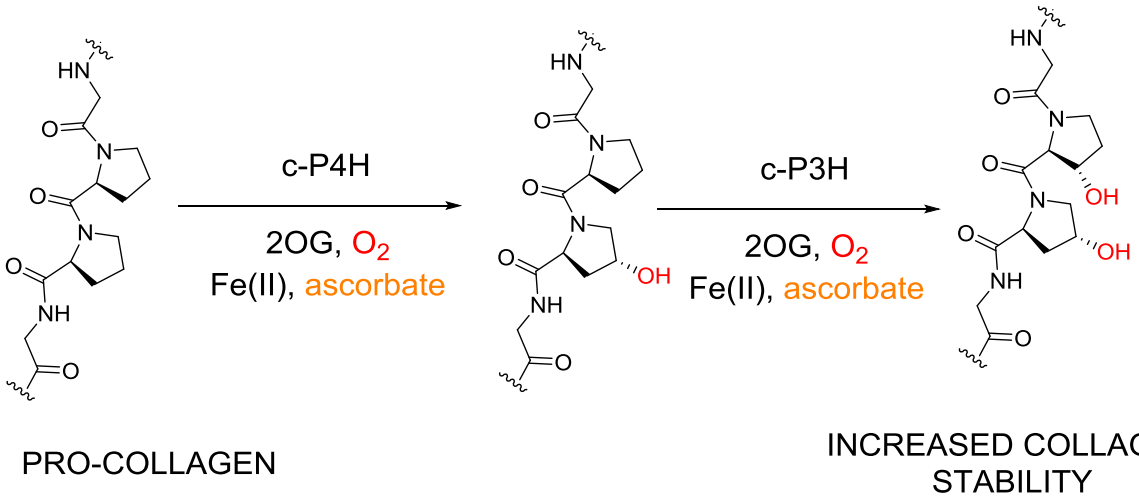


Figure 1.2: 2OG oxygenase-catalysed prolyl hydroxylation of pro-collagen at C4 and C3 positions is Fe(II), 2OG, O_2, and ascorbate dependent and regulates structural stability of the collagen triple helix. The elucidation of this process provided a biochemical explanation for the activity of ascorbic acid against scurvy.

1.2.1. Small molecule hydroxylases and a common fold

Although the initial understanding of the importance of 2OG oxygenases derived from attempts to understand the biology of collagen, the structural and mechanistic insights emerged from bacterial enzymes from the biosynthetic pathways of penicillins and cephalosporins. The crystal structure of isopenicillin N-synthase (IPNS) was the first structure of an enzyme related to the 2OG-dependent oxygenase family (16), followed by the structure of deacetoxycephalosporin-C synthase (DAOCS) (17). DAOCS is a 2OG-dependent oxygenase coupling the oxidation of 2OG to succinate to catalyse

the ring expansion of penicillin-N to deacetoxycephalosporin C (Figure 1.3a) (17). IPNS does not have a requirement for 2OG but instead catalyzes the four-electron oxidation of a tripeptide 1-δ-(α-aminoadipoyl)-L-cysteinyl-D-valine (ACV) to produce the isopenicillin N nucleus in the presence of Fe(II) and O_2 (Figure 1.3b) (18). Nonetheless, structural studies revealed that they share common structural features which apply to all known 2OG oxygenases.

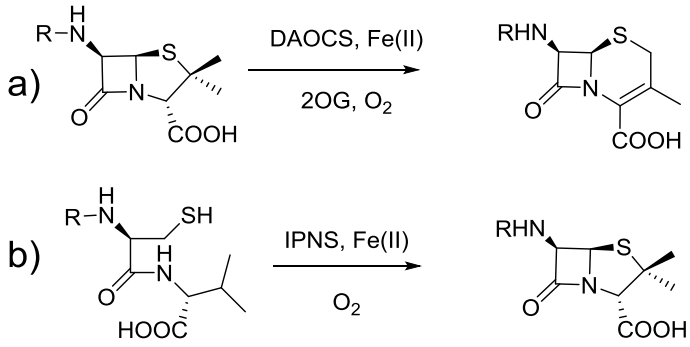


Figure 1.3: Reaction mechanisms catalysed by a) DAOCS and b) IPNS involved in the biosynthesis of β-lactam antibiotics. DAOCS is a 2OG-dependent oxygenase coupling the oxidation of 2OG to succinate to catalyse ring expansion of penicillin-N to deacetoxycephalosporin C; IPNS catalyses the four-electron oxidation of a tripeptide 1-δ-(α-aminoadipoyl)-L-cysteinyl-D-valine (ACV) to produce the isopenicillin N nucleus in the presence of Fe(II) and O_2, but does not require 2OG.

1.2.2. Double stranded beta helix fold – a common structural feature of 2OG oxygenases

The structures of IPNS and DAOCS revealed a common core comprised of eight antiparallel beta-strands folded into a double-stranded beta helix (DSBH, Figure 1.4), resembling the geometry of a jelly-roll, from which it derived its popular name (17,19). The structural similarity was striking given the low overall similarity of primary sequences of the two proteins, and suggested convergent evolution towards a common fold, which indicated its catalytic usefulness.

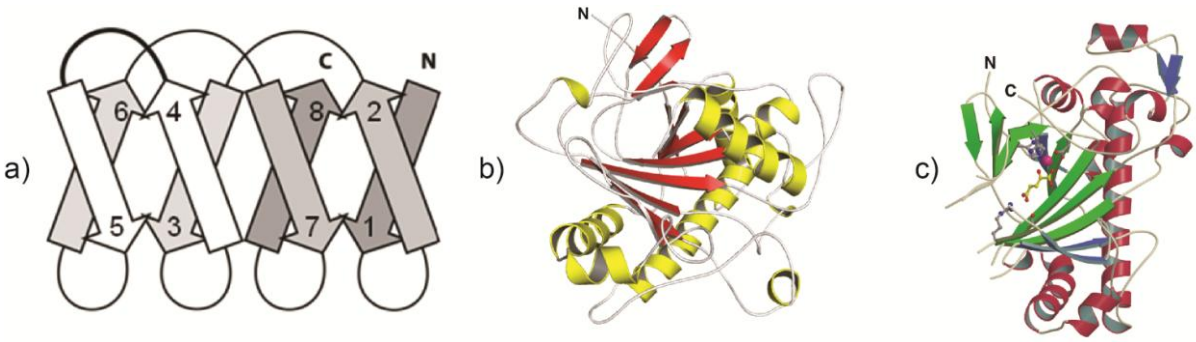


Figure 1.4: The 'double-stranded beta helix' (DSBH, figure a), commonly referred to as the 'jelly-roll' was identified in the crystal structures of b) isopenicillin N synthase (pdb: 1IPS) and c) DAOCS (pdb: 1RXG).

IPNS and DAOCS are only two representatives of the structurally and functionally diverse family of Cupin-domain containing proteins. According to an extensive bioinformatic analysis by Iyer *et al*. (20), cupins can be diversified into at least 8 sub-groups i) 'classical 2OG- and Fe(II) dependent dioxygenases (2OGFeDOs), ii) jumonji-related dioxygenases (JmjC), iii) cyclic nucleotide (cNMP)-binding domains, iv) Cu(II)-dependent monooxygenases, v) thiamine pyrophosphokinase substrate-binding domain, vi) ethanol regulon transcriptional regulator (AlcR)-N domain, vii) nucleophosmin/nucleophosmin superfamily and viii) viral capsid proteins. Proteins from sub-groups i-iii contain a jelly-roll configuration with extended internal cavity, which can accommodate the active site within the DSBH (see below). Sub-groups iv, vii and viii form a distinct clade that is characterised by flatter sheets and a limited interior cavity, which results in binding the peptide/nucleic substrates on the surface, rather than in the interior, of the DSBH (20).

1.2.3. Structural characteristics of the DSBH

2OGFeDOs and JmjC-containing 2OG oxygenases are the best-studied members of the cupin family. Their DSBH is not symmetric; it contains a 'major' and a 'minor' sheet, and is usually more 'open' on one end than on the other (21). The catalytic iron is bound by three residues located at conserved positions in the active site of 2OG oxygenases (βII and βVII), closer to the open end of the DSBH: two histidyl residues and an aspartyl residue are arranged in a HxD...H motif, referred to as the 'facial triad'. The acidic residue could also be glutamyl and the facial triad is therefore most often described as HxD/E...H; it is found in all structurally characterised non-heme 2OG oxygenases (Figure 1.5) (9,21).

The geometry around the catalytic iron is octahedral; three coordination sites are occupied by the facial triad residues, two by a bi-dentate 2OG ligand and one by water, which can be exchanged for O_2. Theoretically, a large number of possible isomers exists for 6-coordinate Fe(II) (22); the three protein-derived ligands can be arranged in *fac*- or *mer*- configurations. The requirement for the nucleophilic attack on the 2OG keto group excludes all *mer* isomers (22). Three *fac* isomers are chiral, giving rise to 6 possible isomers, of which only one has the keto group *trans* to the acidic substituent. This configuration facilitates the electron flow between Fe-O_2 π-bonding orbitals and the

antibonding π-orbitals of the 2OG keto group. Therefore, all known 2OG oxygenases for which structures are available show an identical configuration, classified as OC-6-35 *fac* (22); both stereoisomers (OC-6-35-A and OC-6-35-C) are observed crystallographically, with approximately equal occurrence, suggesting that both can support 2OG oxygenase activity (22).

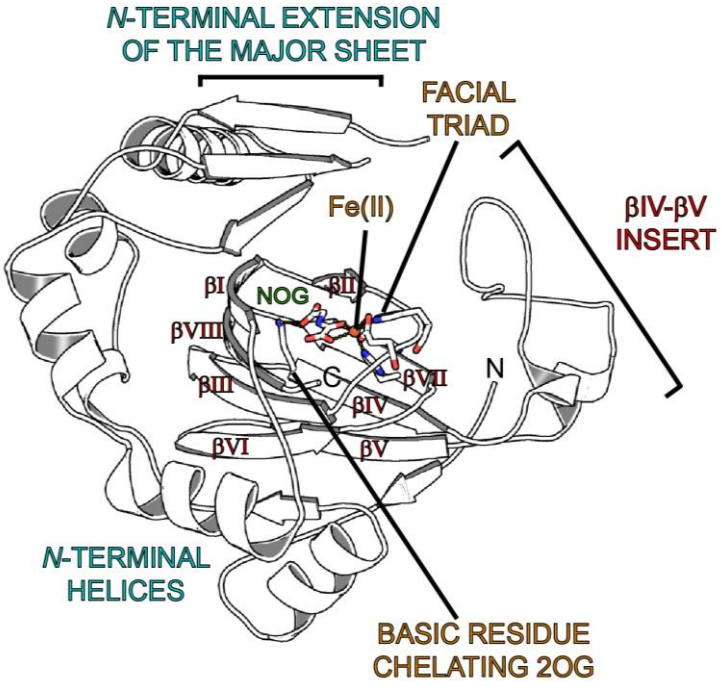


Figure 1.5: The double-stranded beta-helix (DSBH) motif is characteristic of 2OG oxygenase structures. The structure of the JmjC domain-containing Factor Inhibiting HIF (FIH) (PDB: 1H2K) demonstrates the key features of a DSBH: eight β-strands (I-VIII) comprise the DSBH which is flanked by *N*- and *C*-terminal (not shown) helices and strands (note that the *C*-terminal dimerisation helices have been omitted for clarity).The DSBH is comprised of a major β-sheet (strands I, III, VI and VIII) and a minor β-sheet (strands II, IV, V and VII). The active site metal (Fe(II)) is chelated by the 'facial triad' residues, comprised of two histidyl residues and an acidic residue (Asp/Glu) located on strands II and VII. The 2OG molecule chelates Fe(II) in a bidentate manner; the *C*5-carboxylate interacts with a basic residue at the back of the catalytic site; in 2OGFeDO 2OG oxygenases the basic residue is located on strand βVIII, in JmjC-domain containing 2OG oxygenases it is located on strand βIV (see text).

1.2.4. The 2OG-binding residue defines the 2OG oxygenase subfamily

The *C*5-carboxylate of 2OG interacts with a basic residue (lysine or arginine) from the back of the DSBH. The position of the basic residue within the DSBH is one of the defining characteristics of the 2OG oxygenase sub-family; JmjC enzymes contain a basic residue (lysine or arginine) on βIV forming a salt-bridge with the 2OG C5-carboxylate, whereas the 2OGFeDO 2OG oxygenases contain a conserved Arg on βVIII, performing an identical role; this suggests that 2OG chelating residues evolved independently. The histone demethylase JMJD3 and its homologues UTX and UTY differ from these canonical arrangements in having a basic residue at βI (23).

Other structural differences between the 2OGFeDO and the JmjC oxygenases include the *N*-terminal strand-helix extension present in the JmjC enzymes, an additional strand-helix unit inserted into the hairpin of N-1/βI, as well as the presence of the βIV-βV insert in the JmjC enzymes (Figure 1.6).

The DSBH of JmjC-domain containing 2OG oxygenases is further characterised by the conserved strand located antiparallel to βI and two *N*-terminal strands surrounding the core DSBH (labelled *N*-1/-3), also known as the *N*-terminal extension of the major sheet. They form the 'mouth' of substrate-binding cavity; together with the residues from the first and second β-strands of the DSBH, they are involved in prime substrate binding (21).

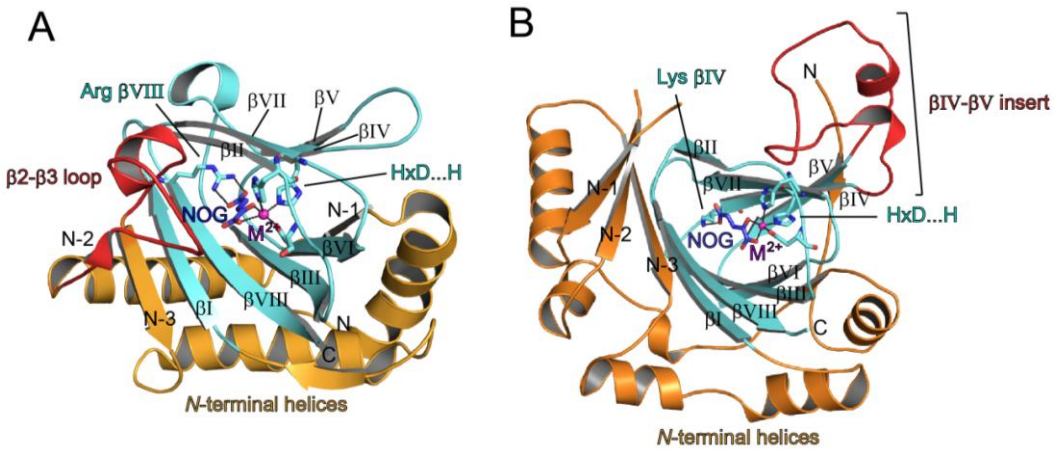


Figure 1.6: Structural comparison of the DSBH from two paradigmatic substructures of 2OG oxygenases: A) PHD2 (2OGFeDO family, PDB: 3HQR) and FIH (JmjC-domain containing 2OG oxygenases, PDB: 1H2K) family. Both structures contain an active site metal chelated by a 2OG analogue NOG and the facial triad (HxD…H). In PHD2, the C5-carboxylate is bound by an arginyl residue located on β8; in FIH the C5-carboxylate is bound by a lysyl residue from β4. PHD2 contains a typical β2-β3 loop *N*-terminal to the DSBH, which is involved in substrate binding; this loop is missing in FIH, which instead uses an extended β4-β5 insert to perform a similar function. FIH contains three sheets *N*-terminal to the DSBH which provide a 'lip' to the active site and assist in substrate positioning. In PHD2 these sheets are replaced by the β2-β3 loop (β2-β3 finger).

1.2.5. The catalytic cycle of 2OG oxygenases

The main feature of the catalytic cycle of all 2-OG oxygenases is the capacity of iron to cycle between the oxidation states +2 and +4, the latter being a highly oxidising ferryl species. Spectroscopic and crystallographic evidence suggests that iron exists in a +2 oxidation state at the start of each catalytic cycle and is coordinated by the facial triad and three water molecules (24). Upon binding of the co-substrate 2OG, the site becomes 5–coordinate and its affinity for O_2 increases. Fe^{2+} thus only binds oxygen in the presence of the required cofactor, which promotes strong coupling between the

reduction of O_2 and the oxidation of substrate and then decarboxylation of 2OG (25). Binding of O_2 is proposed to result in the formation of a Fe^{3+}-O_2^- superoxo species, which leads to oxidation of 2OG (activated by the presence of a metal centre) and its decomposition into succinate and CO_2, concomitant with a heterolytic O-O bond cleavage. The highly oxidising Fe^{4+} (ferryl) species has the capacity to insert an oxygen atom into an unactivated C-H bond of the substrate (Figure 1.7) (26). If the substrate is a methyl group located on a positively-charged nitrogen atom, as is the case in methylated lysyl residues, then the oxidation results in an unstable α-amino alcohol, which collapses, releasing a demethylated lysine and formaldehyde, providing a variation of the 2OG oxygenase mechanism to include demethylation (27).

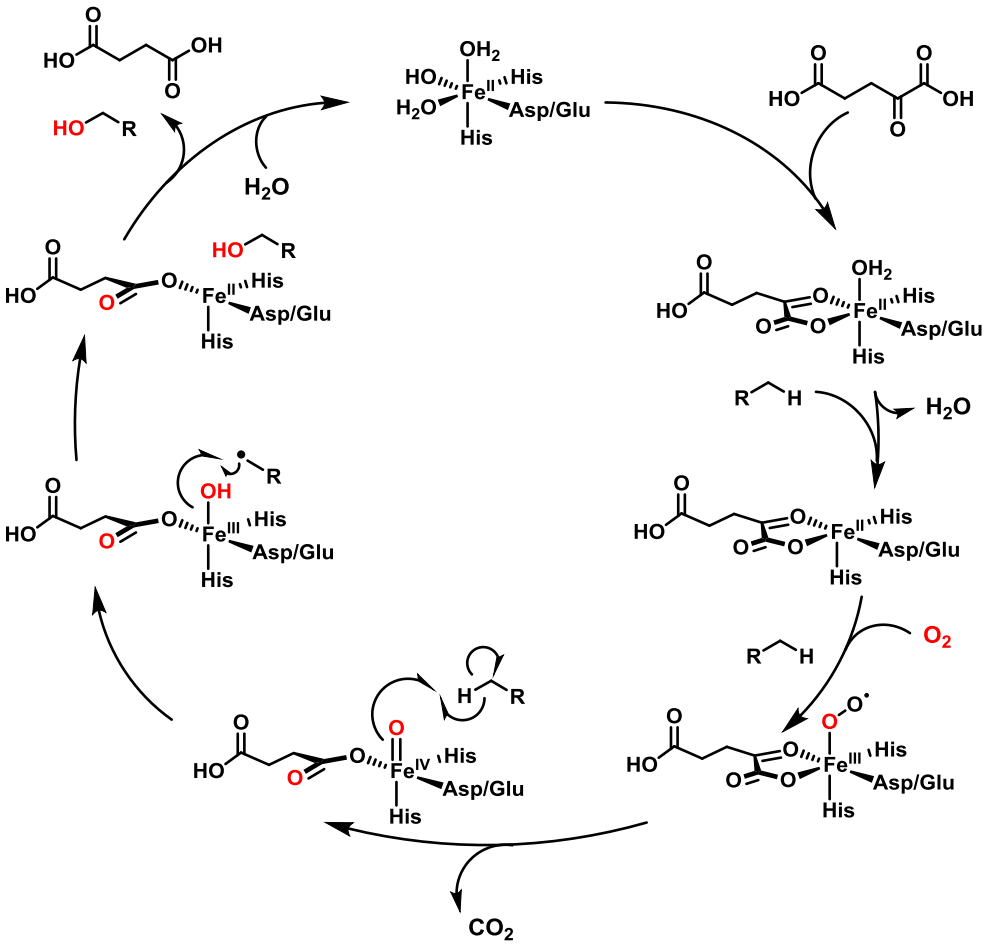


Figure 1.7: The proposed catalytic cycle of 2OG- and Fe(II)-dependent oxygenases. The facial triad (HxD/E…H) octahedrally coordinates the catalytic Fe(II), which initially binds three water molecules. Upon binding of 2OG and substrate the water molecules are displaced; the last vacant coordination site can be occupied by molecular oxygen upon displacement of H_2O. 2OG is decarboxylated with the release of CO_2, forming a highly reactive Fe(IV)=O ferryl intermediate which leads to hydroxylation of the primer substrate. The catalytic cycle ends with the release of hydroxylated substrate and succinate, which are replaced by water molecules.

1.3. The versatility of 2OG oxygenase biological substrates

Since the discovery of hydroxyproline and hydroxylysine residues in collagen fibrils, the scope of protein hydroxylation and *N*-methyl demethylation *via* hydroxylation has expanded considerably. For a detailed account, the reader is invited to consult any of the excellent reviews on the topic (1,7,28). In brief, the known biological substrates of 2OG oxygenases include proteins, nucleic acids and fatty acids involved in transcription (PHDs, FIH, KDMs), splicing (JMJD6), and control of fatty acid metabolism (PHYH, BBOX1) (28).

It is an interesting feature of the 2OG oxygenase family that classification into subfamilies based on sequence similarity, substrate specificity, biological roles, or even structural similarity, is challenging. As discussed above, the majority of 2OG oxygenases belong either to i) 2OGFeDO or ii) JmjC family. As demonstrated in Figure 1.8, the 2OGFeDO subfamily comprises DNA demethylases/tRNA hydroxylases from the ALKB/ABH subfamily; pro-collagen prolyl (P4H and Leprecan (P3H) subfamilies) and lysyl (PLOD subfamily) hydroxylases, HIF-α prolyl hydroxylases (EGLN subfamily), small molecule hydroxylase (gamma-butyrobetaine hydroxylase BBOX1), the rpS23 prolyl hydroxylase OGFOD1, endothelial growth factor asparagine hydroxylase (ASPH), and phytanoyl-CoA dioxygenases (PHYD subfamily). The figure demonstrates that it is difficult to infer biological function, and the chemistry of the substrate, based on phylogenetic proximity.

A similar observation can be made of the JmjC subfamily of 2OG oxygenases. They comprise of a large subfamily of histone demethylases (KDMs/JMJDs), including the distinct subfamilies of JMJD2s, JARIDs, FBXLs, JMJD1A/B and JMJD3/UTX/UTY, which catalyse a variety of oxidative demethylations of di-/tri-methyl lysines on histone H3 and H4. However, the role of JmjC-containing 2OG oxygenases is not limited to lysine demethylation. A distinct subfamily of JmjC oxygenases comprises FIH-like oxygenases, with no observed demethylase activity (FIH, JMJD4) or proposed demethylase activity (JMJD5, JMJD6, NO66, MINA53). Within the JmjC-containing hydroxylase subfamily, substrate selectivity and biological functions vary: FIH is a HIF-α asparaginyl hydroxylase mediating hypoxia sensing, whereas related JMJD4 and JMJD6 are lysyl hydroxylases with functions

in translation termination and splicing, respectively. As will be discussed later, NO66 and MINA53 are histidyl hydroxylases whereas their closest bacterial homologue, ycfD, is an arginyl hydroxylase. Although the wide range of JmjC-domain containing oxygenase substrates is challenging to unify, it could be regarded as a unique advantage of an enzyme subfamily that could be exploited biotechnologically to develop catalysts for a wide range of stereoselective hydroxylations. Indeed, substrate selectivity studies on FIH demonstrated that it can catalyse serine, histidine and valine hydroxylation (29).

Since only limited knowledge about individual 2OG oxygenases with unknown function can be inferred from sequence/structure similarity with 2OG oxygenases with known function, the elucidation of biological function of uncharacterised 2OG oxygenases (identified bioinformatically), must be followed by co-immunoprecipitation, or other cell-based biochemical studies.

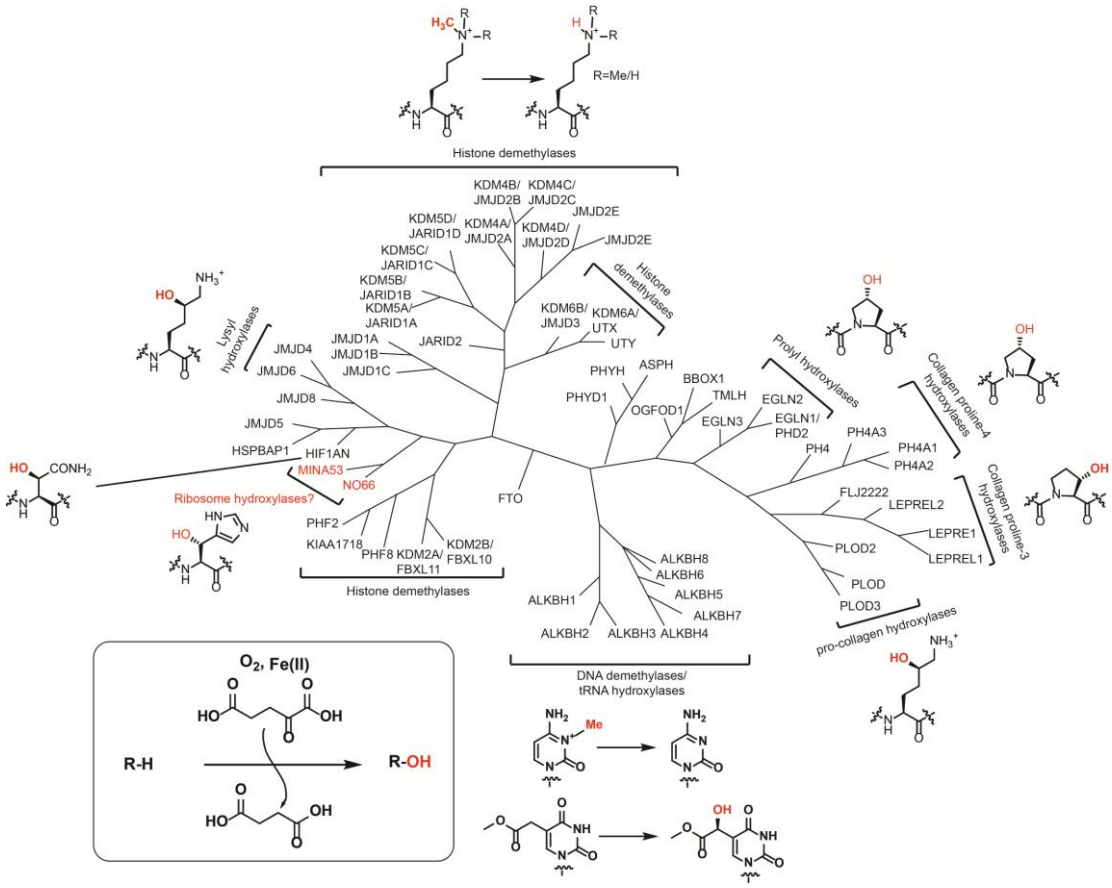


Figure 1.8: Phylogenetic tree of human 2OG oxygenases. The tree shows evolutionary relationships between 2OG oxygenases generated on the basis of sequence alignments (source: Structural Genomic Consortium (SGC), Oxford).

1.3.1. 2OG oxygenases are hypoxia sensors in animals

Arguably one of the most studied roles of 2OG oxygenases in humans and other animals is the regulation of the physiological response to hypoxia (30) *via* the hypoxia-inducible factor (HIF) system. The HIF pathway is an evolutionarily-conserved oxygen sensing mechanism, which acts to maintain a stable cellular supply of oxygen (31); the system has been found in the simplest animal *Trichoplax adherens*, but not in simpler eukaryotes or prokaryotes (32). HIF is a heterodimeric transcription factor composed of two basic helix-loop-helix protein subunits, HIF-α and HIF (33). Three human HIF-α isoforms (HIF-1α, HIF-2α, HIF-3α) and three HIF-β isoforms (Arnt1, Arnt2, Arnt3) are known. Both HIF-1α and HIF-2α can form transcriptionally-active heterodimers with HIF-1β (34); HIF-3α has no known role as a transcriptional activator. HIF-1/2α and HIF-1β are constitutively expressed (35), but HIF-1/2α are rapidly degraded under normoxic conditions, leading to its capacity as a sensor protein (Figure 1.9) (36,37).

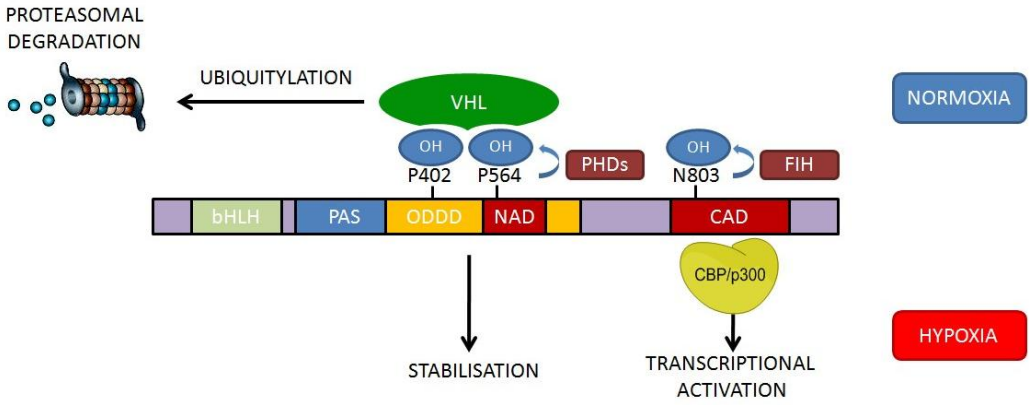


Figure 1.9: The structure of HIF-1α confers its function as a key regulator of hypoxic response (37). Two prolyl residues in oxygen-dependent degradation domain (ODDD) and N-terminal activation domain (NAD) are substrates for PHD enzymes; E3 ubiquitin ligase VHL recognises hydroxylated P402 and P564 and marks them for proteolytic degradation by the proteasome. In hypoxia, unhydroxylated prolyl residues are not recognised by VHL, thereby stabilising the HIF-α protein. Residue N803 in the C-terminal activation domain (CAD) is hydroxylated by FIH, enabling interaction with transcriptional co-activators CBP/p300. bHLH: basic helix-loop-helix, NAD /CAD: *N*-/*C*- terminal activation domain.

1.3.2. Prolyl hydroxylase domain (PHD) enzymes

The proteolytic stability of HIF-α and its transcriptional activity are regulated through an oxygen-dependent mechanism. Two prolyl residues (P402 and P564) within the oxygen-dependent degradation domains (ODDDs) of HIFα are substrates for hydroxylation by prolyl hydroxylase

domain (PHD) enzymes (30). Three PHD enzymes (PHD1-3) have been implicated in regulation of HIFα hydroxylation, although PHD2 is the most important, at least for acute oxygen responses (38). The apparent K_m for oxygen of PHD hydroxylases (230-250 μM) is higher than the physiological oxygen concentration (~210 μM in water (39)), so that oxygen concentration is limiting for the enzyme activity over the entire physiological range, allowing it to function as an oxygen sensor (40). Although the HIF-α substrate P564 prolyl ring in the active site of PHD2 is observed to adopt the C^4-*endo* conformation (41), the catalytic product of PHD2-catalysed *trans*-4-prolyl hydroxylation is biased towards towards the C^4-*exo* conformation, likely due to the stereoelectronic *gauche effect* exerted by the electron-withdrawing hydroxyl group (2); this conformation of the hydroxylated prolyl ring increases the affinity of the von Hippel-Lindau tumour suppressor protein (pVHL) ~1000-fold, acting as the recognition component of an E3 ubiquitin ligase, leading to ubiquitylation of HIF-1α and subsequent proteasomal degradation (42) (Figure 1.9).

In hypoxia, the rate of PHD-dependent HIF-α hydroxylation is decreased due to the lack of O_2, and HIF-1a is no longer recognised by pVHL, resulting in accumulation of the HIF-1α protein which can heterodimerise with HIF-1β. The HIF-1α/β transcription factor binds to the *hypoxia-response elements* (HRE, Figure 1.10), conserved DNA sequence motifs (G/ACGTG) in the regulatory regions of hypoxia-inducible genes (30).

1.3.3. Factor inhibiting HIF (FIH) asparaginyl hydroxylase

To initiate transcription, the HIF-1 activity is enhanced by the transcriptional co-activators p300 and CREB binding protein (CBP). Hydroxylation of an asparaginyl residue (N803) in the *C*-terminal activation domain (CAD) by factor inhibiting HIF (FIH), a JmjC-domain containing 2OG oxygenase, blocks transcriptional activation by preventing the HIF-1α–p300 interaction (43-45). The K_m of FIH for oxygen binding is 90 μM (46), lower than for PHD (see above), resulting in limited activity of FIH only in severe hypoxia (47). Thus while PHD-dependent hydroxylation regulates HIFα stability, FIH-dependent hydroxylation regulates its transcriptional activity (Figure 1.10). HIFα contains one

transcriptional activation domain, the *C*-terminal activation domain, CAD (Figure 1.9), which is directly regulated by FIH. Activation of the different regions of HIFα (oxygen-dependent degradation domain (ODDD) or CAD, respectively) can lead to a differential expression of hypoxia-inducible genes (47). Given the capacities of PHDs and FIH to operate under different oxygen tensions, this may contribute to 'fine-tuning' of the hypoxic response. For example, CA-9 and PHD3 are CAD-inducible genes, whose mRNA is expressed when FIH activity is inhibited; PGK1 expression, on the other hand, is not inhibited by FIH activity (48). Over a hundred genes are regulated by HIF-1α (49); the most important transcriptional targets in the hypoxic response include erythropoietin (EPO), vascular endothelial growth factors/growth factor A **(**VEGF-A), glucose transporter 1/3 (Glut1/3), phosphoglycerate kinase 1 (PGK1) and carbonic anhydrase (CA-IX) (50). HIF-1α is upregulated in numerous types of hypoxic tumours (51), and affects erythropoiesis, invasiveness, immortalisation, pH regulation, vascularisation, metabolic reprogramming and genetic instability (Figure 1.10) (52).

Unlike PHD2, which appears to only catalyse the hydroxylation of HIF-α, FIH has been shown to catalyse the hydroxylation of multiple substrates, including the ankyrin-repeat domain (ARD) family (53). The ARD proteins are primarily structural proteins; they are thought to unfold when binding to the FIH active site (54) and the FIH-catalysed hydroxylation stabilises the ankyrin fold through introduction of a hydrogen bond in a turn, and does not function in a signalling capacity (55); furthermore, it never proceeds to completion and increases the affinity of ARD proteins for FIH, suggesting that it potentially serves to 'sequester' FIH from the hypoxia response pathway (56).

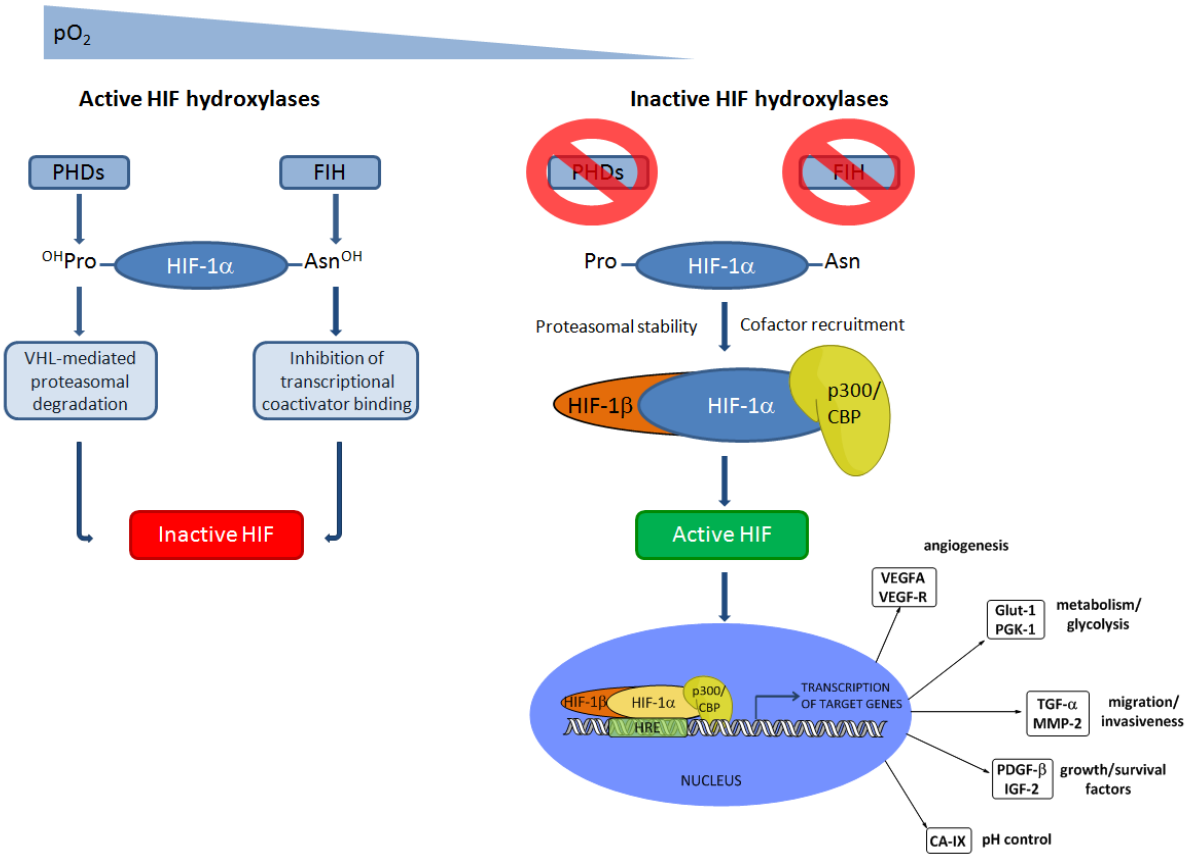


Figure 1.10: Regulation of hypoxia inducible factor (HIF) by oxygen levels. HIF prolyl hydroxylases (PHD) are most acutely sensitive to decreased oxygen levels, leading to inhibition of proteasomal degradation of HIF in hypoxia. At very low oxygen levels, factor-inhibiting HIFα (FIH) is inhibited, leading to recruitment of transcriptional co-activators p300 and CREB binding protein (CBP). Subsequent activation of HIF leads to induction of hypoxia-inducible genes by binding to *hypoxia response elements* (HRE) located in promoter regions of hypoxia inducible genes. The HIF system thereby regulates angiogenesis (via expression of vascular endothelial growth factors/growth factor receptors (VEGFA/R), metabolic reprogramming, (via expression of glucose transporter 1 (Glut-1) and phosphoglycerate kinase 1 (PGK1)), migration (via transforming growth factor a (TGF-a) and matrix metalloproteinase 2 (MMP-2)), growth (via platelet-derived growth factor-b (PDGF-b) and insulin growth factor 2 (IGF-2)) and control of cellular pH (via carbonic anhydrase-IX (CA-IX)) (57).

1.4. Oxygen sensing in prokaryotes

The HIF system is apparently limited to animals (32). In prokaryotes, no 2OG oxygenase has been implicated in oxygen sensing, which is thought to occur *via* two major pathways: (i) sensing proteins function as transcription factors which are stabilized in the absence of oxygen or (ii) proteins that have the capacity to harness changes in protein conformation, which occur due to a change in the heme-iron spin state upon binding of O_2 (Figure 1.11) (58). Prokaryotic oxygen sensors therefore include i) [4Fe-4S] cluster-containing proteins, e.g. the transcription factor product of the *fumarate and nitrate reductase* gene *FNR* (59,60), and ii) heme-containing sensor proteins, including Per-ARNT-Sim (PAS) domain proteins, cGMP-Adenylyl cyclase-FhlA (GAF) domain containing proteins

and globin domain-containing proteins (58). The sensing heme centres are found in proximity to kinase signaling domains, which become activated upon Fe(II)-spin change (e.g. the heme iron in FixL adopts a high-spin conformation in the absence of O_2, activating the kinase domain for autophosphorylation leading to downstream gene activation (61)).

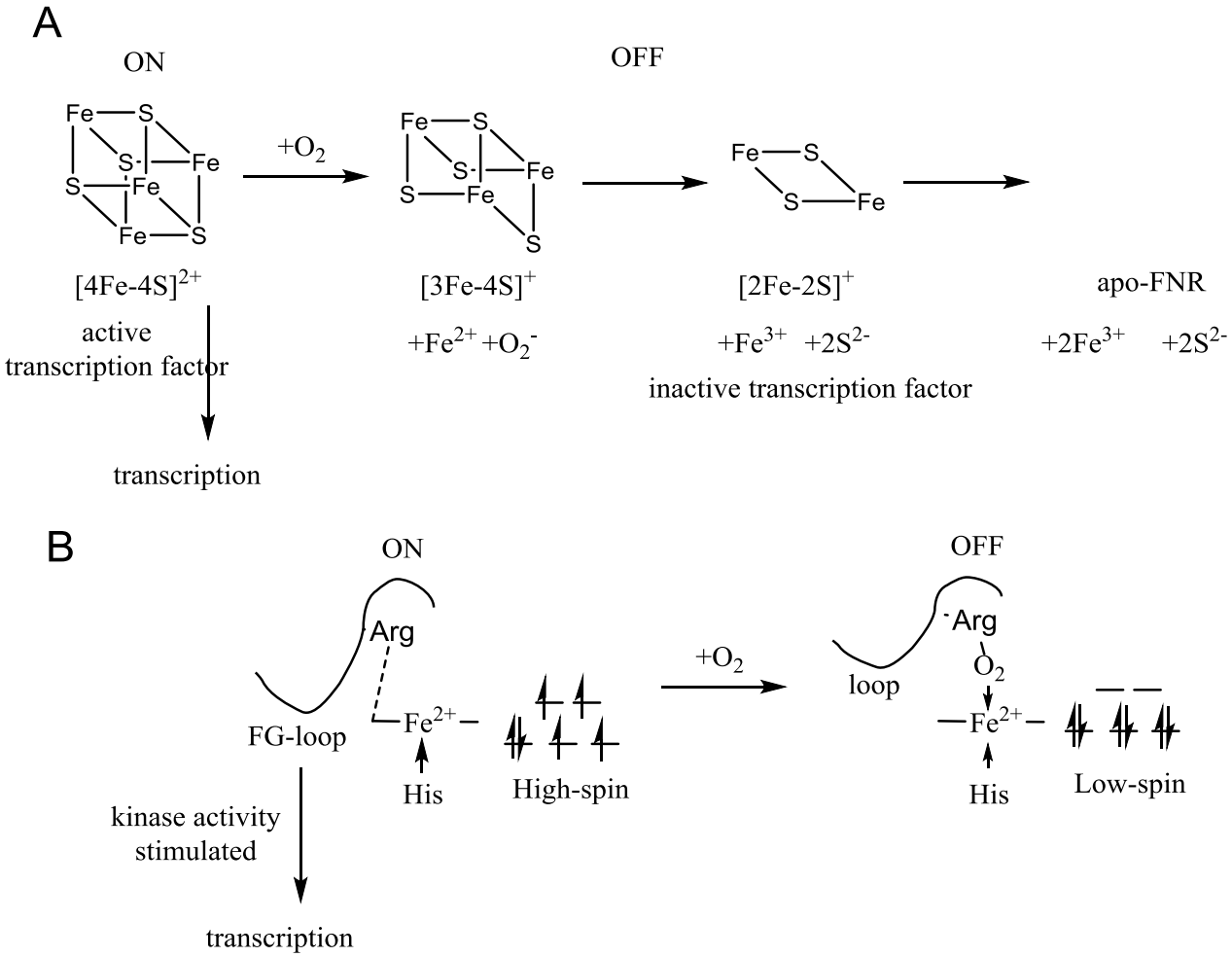


Figure 1.11: Oxygen sensors in prokaryotes include $[4Fe\text{-}4S]^{2+}$ cluster-containing proteins and heme-containing sensor proteins. A) Transcription factor FNR contains a $[4Fe\text{-}4S]^{2+}$ cluster which upon reaction with O_2 is oxidised from an active dimer ('ON') to an inactive monomer ('OFF'). The dimeric form leads to transcription of genes associated with anaerobic respiration. B) Heme-based oxygen sensing harnesses the conformational changes that ensue upon O_2-binding to Fe(II), which lead to a change in iron co-ordination state from high-spin to low-spin. For example, in FixL protein, the heme is located in a PAS domain ('sensory domain'), which is linked to a transmitter domain containing histidine kinase activity, which becomes activated upon the change of protein conformation. O_2-binding leads to a conformational change, preventing the activity of neighbouring histidine kinase domain, which auto-phosphorylates the transmitter domain in the absence of O_2, leading to transcription of target genes (58).

1.4.1. Prokaryotic homologue of FIH – a prokaryotic oxygen sensor?

Six putative 2OG oxygenases have been identified in *E. coli*, of which AlkB, a DNA repair enzyme (62,63) and taurine dioxygenase (TauD) (64) have been well characterised; the functions of csiD, ybiU and ybiX remain unknown. Interestingly, the putative *E. coli* 2OG oxygenase, ycfD, was found to share protein sequence homology with several human 2OG oxygenases, including the HIF asparaginyl hydroxylase FIH (65,66) lysyl hydroxylase of splicing-related protein U2AF56 (JMJD6 (67)), and 2OG oxygenases with proposed histone demethylase activity (Myc-induced nuclear antigen (MINA53) and the Nucleolar protein NO66 (68,69)).

The identification of the *ycfD* gene as sharing relatively high homology with a characterised human hypoxia sensor FIH suggests a potential role as an oxygen sensor in prokaryotes, thereby providing a previously uncharacterised oxygen-sensing mechanism in eukaryotes. Furthermore, its homology with human genes *NO66* and *MINA53*, the function of which has not been established, suggests that understanding the function of the *ycfD* gene may shed light on the function of its human homologues *NO66* and *MINA53*.

1.4.2. *E. coli ycfD* is an arginyl hydroxylase homologous to NO66 and MINA53

Work conducted in the Schofield laboratory prior to the work described in this thesis demonstrated that ycfD hydroxylates the peptide RLLPAVSEATI**R**RL on the arginyl-residue (R12) at the -3 position relative to the *C*-terminus of the substrate peptide; furthermore, it catalyses the hydroxylation of *C*-terminal arginyl residues at the C-3 position to produce an unprecedented 2*S*,3*R*-hydroxyarginine derivative (Figure 1.12), thereby extending the range of chemical modifications catalysed by the JmjC-containing hydroxylases to arginine (70).

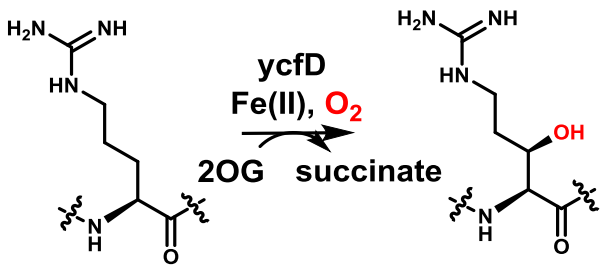
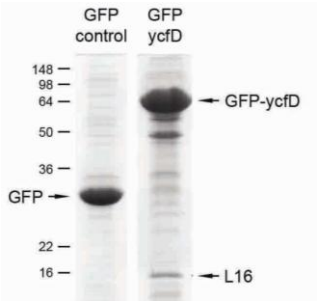


Figure 1.12: Arginyl hydroxylation catalysed by the *E .coli* ycfD.

Mechanistically, ycfD shares similarity to human 2OG dependent hydroxylases rather than 2OG oxygenases with small molecule substrates, as demonstrated by the high level (>90%) of ^{18}O incorporation into the product under an $^{18}O_2$ atmosphere rather than from water, and the fact that it catalyses arginyl, but not arginine hydroxylation. In *E. coli*, ycfD co-immunoprecipitates with a series of ribosomal proteins (Figure 1.13), including the 50S ribosomal proteins L16, L10, L9, L22 and L13 (70). L16 had been shown to be oxidatively modified on R81 (71); consistent with that observation, ycfD hydroxylates peptide fragments based on the L16 protein sequence containing R81 *in vitro*.



ycfD Interactor	Accession	Protein Score	# Peptides Detected
gatZ	B1IYY5	1534	52
EF-Tu	P0A6N2	872	34
50S ribosomal protein L16	P0ADY7	610	19
50S ribosomal protein L10	P0A7J3	400	15
50S ribosomal protein L9	P0A7R1	365	21
YaeH	Q0TLG8	316	29
50S ribosomal protein L22	Q0TCE6	111	15
50S ribosomal protein L13	P0AA10	97	11

Figure 1.13: Co-immunoprecipitation studies with GFP-labelled ycfD identified a number of ribosomal proteins as interaction partners of ycfD. SDS-PAGE gel demonstrated the presence of a number of interacting proteins. LC-MS-MS analysis of interacting proteins revealed gatZ, Ef-Tu, L16, L10, L9, YaeH, L22 and L13 as the interacting proteins with highest abundance. Reproduced, with permission, from (70) .

A crystal structure of ycfD (Schofield Lab, unpublished data) reveals a strikingly similar overall fold to NO66 and MINA53 (Schofield Lab, unpublished data and (72)): the JmjC catalytic domain is followed by a helical dimerisation domain and a *C*-terminal domain (Figure 1.14). Whilst a similar dimerisation domain is present in FIH (73,74), the *C*-terminal domain has no structurally homologous domains in other proposed/characterised JmjC protein hydroxylases (FIH (73,74), JMJD5 (75) and JMJD6 (67)) or KDMs (76). Their structural similarity may therefore suggest a conserved biological function. Indeed, work conducted in the Schofield and Ratcliffe laboratories prior to the work described in this thesis identified peptide fragments of the eukaryotic ribosomal proteins rpL8 and rpL27a (evolutionarily unrelated to L16) as substrates of NO66 and MINA53 for histidyl hydroxylation *in vitro (70).*

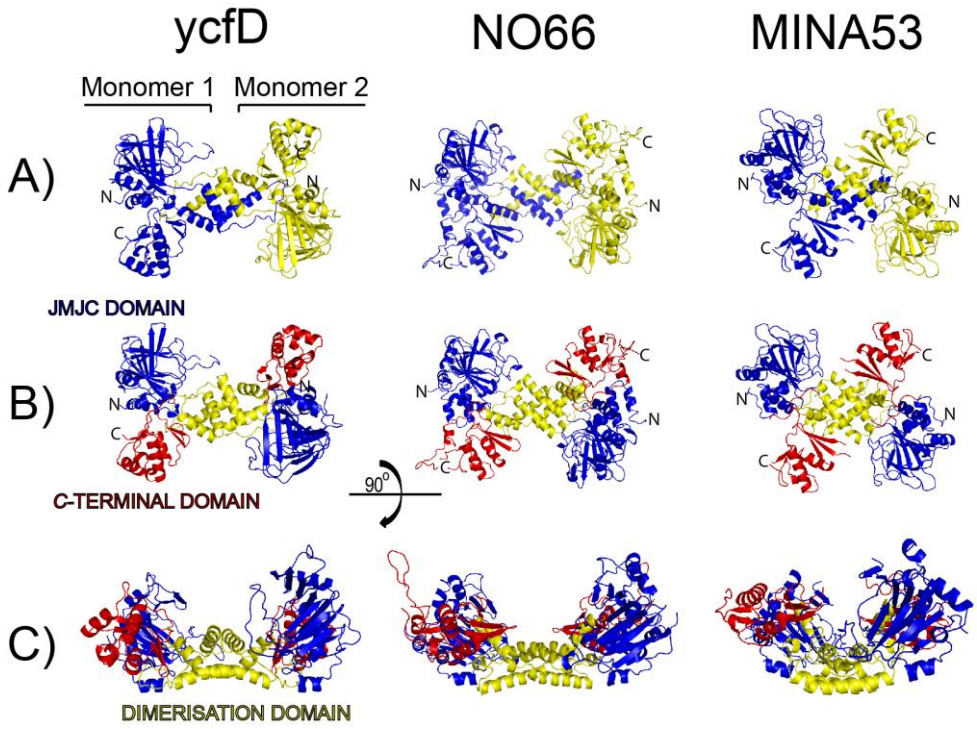


Figure 1.14: The overall folds of structurally similar JmjC-domain containing oxygenases ycfD (*E. coli*), NO66 and MINA53 (both *Homo sapiens*). A) A top-down view on the tertiary structure of ycfD, NO66 and MINA53. Two monomers are coloured yellow and blue. *N*- and *C*- termini are indicated; B) domain architecture of ycfD, NO66 and MINA53 is comprised of a JmjC (blue), dimerisation (yellow) and *C*-terminal (red) domains which have been conserved throughout evolution. C) A 'side-on' view of the dimerisation interface of the structures reveals subtle differences in the arrangement between monomeric units as well as location and length of loop extensions some of which protrude from the JmjC and *C*-terminal domains.

1.4.3. Ribosomal hydroxylases – a new biological function of 2OG oxygenases?

The identification of ribosomal proteins as substrates for prokaryotic and human 2OG oxygenases was significant. Although 2OG oxygenases are involved in cellular processes ranging from transcription (by transcription factor stabilisation *via* HIF-α hydroxylation (30) and by demethylation of histone proteins by histone demethylases (77)) and splicing (via modification of splicing-related protein U2AF65 by JMJD6 (67), the translation machinery has not been reported as a hydroxylation target for 2OG oxygenases.

The significance of the above finding is two-fold: i) 2OG-oxygenase catalysed hydroxylation of ribosomal protein is unprecedented; if true, it may represent a previously unknown oxygen-controlled cellular process, which could be exploited therapeutically; ii) the biological function for 2OG oxygenases ycfD, NO66 and MINA53 with previously unassigned biochemical functions would be elucidated.

The work described in this thesis aimed to address the following questions, which arise from the identification of a potential ribosomal hydroxylases:

i) Does ribosomal hydroxylation by ycfD (and human homologues) occur in cells?

ii) Is ribosomal hydroxylation affected by atmospheric oxygen levels?

iii) Does ribosomal hydroxylation affect cellular proliferation?

iv) Does ribosomal hydroxylation affect the process of translation (accuracy/rate)?

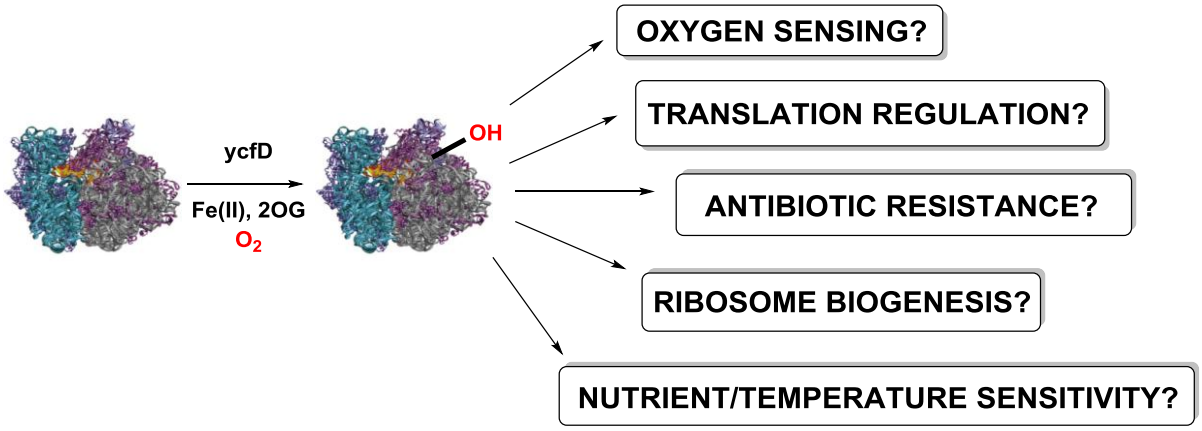
v) Does ribosomal hydroxylation mediate resistance/sensitivity to antibiotics?

vi) What is the mechanism of substrate recognition?

vii) Is the substrate recognition mechanism conserved through to human homologues?

viii) Does ribosomal protein hydroxylation occur in humans?

ix) Does ribosomal protein hydroxylation in humans vary with oxygen levels/disease state?



1.5. The ribosome and the process of translation

The ribosome is a macromolecular ribonucleoprotein complex which catalyses templated peptide bond formation during protein synthesis. It uses messenger RNA (mRNA) as a chemical template written in three nucleotide units (codons), which are recognised by complementary nucleotides (anticodons) on the stem loop of a transfer RNA (tRNA) molecule. The essential role of the ribosome is the catalysis of the peptidyl transfer reaction. Peptidyl tRNA carries an ester-activated amino acid on the 3'-end, which undergoes a nucleophilic attack by the α-amino group of the aminoacyl tRNA chain undergoing elongation. The ribosome increases the rate of the reaction 10^5-fold (78), with high

accuracy (reported values range between 1:450 error rate *in vitro* to $>10^6$ *in vivo*, the difference has not yet been settled) and rate (6.6 s^{-1} *in vitro*, 22 s^{-1} *in vivo* (78)).

In bacteria, ribosomes consist of a large (50S) subunit and a small (30S) subunit, which together form a functional 70S ribosomal particle of approximately 2.5 MDa (Figure 1.15). The 50S subunit comprises 23S rRNA (~2900 nucleotides), 5S rRNA (~120 nucleotides) and 35 proteins. The 30S subunit comprises 16S rRNA (~1500 nucleotides) and 25 proteins (79).

The first structures of the ribosome at atomic resolution were solved in 2000 by Steitz *et al.* who studied the structure of the archaeon *Haloarcula marismortui* (80). Shortly after Ramakrishnan *et al.* published the first structure of the 30S subunit from the thermophilic bacterium *T. thermophilus* (81). Subsequently, groups around the world used various approaches to crystallize and solve 70S ribosomal structures from the prokaryotes *E. coli* (82) and *T. thermophilus* (83), and recently of eukaryotic 40S and 60S from *T. thermophila* (84,85) and *S. cerevisiae* (86). Thermophilic (*T. thermophilus*), halophilic (*H. marismortui*) and radiation-tolerant (*R. deinococcus*) organisms played a crucial role in obtaining high-quality ribosome crystals, presumably due to the heat- and radiation resistant adaptation of their molecular assemblies.

The first structures revealed that 30S subunit consists of a 'head' and a 'body', which line a 'cleft' where mRNA binds and where the codon-anticodon interaction takes place. Three molecules of tRNA can bind to 30S subunit simultaneously, to tRNA binding sites called A-, P-, and E-site; each position has an associated catalytic function. The A-(aminoacyl) site binds the incoming aminoacyl tRNA, the P- (peptidyl) site binds the elongated peptidyl-tRNA which carries the nascent polypeptide chain. The E- (exit) site binds the deacylated tRNA, which is subsequently ejected from the ribosome and recycled by tRNA synthetases (Figure 1.15).

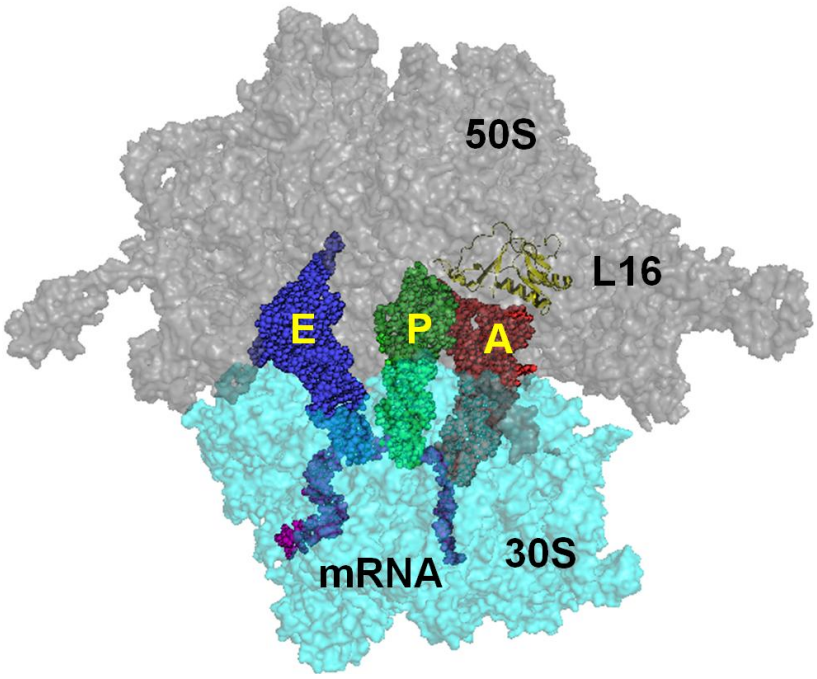


Figure 1.15: The crystal structure of the 70S ribosome in complex with mRNA and three tRNA molecules. The structure demonstrates the relative locations of the small (30S) and large (50S) subunit and the binding sites of mRNA and tRNA molecules. Three tRNA molecules can simultaneously bind to the ribosome in the A-, P-, and E-sites, respectively (indicated by yellow letters). The position of ribosomal protein L16, the proposed hydroxylation substrate of ycfD, is shown as a yellow cartoon. The figure was compiled using the PDB entries 2WDK, 2WDL.

While the 30S subunit serves to recognise cognate codon-anticodon interactions, the 50S subunit enables peptide bond formation. It harbours the peptidyl-transferase centre (PTC), where the aminoacylated 3'-ends of A- and P-site tRNAs come into close proximity, which results in a spontaneous nucleophilic attack by the free α-amine of A-site aminoacyl tRNA on the activated ribose ester of the P-site tRNA peptidyl chain. Nucleophilic attack leads to the elongation of the nascent polypeptide chain by a single amino acid, concomitant with the transfer of the chain onto the A-site tRNA molecule, which during the process (by a movement named 'ratcheting') migrates to the P-site, while the P-site molecule is relocated to the E-site.

The ribosome binds several extrinsic factors that facilitate translation. These are broadly classified as *initiation*, *elongation* and *termination* factors, referring to the part of translation process they facilitate; a translation cycle consists of three major steps: initiation, elongation and termination.

1.5.1. Initiation

In bacteria, the first step in initiation is positioning of the ribosome on the Shine-Dalgarno sequence upstream of the start codon on mRNA. Next, *N*-formylmethionine tRNA (fMet-tRNA^{fMet}) scans through the mRNA until it reaches the start codon (ATG) in the P-site, which requires initiation factors IF1-3 for correct positioning. The *N*-terminal residue on newly-synthesised protein is therefore methionine; however, the N-terminal methionine (NTM) is frequently post-translationally removed by methionine aminopeptidases (MAPs).

1.5.2. Elongation

After initiation, the ribosome is primed for elongation. The fMet-tRNA^{fMet} is located in the P-site; the A-site is empty. A ternary complex comprising of elongation factor Ef-Tu, GTP, and aminoacyl tRNA is presented to the ribosome for decoding. When cognate aminoacyl-tRNA is recognised in the decoding centre (see below), Ef-Tu undergoes GTP hydrolysis and dissociates from the complex, facilitating the migration of the 3'-end of aminoacyl tRNA into the PTC, known as 'accomodation'. In the PTC, the peptide bond formation takes place (see below). Upon successful peptide bond formation, translocation of tRNA and mRNA relative to the ribosome is facilitated by EF-G-catalyzed GTP hydrolysis (Figure 1.16) by a process called 'ratcheting'.

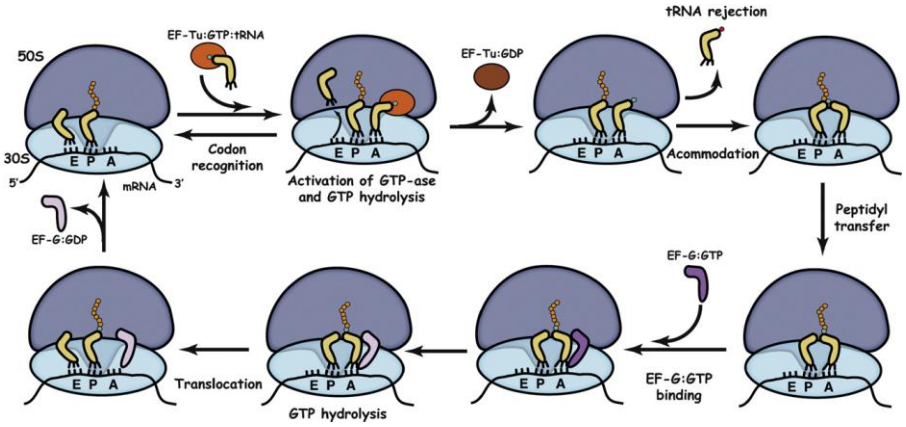


Figure 1.16: The elongation cycle. The elongation cycle of a translating ribosome comprises of a codon recognition step during which a ternary complex (Ef-Tu.GTP.tRNA) binds to the A-site. Cognate codon-anticodon interaction results in GTP hydrolysis catalyzed by GTPase Ef-Tu, followed by the release of Ef-Tu and the peptidyl transfer of the nascent peptidyl chain onto from the P-site tRNA onto the A-site tRNA amino acid. Translocation is enabled by binding of EF-G.GTP binary complex. Due to GTPase activity of EF-G, GTP hydrolysis provides the energy required for translocation step during which A- and P-site tRNAs translocate to P- and E-sites, respectively. Reproduced with permission from (87).

1.5.3. Decoding

Recognition of cognate tRNA is achieved in the decoding centre in the 30S subunit. Watson-Crick base pairing importantly contributes to correct recognition of codon-anticodon interactions. Furthermore, crystallographic studies revealed the function of three universally conserved rRNA bases, A1492, A1493 and G530 (Figure 1.17) (88). These were first implicated in the peptidyl transfer reaction in 1986, using studies of tRNA-dependent nucleotide protection from attack by chemical probes (89). Their importance was beautifully demonstrated in a crystallographical study using antibiotics which interfere with translational accuracy (88). Cognate codon-anticodon interaction leads to the formation of an mRNA-tRNA 'minihelix', which has distinct minor and major grooves. Conserved bases can recognise the minor groove of this helix, providing further discrimination of cognate from non-cognate mRNA-tRNA interactions. In the *apo* ribosome structure they are stacked in helix h44 of 23S rRNA (88). Upon binding of tRNA, they 'flip-out' from their original conformation to interact with the minor groove of mRNA-tRNA minihelix. This recognition does not occur at the third base of a codon. This position is called the 'wobble base', reflecting the lack of stringency of molecular recognition; it results in the degeneracy of the genetic code; multiple codons encoding a unique amino acid frequently differ in the third position, e.g. GGA, GGC, GGG and GGU coding for glycine.

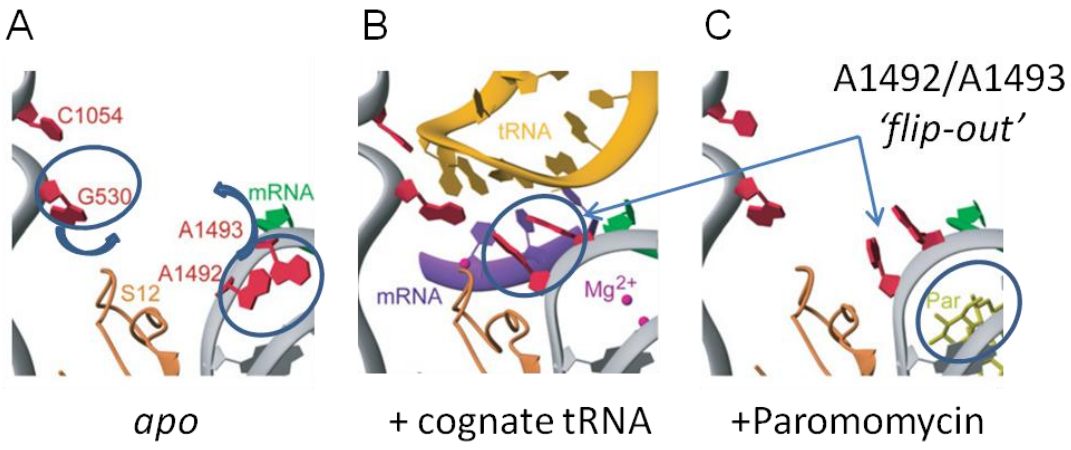


Figure 1.17: The decoding process in the 30S subunit involves conformational changes of three universally conserved nucleotides upon cognate tRNA binding, which also occur in the presence of aminoglycoside antibiotics. A) in the absence of tRNA and antibiotics, G530, A1492 and A1493 are in the *apo* conformation. B) In the presence of cognate tRNA, A1492 and A1493 'flip-out' and G530 rotates about the ribose-base bond, so as to interact with the minor groove on the mRNA-tRNA minihelix. C) Upon binding of aminoglycoside antibiotics, e.g. paromomycin, nucleotides A1492 and A1493 'flip-out' from their original conformation. Adapted from (90).

Theoretical calculations suggest that these interactions should lead to an accuracy of 10^3-10^4 (91). *In vivo*, this accuracy is not observed, suggesting that the binding energy of codon-anticodon interaction is dissipated in other processes. The binding energy derived from correct codon-anticodon pairing is utilised for inducing a conformational change in the ribosome (domain closure in 30S) and to distort the conformation of the tRNA anticodon stem (79). Ef-Tu-catalysed hydrolysis of GTP leads to tRNA 'relaxation' in the PTC, where peptide bond formation takes place. Entropic effects of proximity play a crucial role in the rate of peptidyl transfer, leading them to label the ribosomes as an 'entropy trap' (78). An extensive network of hydrogen bonds positions and orients the α-amine of aminoacyl-tRNA in optimal orientation for a successful reaction, as well as shields it from bulk solvent (78) (Figure 1.18). Ribosomal rRNA and ribosomal proteins play no direct role in the chemical reaction.

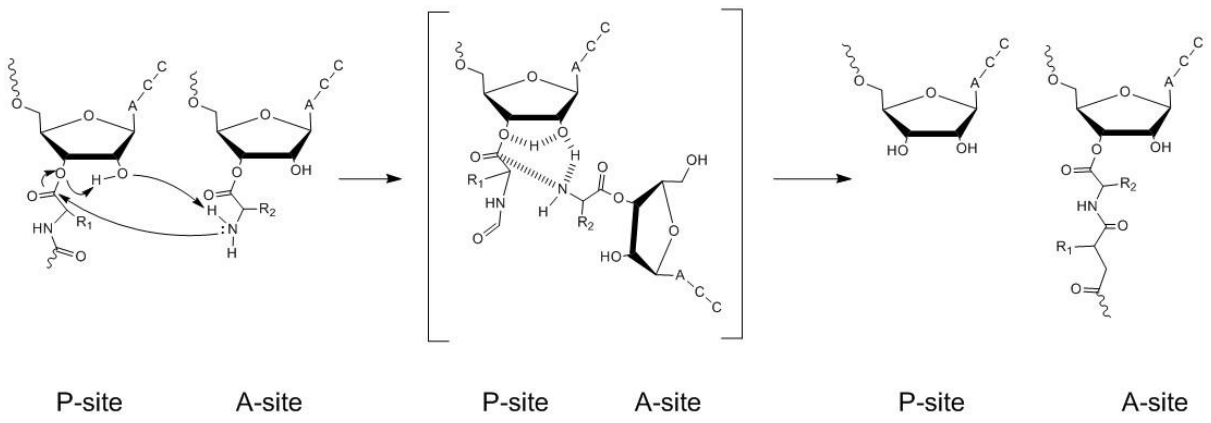


Figure 1.18: The peptidyl transferase step catalysed by the peptidyl transferase centre (PTC) located in the 50S subunit of the ribosome. An extensive network of hydrogen bonds positions and orients the α-amine of aminoacyl-tRNA in optimal orientation for a successful reaction, as well as shields it from bulk solvent (78). A water molecule is positioned in a location from which it can interact with the highly polarised transition state. Ribosomal rRNA and proteins play no direct role in the chemical reaction.

Upon a successful peptidyl transfer, the nascent polypeptide chain is transferred from the P-site tRNA to the A-site tRNA molecule. The tRNAs and mRNA therefore need to move relative to the ribosome, in order to allow a new aminoacyl-tRNA to bind to the A-site, and the deacylated tRNA to move from P-site to E-site, from which it will exit the ribosome. This movement is called 'translocation', and requires the formation of hybrid A/P and P/E states. The proposed ycfD substrate residue, R81 on L16, was proposed to be involved in the translocation process by assisting the positioning of tRNA in the P-site after 'stepping' from the A-site (92).

1.5.4. Termination

Translation termination takes place when the ribosome encounters a stop codon. Stop codons are recognized by the class I release factors (RFs) in prokaryotes. Three stop codons are known: ochre, amber, and opal (UAA, UAG and UGA, respectively). RFs have overlapping specificity for recognition of stop codons: RF1 recognizes UAG, RF2 recognizes UGA, and both recognize UAA. The tripeptide motifs (PXT and SPF) serve as recognition residues, chemically mimicking the anticodon (93). They are recognised by the same nucleotides that monitor codon-anticodon interaction (G530, A1492 and A1493) (94).

RFs promote ester hydrolysis of the bond between the nascent polypeptide and peptidyl tRNA, leading to the release of nascent polypeptide (Figure 1.19). The GGQ motif is required for catalytic activity of RF2 (95); its location in the PTC is universally conserved (96). Mutations of both glycines are detrimental for peptide release; mutations of glutamine to alanine, arginine and glutamate are tolerated (95). Mutations of GGQ glycines would lead to severely strained conformations that are not possible for amino acids other than glycine (97-99).

The side chain amide of Q240 is important for the catalytic activity of RF2 (98). A recent crystal structure reveals that Q240 forms a tight binding pocket for nucleophilic water (Figure 1.19) (100). In section it will be discussed that Q240 is subject to post-translational methylation, which is proposed to assist the catalytic function of RF2.

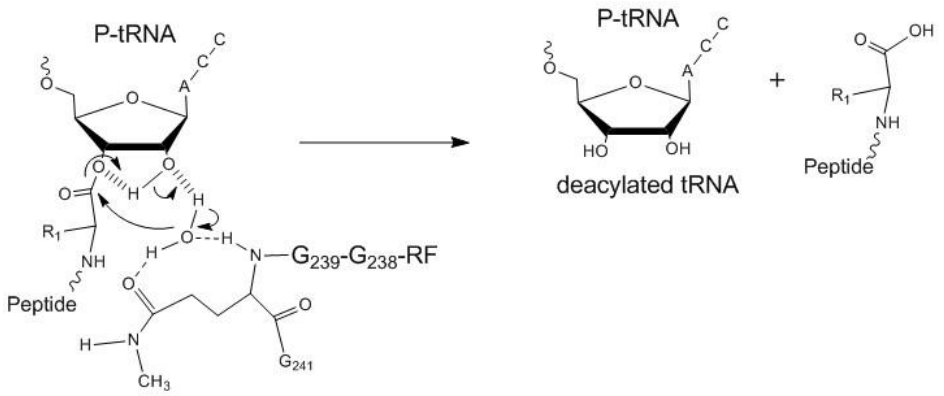


Figure 1.19: The mechanism of peptide release mediated by a release factor. When the ribosome enounters a stop codon, release factors bind to promote ester hydrolysis of the bond between the nascent polypeptide and peptidyl tRNA, which results in the release of nascent polypeptide. Q240 is a key residue in the structure of RF2; it forms a tight binding pocket for nucleophilic water, which attacks the ribose ester, leading to ester hydrolysis and peptide release.

Release factor RF3 from class II facilitates the dissociation of RF1/RF2 from the ribosome, which still contains mRNA and deacylated tRNA in the P-site. RF3 is a GTPase; hydrolysis of GTP likely induces the dissociation of RF1/RF2 from the ribosome. After hydrolysis, RF3 also dissociates from the 50S subunit, and mRNA and deacylated P-site tRNA are removed from the subunit by the ribosome recycling factor (RRF).

1.6. Post-translational modifications on the ribosome

1.6.1. rRNA modifications

The ribosome is subject to a large number of post-translational modifications both on rRNA and on ribosomal proteins. The majority of rRNA modifications are involved in ribosome biogenesis and serve as quality control check-points during ribosome maturation (101). They cluster around functionally important regions, including the decoding centre and the PTC (102,103). Some modifications of 23S rRNA are found to be essential for viability (104), whereas modifications of 16S rRNA are not (105). The rRNA can be modified in response to stress conditions, including heat shock (106); for example, the methyltransferase FtsJ is heat-inducible and catalyses the methylation of U2552 ribose (106). The deleted mutant is heat-sensitive, and leads to a decrease in programmed -1 and +1 frameshifting and a decrease in readthrough of UAA and UGA stop codons (107) suggesting that the modification is important under stress conditions (106). Modifications of rRNA can be induced upon more specific conditions, such as in the presence of macrolide antibiotics. Adenine A2058 is methylated by Erm-type methyltransferases in the presence of macrolides; the modification, located at the exit tunnel, leads to a decrease in translation rate under normal conditions and is unfavourable in the absence of macrolides (108).

Three types of rRNA modifications are known: i) pseudouridinylation (Ψ); ii) 2'OH methylation (Nm) and iii) base methylation (mN) (Figure 1.20). In *E. coli* rRNA there are ten Ψ, four Nm's and 19 mNs modifications (102). In eukaryotes the extent of modification is markedly increased (in humans, there are 91 Ψ, 105 Nms and 10 nMs). Modifications in *E. coli* are mediated by site-/region-specific

enzymes; in eukaryotes the Ψ and Nm modifications are catalysed by small nucleolar RNA-protein complexes (snoRPs) divided into two families, of which one catalyzes all pseudouridinylations and the other all 2'-O-methlyations. In *E. coli*, all of the pseudouridyl synthetases have been identified. Only RluD deletion leads to a growth defect, and aberrant ribosome biogenesis, suggesting that pseudouridines 1911, 1915 and 1917 are important for subunit association (109).

Sites of rRNA modifications are highly clustered around functional sites in the ribosome structure, both in 30S and 50S subunit (Figure 1.20). In both subunits the modifications are oriented towards the subunit interfaces. rRNA modifications do not occur in ribosomal protein-rich areas, suggesting that modifications do not affect rRNA-ribosomal protein interactions. There are a total of 11 identified modification sites in 30S and 24 in 50S. In 30S, most modifications are clustered around inter-subunit bridges, at the convergent point of the head, neck and upper body of 30S (110). One modification is found in the proximity of the A-site and three in the proximity of the P-site. In the 50S subunit, there are more modifications close to the A-site (five) than the P-site, but the majority of modifications occur in vicinity of the PTC (eleven) and the exit tunnel (eighteen). The modifications cluster within domain V and neighbouring domains II and IV. The modifications of the central loop region of domain V line the active site of the PTC, in proximity of 3' ends of A- and P-site tRNA binding sites (110) (Figure 1.20). Functions of the modifications are largely unknown. They may affect steric properties, hydrogen-bonding potential and base-stacking or increase rigidity of the rRNA. Methylation of 3'OH was proposed to affect solvation energy and favour 3'endo configuration, as well as to prevent hydrolysis of internucleotide bonds (111). One interpretation of this clustering is that the modifications have a structural role, providing rigidity to otherwise flexible regions (112).

1.6.2. Post-translational modifications of ribosomal proteins

Prokaryotic ribosomal protein modifications can be divided into six main classes: i) *N*-terminal methionine cleavage; ii) *N*-terminal and side-chain methylation; iii) acetylation; iv) β-methylthiolation; v) phosphorylation; and vi) hydroxylation.

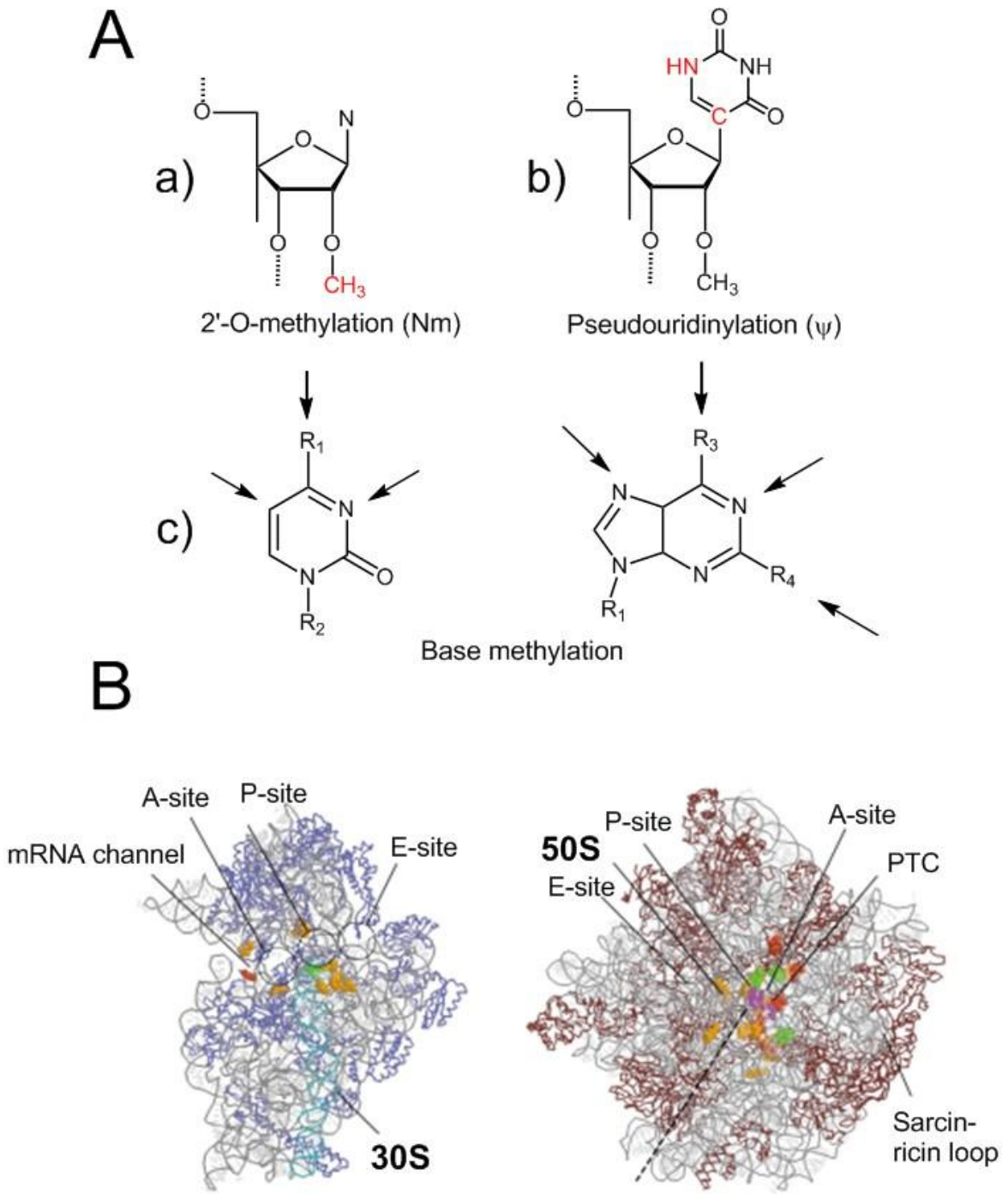


Figure 1.20: rRNA modifications cluster around the functional regions of the ribosome (adapted after (110). A) The *E. coli* ribosome is subject to ten Ψs four Nms and 19 mNs, which cluster around the decoding centre of the 30S and the peptidyl transferase centre (PTC) of the 50S. B) Grey ribbons represent rRNA, blue ribbons represent the 30S ribosomal proteins (*T. thermophilus*; PDB: 1FJF), red ribbons represent the 50S ribosomal proteins (*H. marismortui*; PDB: 1FFK/1FFZ). Orange surface indicates the positions of modified nucleotides with van der Waals radii. Adapted, with permission, from (110).

1.6.2.1. *N*-Terminal methionine removal

As discussed above, the translation of every protein starts with fMet-tRNA^{fMet}; consequently, formylated methionine is the starting amino acid of every newly-synthesised protein. The formyl group is cleaved by the action of a peptide deformylase followed by methionine aminopeptidases cleavage of the *N*-terminal methionine (NTM, Figure 1.21); the modification is observed in 37 of 55 *E. coli* ribosomal proteins (113). The modification depends on the residue at position 2. Only small hydrophobic residues allow NTM cleavage; larger residues are proposed to hinder binding to methionine peptidase active site. Alanine, leucine, proline and glycine at second position always lead to NTM cleavage; serine is ambiguous, leading to cleavage in one of 5 cases. All other amino acids lead to retention of methionine at the *N*-terminus (113,114). NTM cleavage was proposed to correlate with the gyration radius of residue at position 2, with cut-off radius between 1.3-1.4 Å (115). No specific function has been described for this modification.

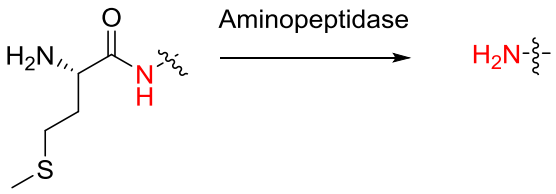


Figure 1.21: Aminopeptidases catalyse the removal of *N*-terminal methionine (NTM).

1.6.2.2. Methylation

Ribosomal protein methylations occur on the α-amino termini and on side-chains, including lysine and glutamine.

i) *N*-Terminal Methylation

Modification of α-amino termini is common. Methylation of ribosomal proteins is catalysed by *S*-adenosylmethionine-dependent methyltransferases, of which five sub-classes are known (114). Methylation can occur on the amine of an *N*-terminal residue or on the side-chain of a basic residue (arginine or lysine). The first reported *N*-terminal methylations were on ribosomal proteins L16 (71) and L33 (116). Although no functional significance of ribosomal protein methylation has been

reported, methylation occurs at functionally important sites in other proteins; for example, methylation of myosin light chain was proposed to occur at actin binding site (117).

N-Terminal methylations are classified according to the sequence motif required for methylation. One group consists of the 'QP methyltransferases', referring to the residues at positions 2-3 required for methylation (e.g. L16); another comprises the 'MAK methyltransferases', referring to residues 1-3 (e.g. IF3, L33 and S11) (114).

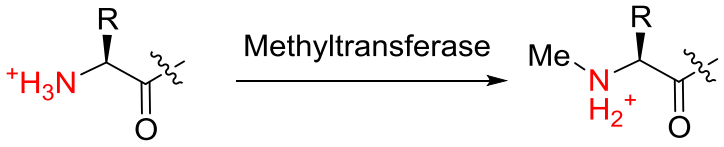


Figure 1.22: N-terminal methylation in prokaryotes is catalysed by 'MAK' and 'PQ' methyltrasnferases, referring to their sequence specificity required to achieve methylation.

ii) Side-chain methylation of L11

The *N*-terminus of a protein can be modified multiple times. For example, L11 is trimethylated on Ala1 (similar to other bacterial MAK methylations) and Lys3, similar to the eukaryotic 'PK' methylations. The enzyme PrmA was identified and characterised as the L11 methyltransferase (118-120), which also catalyses the trimethylation of Lys40, leading to a total of nine methylations of the protein (Figure 1.23); the phenomenon is widely spread in eubacteria (121).

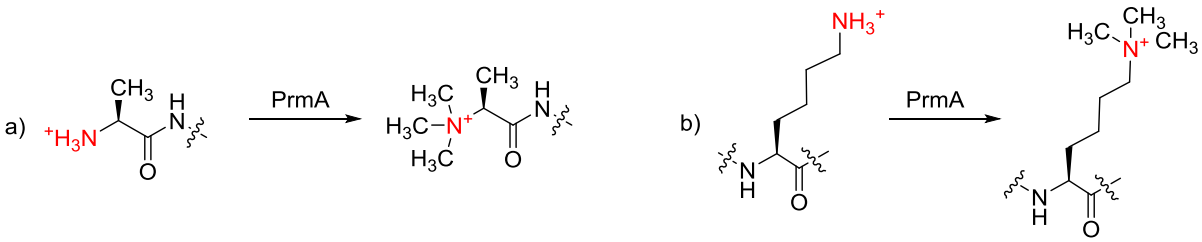


Figure 1.23: Methyltransferase PrmA catalyses trimethylation of L11 at N-terminal alanine and lysines K3 and K40.

Crystal structures of PrmA in complex with trimethylated *N*-terminal α-amino group in the active site and Lys39 bound to the active site have been determined (120). The PrmA-deleted *E. coli* strain does not show any phenotypic difference from WT, even under growth under stress conditions (118,122). Crystal structures of PrmA revealed unusual domain flexibility which allows for catalysis of multiple

substrates (120). Interestingly, the fold of the *N*-terminal domain of PrmA, which recognises and binds L11, is reminiscent of the V-domain of the elongation factor EF-G, which binds in close proximity of L11 in the ribosome. This suggests that one (or more) domain of a multi-domain ribosomal protein modifying enzymes may, in part, serve for substrate-recognition; the recognition may take place due to the tertiary structure resemblance with the ribosome/ribosome-associated factors, which bind the substrate of the modifying enzyme.

N-Terminal methylation is not always stoichiometric but can also compete with NTM cleavage; L33 protein could either be *N*-terminally methylated or the NTM may be cleaved; these processes compete for the substrate (116). The L33 *N*-terminal methylation status could therefore be indicative of metabolite levels required for one of the processes (116).

N-Terminal monomethylation does not affect the chemistry of the *N*-terminus significantly. It may decrease the pKa of the α-amino group by < 1 unit and create slight steric repulsion, thus lowering its reactivity. Trimethylation of α-amino group, on the other hand, affects the charge of the group by creating a full positive charge and abolishing the nucleophilicity of the α-amino group. In the structure of a ribosome, a strong positive charge may interact with the phosphate backbone of rRNA; in the active site of myosine, quaternary ion can be maintained in a hydrophobic environment under conditions which would otherwise lead to deprotonation of amine (123).

iii) Side-Chain Amide Methylation of Ribosomal Protein L3

L3 is monomethylated on the Gln150 amide nitrogen by the methyltransferase PrmB (Figure 1.24) (124).

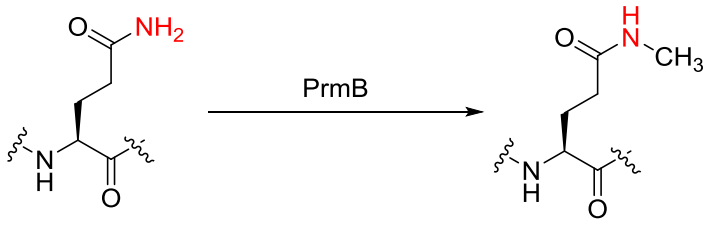


Figure 1.24: Methyltransferase PrmB catalyses the methylation of Q150 on L3.

The *PrmB* deletion strain in *E. coli* is cold-sensitive (slow growth at 22°C) (124). The deletion also affects ribosomal biogenesis leading to a decreased intensity of 30S and 50S peaks, and an increase in the intensity of intermediate peaks in ribosome profiles. Consistent with its function in ribosomal biogenesis, PrmB does not methylate unbound L3 or L3 in completely assembled ribosome. Instead, the highest catalytic rate was observed in the partially assembled ribosome or in the presence of RNA (124). No difference was observed on translation rate or accuracy, which can be reconciled with the fact that L3 is located on the surface of 50S.

iv) Lysine methylation of L7/L12

Methylation of proteins L7/L12 has been described in detail, but the methyltransferase remains unknown. Methylation of L7/L12 varies with environmental conditions; cold-shock increases methylation (125).

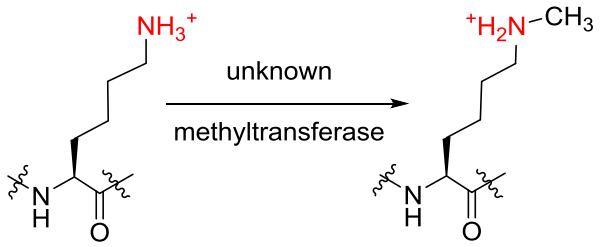


Figure 1.25: Ribosomal proteins L7/L12 are methylated by an unknown methyltransferase; methylation increases during cold-shock.

1.6.2.3. Acetylation

Although at least 50% of cytosolic eukaryotic proteins appear to be acetylated (126), the role of ribosomal protein acetylation remains poorly understood. Nα-Acetylation is catalysed by Nα-acetyltransferases which are acetylcoenzyme A-dependent. The modification is more prevalent in eukaryotes than in bacteria, in which ribosomal proteins S5, S18 and L7, as well as elongation Ef-Tu are acetylated. Ribosomal protein S5, S18 and L7 were identified as substrates for acetyltransferases rimJ (127,128), rimI (128) and rimL (129,130), respectively (Figure 1.26). There is little sequence similarity between acetyltransferases, although they catalyse similar substrates (rimI and J are alanine

acetyltransferases with 19% sequence similarity, whereas RimL is a serine acetyltransferase with 23% similarity to RimJ and 20% to RimI).



Figure 1.26: Acetyltransferases RimI/J/L catalyse the acetylation of ribosomal proteins S18, S5 and L7, respectively. Acetylated product of L7 is referred to as L12.

Acetylation of S5 by RimJ is regulated by the ribosome tertiary structure (131). Mutations in the central pseudoknot of 16S disturbs the 30S assembly process; moreover, it leads to reduction in S5 acetylation, which is located in the immediate vicinity of the pseudoknot, suggesting that RimJ acts during the assembly process (131). Hypo-acetylated S5 is thus a marker of defective ribosomal assembly. Overexpression of RimJ in strains with mutations in S5, which display a cold sensitivity phenotype and reduced translation accuracy reverses the translational defects (132). Crystal structure of the ribosome reveals that acetylation site is on the surface of the ribosome, therefore acetylation during assembly seems plausible. Despite the analogy in the nature of chemical modification, acetylation of S18 by RimI does not seem to affect either growth or quality of translation (133).

Ribosomal protein L12 is *N*-terminally acetylated by RimL to form L7 (134). RimL can act upon L12 at any stage during ribosomal biogenesis, as well as on unbound protein *in vitro* (135), in contrast to other acetyltransferases, e.g. RimJ, which only act on pre-assembled ribosomal proteins (see above). L12 acetylation varies with growth phase of *E. coli*, thereby altering the L7:L12 ratio: the relative amount of L7 increases during growth of *E. coli* and is highest in stationary phase (136). Acetylation is complete when bacteria are grown in minimal medium (136). There are four copies of L7/L12 protein in the *E. coli* ribosome (137) which together with a single copy of L10 form the L8 stalk complex (138). L7 acetylation increases the interaction with L10 (139). The stoichiometry of the stalk complex varies with growth temperature; mesophilic bacteria contain a pentameric (tetrameric L7/L12 plus L10) whilst thermophilic bacteria contain a heptameric stalk complex (hexameric L7/L12 plus L10). Eukaryotes only contain a pentameric stalk complex, whereas mesophilic archaea simultaneusly

contain pentameric and heptameric stalk complexes, the ratio of which depends on the growth phase (140). Ribosomes depleted of L7/L12 can bind Ef-Tu and Ef-G but cannot activate GTP hydrolysis (141). A complex $L12_6$.L10 structure has been solved and used to model the stalk on 50S (142), which can aid the rationalisation of the importance of L7/L12 proteins for initiation (143) but it is unclear why acetylation of L12 does not affect the translational mechanism.

1.6.2.4. β-Methylthiolation

Although β-methylthiolation of S12 at D88 has been identified (144) and ascribed to the catalytic activity by RimO (145,146), no details are available about the role of the modification for the translational process (Figure 1.27).

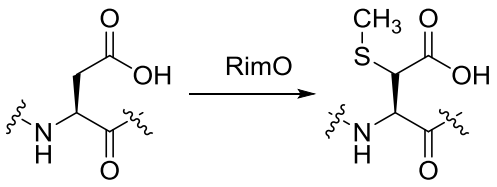


Figure 1.27: RimO catalyses the β-methylthiolation of ribosomal protein S12.

The role of S12 in fidelity of translation has been studied in detail (147); S12 has been established as the most important ribosomal protein for translational accuracy. One study used a mutational approach to demonstrate that variation of most residues on two highly conserved loops in S12 led to increased translational fidelity, suggesting their importance for tRNA selection. Variants K42A and R53A, were substantially resistant to the miscoding effects of paromomycin. Interestingly, variants of P44 and D88 were lethal, suggesting that they are crucial for the process of translation (148). However, β-methylthiolation of S12 was not shown to have strong effects on translation, suggesting that it may be less important for the function of the ribosome.

1.6.2.5. Phosphorylation

None of the published whole-protein mass spectrometry studies of bacterial ribosomal proteins identified phosphorylation of ribosomal proteins. Nonetheless, several studies on phosphoproteomes of prokaryotic cells reported a large number of ribosomal proteins to be phosphorylated (one study

reported as many as 24 ribosomal proteins in *Streptomyces coelicolor* to be phosphorylated). Phosphorylation was proposed to decrease ribosomal subunit association and decrease the rate of poly(U) translation (149). In *E. coli*, the ribophosphoproteome was reported to include S7, L7/L12, L16, L19 and L28 (150). Whole-protein analyses of ribosomal proteins did not detect any of the proposed phosphorylations (113,151).

1.6.2.6. Isopeptide bond formation

Ribosomal protein S11 undergoes a unique modification, whereby an isopeptide bond is formed by Nε of K2 and the carboxy terminus of A1, thereby exposing the A1 α-amino group as the side-chain (Figure 1.28) (152). The modification does not lead to a change in mass and therefore cannot be observed by whole-protein mass spectrometry. The *N*-terminus of S11 is located at the surface of the ribosome (although part of S11 is situated in proximity to the E-site on 30S) and is thus unlikely to affect ribosomal function.

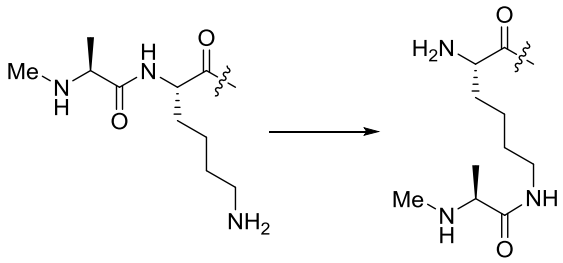


Figure 1.28: Isopeptide bond formation at the *N*-terminus of S11.

1.6.2.7. Ribosomal protein oxidation/hydroxylation

Oxidation of the ribosomal protein L16 was first reported by Brosius and Chen (71). Using micro-dansyl Edman procedure on peptides derived from protease digestion, they observed *N*-terminal *N*-methyl methionine as well as an 'unusual amino acid related to arginine', which during amino acid analysis migrated between ammonia and arginine (*ibid.*). The first comprehensive mass spectrometry study by Arnold and Reilly speculated that the unusual amino acid may be hydroxyarginine (113). Their intact protein analysis of L16 revealed a mass shift of +46 Da relative to the predicted mass of L16, suggesting that another oxidative modification may be present on L16. Nothing is known about the biological role of ribosomal protein hydroxylation. The preliminary work in the Schofield laboratory suggested that R81 may be the substrate of the *E. coli* ycfD (Figure 1.29).

Table 1.1: A summary of ribosomal protein modifications in *E. coli*

Ribosomal protein	Modification	Residue	Enzyme
S5	Ac	A1	RimJ
S11	Me	A1	MAK
S11	Isopeptide bond	A1	unknown
S12	β-Methyl thiolation	D88	RimO
S18	Ac	A1	RimI
L3	Me	Q150	PrmB
L12	Ac/Me	S1/K82	RimL/unknown
L11	9 Me	A1, K3, K40	PrmA
L12	Me	K82	unknown
L16	2OH, Me	M1 (Me), R81 (OH),	ycfD
L33	Me	A1	unknown

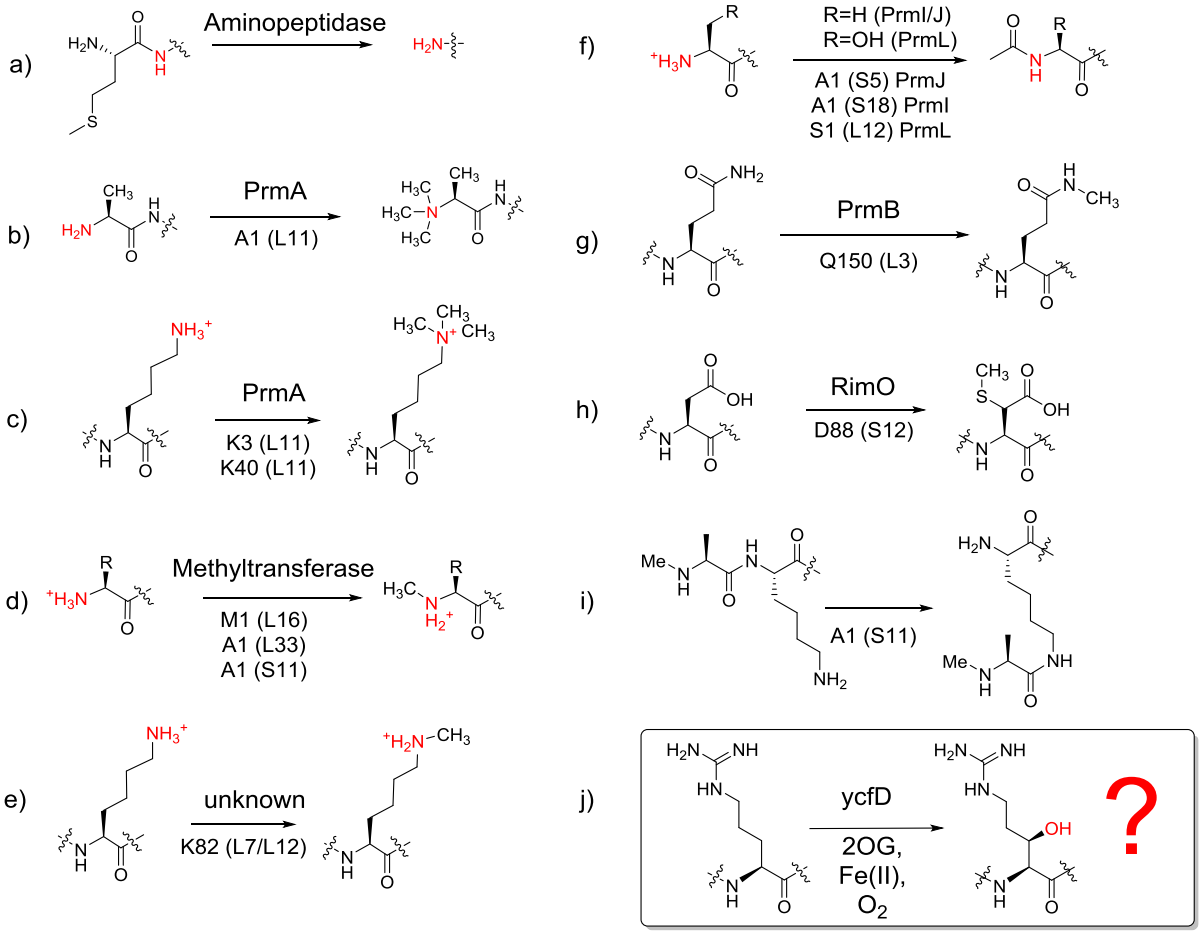


Figure 1.29: Post-translational modifications of bacterial ribosomal proteins. A) *N*-terminal methionine cleavage, b) α-amino trimethylation by PrmA; c) N-ε lysine trimethylation by PrmA; d) α-amino monomethylation by unknown methyltransferases; e) N-ε monomethylation of K82 on L7/L12; f) α-amino acetylation by acetyltransferases PrmI/J/L; g) side-chain amide methylation of Q150 on L3 by PrmB; h) β-methylthiolation of D88 on S12 by RimO; i) isopeptide bond formation at A1 on S11; j) proposed hydroxylation of R81 on L16 by the 2OG oxygenase ycfD.

1.6.3. Modifications of ribosome-associated factors

1.6.3.1. EF-Tu modifications

In *E. coli*, Ef-Tu is mono- and dimethylated on K56 (153). Dimethylation increases during nutrient deprivation (154). During logarithmic growth-phase, Ef-Tu is monomethylated; upon entry into stationary phase, it is converted to dimethyl-lysine; trimethyl-lysine is not observed (154). The methylation decreases the rate of GTPase activity of Ef-Tu (155). A recent report proposed that in *Pseudomonas aeruginosa* EF-Tu is trimethylated, thus chemically resembling phosphorylcholine; the latter is present in platelet-activating factor, leading to the hypothesis that Ef-Tu trimethylation serves as an epitope important for colonization of the respiratory tract and causing pneumonia (156). Such biological activity therefore depends on exo-ribosomal function of Ef-Tu.

1.6.3.2. Release factor (RF) glutamyl methylation

The methyltransferase PrmC methylates the side-chain of the residue Gln-240 of a universally conserved tripeptide motif GGQ of bacterial release factors RF1 and RF2 (Figure 1.30). Crystal structures of the ribosome-RF2 complex suggest that a potential function for the methylation of Gln240 is to rotate the amide nitrogen away from the site of nucleophilic attack, thus ensuring that the carbonyl is facing the catalytic water, enhancing its nucleophilicity (100); the proposal is supported by computational studies (157). PrmC inactivation has no effect on cell growth in rich media but reduces growth considerably on poor carbon sources, potentially suggesting that the expression of some genes needed for optimal growth under such conditions can become growth limiting as a result of inefficient translation termination (158). Interestingly, the modification site is within 15 Å of the proposed site of arginine hydroxylation (Figure 1.30).

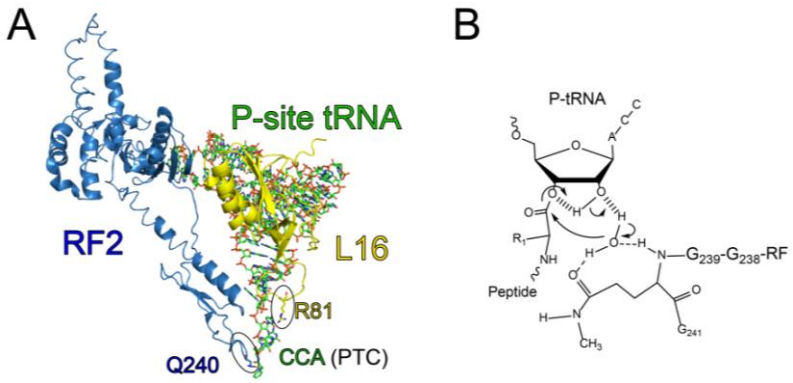


Figure 1.30: The methyltransferase PrmC catalyses the methylation of Q240 of the catalytically important GGQ motif on release factors RF1 and RF2. A) The modified residue (Q240) of RF2 (blue, PDB: 2X9R, (100)) protrudes into the peptidyl transferase centre (PTC) of the 50S subunit. The structural proximity to R81 on L16 (yellow), the proposed hydroxylation substrate of ycfD, is notable. B) The methylation is proposed to orient the side-chain amide of Q240 to enable hydrogen bond formation with the catalytically important nucleophilic water molecule, which leads to peptide release.

1.6.3.3. EF-P lysyl hydroxylation

Elongation factor P (EF-P) stimulates the peptidyl transferase activity of *E. coli* 70S (159). *In vitro*, it stimulates dipeptide formation between fMet-tRNA and puromycin, i.e. the formation of the first peptide bond (160). EF-P is homologous to the eukaryotic initiation factor 5A (eIF5a). A conserved basic residue is located at the tip of the loop in the *N*-terminal domain of EF-P/eIF5a; in eukaryotes, the lysine residue is modified to hypusine, the modification which is required for viability in yeast (161). The modification is carried out by the successive action of deoxyhypusine synthase and deoxyhypusine hydroxylase (162). No homologous enzymes are found in bacterial genomes. Instead, K34 on EF-P is modified by the sequential action of YjeK and YjeA, which lysinylate K34, and YfcM, which hydroxylates the lysinylated product at C4 or C5 position (Figure 1.31) (163). The lysinylated K34 may extend towards the PTC and stabilise the CCA-end of the P-site tRNA (163). Deletion of efp, yjeA or yjeK results in growth defects, increased sensitivity to antibiotic gentamicin and loss of virulence (164). The modification may assist the synthesis of proteins containing poly-proline motifs (165). The location of K34 modification is within 15 Å of L16 hydroxylation and RF2 methylation, suggesting that a ‘triangle’ of post-translational modifications surrounds the PTC centre of the ribosome (Figure 1.32). The modification may assist the synthesis of proteins containing poly-proline motifs (165). The location of the K34 modification is within 15 Å of the proposed ycfD-catalysed R81 hydroxylation on L16 and Q240 methylation on RF2. Thus, an apparent ‘triangle’ of post-translational modifications surrounds the PTC centre of the ribosome (Figure 1.32).

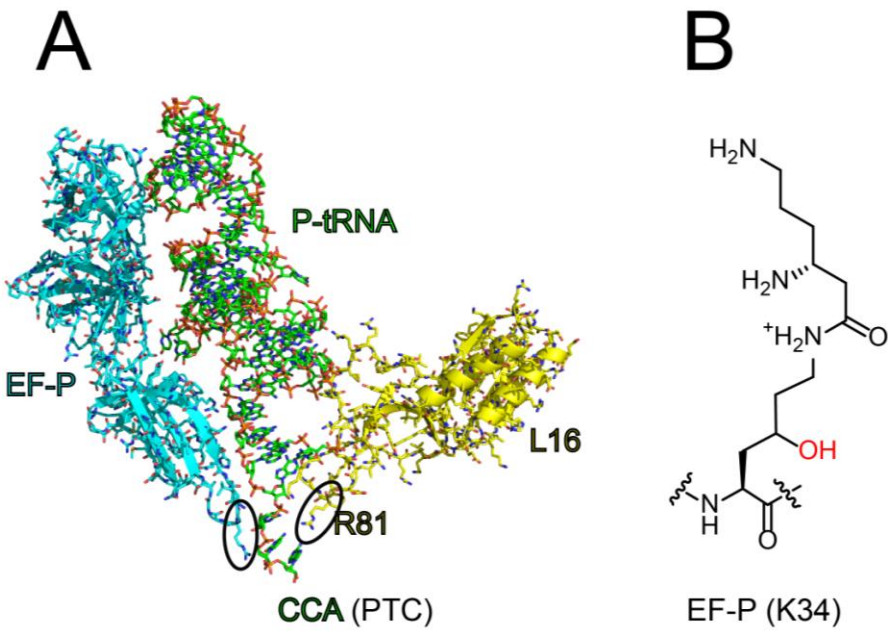


Figure 1.31: The oxygenase yfcM catalyses the hydroxylation of K34 of the catalytically important basic residue motif on a beta turn of the elongation factor EF-P. A) A view of the orientation of the EF-P molecule in the crystal structure of *T. thermophilus* 70S (PDB: 3HUW (166)). The analogous residue to the modified residue (K34) of EF-P (blue) in *E. coli* is arginine. It protrudes into the peptidyl transferase centre (PTC) of the 50S subunit between the P- and E-sites making several interactions with the 3' acceptor stem of the P-site bound tRNA. The structural proximity to R81 on L16 (yellow), the proposed hydroxylation substrate of ycfD, is notable. B) The chemical structure of modified K34 of EF-P.

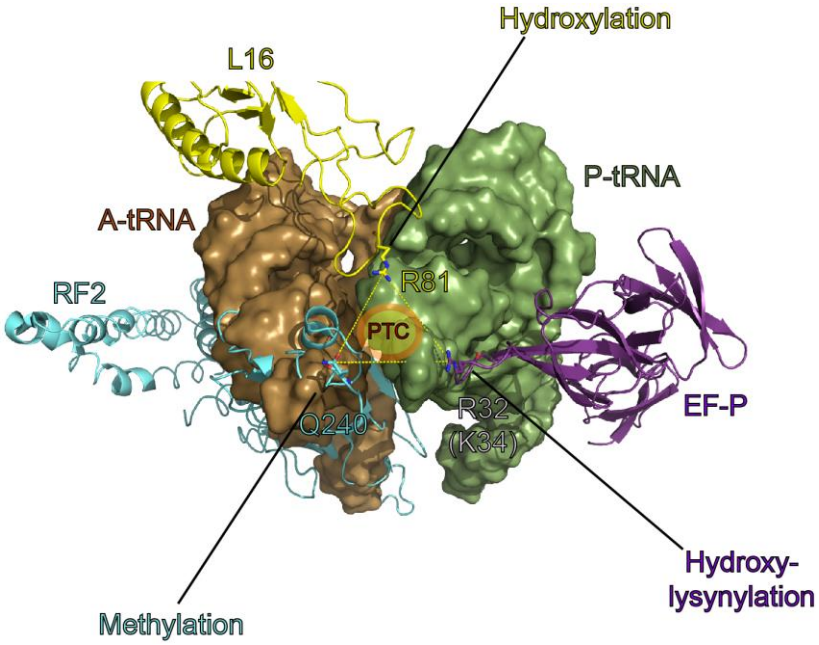


Figure 1.32: The 'triangle' of post-translational modifications of ribosomal proteins and ribosome-associated factors. The figure was created using the PDB entries 3HUX, 3HUW, 2X9T, 2WDK, 2WDL.

1.7. Summary

The *E. coli* 2OG oxygenase ycfD is a JmjC-domain containing 2OG oxygenase which functions as a 2*S*,3*R*-arginyl hydroxylase of the 50S protein L16 fragment peptide *in vitro*. It is distantly homologous to the HIF-1α asparaginyl hydroxylase FIH, which functions as a hypoxia sensor in humans and other animals. Furthermore, it shares structural homology with JmjC-containing 2OG oxygenases NO66 and MINA53 with unknown functions.

It is unknown whether the activity of ycfD *in vitro* against a ribosomal protein peptide fragment is biologically relevant. A demonstration of its function as a ribosomal hydroxylase would expand the known biological functions of 2OG oxygenases to include the hydroxylation of the translational apparatus, in particular ribosomal proteins. Concomitantly, it would expand the plethora of ribosome modifications to include ribosomal protein hydroxylation. A large number of post-translational modifications of bacterial ribosomes have been reported, including *N*-terminal methionine cleavage, methylation, acetylation, β-methylthiolation, and phosphorylation. The functions of most of the modifications remain unknown. However, deletions of a limited number of ribosome-modifying enzymes reveal a growth phenotype, such as cold-sensitivity (PrmB, RimJ) and translation defects, such as reduced translation accuracy (RimJ), which suggest functional importance of the modifications. The ycfD hydroxylation substrate R81 of L16 is located in the proximity of the PTC on the 50S subunit, within ~15 Å of known post-translational modifications on RF-2 and EF-P, deletions of which lead to severe phenotypes.

1.8. Aims of the work described in this thesis

1) To investigate the potential role of *ycfD* as a ribosomal protein hydroxylase in cells, including its specificity and oxygen dependence, by liquid-chromatography-mass spectrometry (LC-MS:
    a. Optimise LC-MS conditions for coverage and accuracy;
    b. Analyse ribosomal protein masses from the wild-type and *ycfD*-deleted (KO) *E. coli* strains;
    c. Assign the ycfD-dependent modification by LC-MS-MS;
    d. Demonstrate the dependence of the modification on catalytic activity of ycfD by ectopic over-expression of WT and catalytically inactive ycfD variants;
    e. Study the ycfD oxygen-dependence in cells by culturing *E. coli* WT cells in a hypoxic chamber at variable oxygen concentrations.
2) To investigate the biochemical function of *ycfD* in *E. coli* under conditions that lead to exacerbated phenotypes of other ribosomal modifications
    a. Limited nutrient conditions;
    b. Growth at low and high temperature;
    c. Global translation rate;
    d. Antibiotic sensitivity.
3) To analyse the evolutionary conservation of ycfD homologues:
    a. Occurrence of ycfD in prokaryotic phyla;
    b. Biochemical characterisation of extremophilic ycfD homologue.
4) To investigate the structural features of ycfD catalysis of ribosomal protein L16 by X-ray crystallography:
    a. Obtain a soluble ycfD-substrate fragment complex;
    b. Analyse the structural features of ycfD catalysis, as compared to human homologues.
5) To investigate the potential role of the ycfD human homologue MINA53 as a ribosomal protein hydroxylase:
    a. Develop an LC-MS assay to measure the biochemical activity of MINA53 and its variants in cells, including under limited oxygen availability;
    b. Study the effects of cancer mutations of *MINA53* on biochemical activity of MINA53 by LC-MS.

1.9. References

1. Loenarz, C., and Schofield, C. J. (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* **36**, 7-18
2. Loenarz, C., Mecinovic, J., Chowdhury, R., McNeill, L. A., Flashman, E., and Schofield, C. J. (2009) Evidence for a stereoelectronic effect in human oxygen sensing. *Angew Chem Int Ed Engl* **48**, 1784-1787
3. Bretscher, L. E., Jenkins, C. L., Taylor, K. M., DeRider, M. L., and Raines, R. T. (2001) Conformational stability of collagen relies on a stereoelectronic effect. *J Am Chem Soc* **123**, 777-778
4. Santen, R. A. v., and Neurock, M. (2006) *Molecular heterogeneous catalysis : a conceptual and computational approach*, Wiley-VCH, Weinheim
5. Dunford, H. B., and North Atlantic Treaty Organization. Scientific Affairs Division. (1982) *The biological chemistry of iron*, D. Reidel, Boston, U.S.A.
6. Schofield, C. J., and Zhang, Z. (1999) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol* **9**, 722-731
7. Clifton, I. J., McDonough, M. A., Ehrismann, D., Kershaw, N. J., Granatino, N., and Schofield, C. J. (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* **100**, 644-669
8. Lange, S. J., and Que, L., Jr. (1998) Oxygen activating nonheme iron enzymes. *Curr Opin Chem Biol* **2**, 159-172
9. Hegg, E. L., and Que, L., Jr. (1997) The 2-His-1-carboxylate facial triad--an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem* **250**, 625-629
10. Que, L., Jr., and Ho, R. Y. (1996) Dioxygen Activation by Enzymes with Mononuclear Non-Heme Iron Active Sites. *Chem Rev* **96**, 2607-2624
11. Lind, J. (1983) A Treatise of the Scurvy. *Nutr Rev* **41**, 155-157
12. Rosenfeld, L. (1997) Vitamine-vitamin. The early years of discovery. *Clin Chem* **43**, 680-685
13. Stetten, M. R. (1949) Some Aspects of the Metabolism of Hydroxyproline, Studies with the Aid of Isotopic Nitrogen. *J Biol Chem* **181**, 31-37
14. Hutton, J. J., Jr., Kaplan, A., and Udenfriend, S. (1967) Conversion of the amino acid sequence gly-pro-pro in protein to gly-pro-hyp by collagen proline hydroxylase. *Arch Biochem Biophys* **121**, 384-391
15. Murad, S., Grove, D., Lindberg, K. A., Reynolds, G., Sivarajah, A., and Pinnell, S. R. (1981) Regulation of Collagen-Synthesis by Ascorbic-Acid. *P Natl Acad Sci USA* **78**, 2879-2882
16. Roach, P. L., Clifton, I. J., Fulop, V., Harlos, K., Barton, G. J., Hajdu, J., Andersson, I., Schofield, C. J., and Baldwin, J. E. (1995) Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* **375**, 700-704
17. Valegard, K., van Scheltinga, A. C., Lloyd, M. D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H. J., Baldwin, J. E., Schofield, C. J., Hajdu, J., and Andersson, I. (1998) Structure of a cephalosporin synthase. *Nature* **394**, 805-809
18. Schofield, C. J., Baldwin, J. E., Byford, M. F., Clifton, I., Hajdu, J., Hensgens, C., and Roach, P. (1997) Proteins of the penicillin biosynthesis pathway. *Curr Opin Struct Biol* **7**, 857-864
19. Roach, P. L., Clifton, I. J., Hensgens, C. M., Shibata, N., Schofield, C. J., Hajdu, J., and Baldwin, J. E. (1997) Structure of isopenicillin N synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **387**, 827-830
20. Iyer, L. M., Abhiman, S., de Souza, R. F., and Aravind, L. (2010) Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. *Nucleic Acids Res* **38**, 5261-5279
21. McDonough, M. A., Loenarz, C., Chowdhury, R., Clifton, I. J., and Schofield, C. J. (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struct Biol* **20**, 659-672

22. Hangasky, J. A., Taabazuing, C. Y., Valliere, M. A., and Knapp, M. J. (2013) Imposing function down a (cupin)-barrel: secondary structure and metal stereochemistry in the alpha KG-dependent oxygenases. *Metallomics* **5**, 287-301
23. Kruidenier, L., Chung, C. W., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H., Eberhard, D., Hutchinson, S., Jones, E., Katso, R., Leveridge, M., Mander, P. K., Mosley, J., Ramirez-Molina, C., Rowland, P., Schofield, C. J., Sheppard, R. J., Smith, J. E., Swales, C., Tanner, R., Thomas, P., Tumber, A., Drewes, G., Oppermann, U., Patel, D. J., Lee, K., and Wilson, D. M. (2012) A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404-408
24. Solomon, E. I., Brunold, T. C., Davis, M. I., Kemsley, J. N., Lee, S. K., Lehnert, N., Neese, F., Skulan, A. J., Yang, Y. S., and Zhou, J. (2000) Geometric and electronic structure/function correlations in non-heme iron enzymes. *Chem Rev* **100**, 235-349
25. Costas, M., Mehn, M. P., Jensen, M. P., and Que, L., Jr. (2004) Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem Rev* **104**, 939-986
26. Topol, I. A., Nemukhin, A. V., Salnikow, K., Cachau, R. E., Abashkin, Y. G., Kasprzak, K. S., and Burt, S. K. (2006) Quantum chemical modeling of reaction mechanism for 2-oxoglutarate dependent enzymes: effect of substitution of iron by nickel and cobalt. *J Phys Chem A* **110**, 4223-4228
27. Klose, R. J., Kallin, E. M., and Zhang, Y. (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**, 715-727
28. Loenarz, C., and Schofield, C. J. (2008) Expanding chemical biology of 2-oxoglutarate oxygenases. *Nat Chem Biol* **4**, 152-156
29. Yang, M., Hardy, A. P., Chowdhury, R., Loik, N. D., Scotti, J. S., McCullagh, J. S. O., Claridge, T. D. W., McDonough, M. A., Ge, W., and Schofield, C. J. (2013) Substrate Selectivity Analyses of Factor Inhibiting Hypoxia-Inducible Factor. *Angew Chem Int Edit* **52**, 1700-1704
30. Schofield, C. J., and Ratcliffe, P. J. (2004) Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Biol* **5**, 343-354
31. Guzy, R. D., Hoyos, B., Robin, E., Chen, H., Liu, L., Mansfield, K. D., Simon, M. C., Hammerling, U., and Schumacker, P. T. (2005) Mitochondrial complex III is required for hypoxia-induced ROS production and cellular oxygen sensing. *Cell Metab* **1**, 401-408
32. Loenarz, C., Coleman, M. L., Boleininger, A., Schierwater, B., Holland, P. W., Ratcliffe, P. J., and Schofield, C. J. (2011) The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO Rep* **12**, 63-70
33. Wang, G. L., Jiang, B. H., Rue, E. A., and Semenza, G. L. (1995) Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc Natl Acad Sci U S A* **92**, 5510-5514
34. Lisy, K., and Peet, D. J. (2008) Turn me on: regulating HIF transcriptional activity. *Cell Death Differ* **15**, 642-649
35. Huang, L. E., Arany, Z., Livingston, D. M., and Bunn, H. F. (1996) Activation of hypoxia-inducible transcription factor depends primarily upon redox-sensitive stabilization of its alpha subunit. *J Biol Chem* **271**, 32253-32259
36. Yu, A. Y., Frid, M. G., Shimoda, L. A., Wiener, C. M., Stenmark, K., and Semenza, G. L. (1998) Temporal, spatial, and oxygen-regulated expression of hypoxia-inducible factor-1 in the lung. *Am J Physiol* **275**, 818-826
37. Semenza, G. L. (2003) Targeting HIF-1 for cancer therapy. *Nat Rev Cancer* **3**, 721-732
38. Berra, E., Benizri, E., Ginouves, A., Volmat, V., Roux, D., and Pouyssegur, J. (2003) HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1alpha in normoxia. *EMBO J* **22**, 4082-4090
39. Tromans, D. (1998) Temperature and pressure dependent solubility of oxygen in water: a thermodynamic analysis. *Hydrometallurgy* **48**, 327-342

40. Ratcliffe, P. J. (2006) Understanding hypoxia signalling in cells--a new therapeutic opportunity? *Clin Med* **6**, 573-578
41. Chowdhury, R., McDonough, M. A., Mecinovic, J., Loenarz, C., Flashman, E., Hewitson, K. S., Domene, C., and Schofield, C. J. (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* **17**, 981-989
42. Ivan, M., Kondo, K., Yang, H., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S., and Kaelin, W. G., Jr. (2001) HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science* **292**, 464-468
43. Hewitson, K. S., McNeill, L. A., Riordan, M. V., Tian, Y. M., Bullock, A. N., Welford, R. W., Elkins, J. M., Oldham, N. J., Bhattacharya, S., Gleadle, J. M., Ratcliffe, P. J., Pugh, C. W., and Schofield, C. J. (2002) Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J Biol Chem* **277**, 26351-26355
44. Freedman, S. J., Sun, Z. Y. J., Poy, F., Kung, A. L., Livingston, D. M., Wagner, G., and Eck, M. J. (2002) Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. *P Natl Acad Sci USA* **99**, 5367-5372
45. Lando, D., Peet, D. J., Whelan, D. A., Gorman, J. J., and Whitelaw, M. L. (2002) Asparagine hydroxylation of the HIF transactivation domain a hypoxic switch. *Science* **295**, 858-861
46. Koivunen, P., Hirsila, M., Gunzler, V., Kivirikko, K. I., and Myllyharju, J. (2004) Catalytic properties of the asparaginyl hydroxylase (FIH) in the oxygen sensing pathway are distinct from those of its prolyl 4-hydroxylases. *J Biol Chem* **279**, 9899-9904
47. Stolze, I. P., Tian, Y. M., Appelhoff, R. J., Turley, H., Wykoff, C. C., Gleadle, J. M., and Ratcliffe, P. J. (2004) Genetic analysis of the role of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible factor (HIF) in regulating HIF transcriptional target genes. *J Biol Chem* **279**, 42719-42725
48. Dayan, F., Roux, D., Brahimi-Horn, M. C., Pouyssegur, J., and Mazure, N. M. (2006) The oxygen sensor factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes through the bifunctional transcriptional character of hypoxia-inducible factor-1alpha. *Cancer Res* **66**, 3688-3698
49. Greijer, A. E., van der Groep, P., Kemming, D., Shvarts, A., Semenza, G. L., Meijer, G. A., van de Wiel, M. A., Belien, J. A., van Diest, P. J., and van der Wall, E. (2005) Up-regulation of gene expression by hypoxia is mediated predominantly by hypoxia-inducible factor 1 (HIF-1). *J Pathol* **206**, 291-304
50. Maynard, M. A., and Ohh, M. (2007) The role of hypoxia-inducible factors in cancer. *Cell Mol Life Sci* **64**, 2170-2180
51. Zhong, H., De Marzo, A. M., Laughner, E., Lim, M., Hilton, D. A., Zagzag, D., Buechler, P., Isaacs, W. B., Semenza, G. L., and Simons, J. W. (1999) Overexpression of hypoxia-inducible factor 1alpha in common human cancers and their metastases. *Cancer Res* **59**, 5830-5835
52. Semenza, G. L. (2007) Evaluation of HIF-1 inhibitors as anticancer agents. *Drug Discov Today* **12**, 853-859
53. Ferguson, J. E., 3rd, Wu, Y., Smith, K., Charles, P., Powers, K., Wang, H., and Patterson, C. (2007) ASB4 is a hydroxylation substrate of FIH and promotes vascular differentiation via an oxygen-dependent mechanism. *Mol Cell Biol* **27**, 6407-6419
54. Coleman, M. L., McDonough, M. A., Hewitson, K. S., Coles, C., Mecinovic, J., Edelmann, M., Cook, K. M., Cockman, M. E., Lancaster, D. E., Kessler, B. M., Oldham, N. J., Ratcliffe, P. J., and Schofield, C. J. (2007) Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038
55. Kelly, L., McDonough, M. A., Coleman, M. L., Ratcliffe, P. J., and Schofield, C. J. (2009) Asparagine beta-hydroxylation stabilizes the ankyrin repeat domain fold. *Mol Biosyst* **5**, 52-58
56. Coleman, M. L., and Ratcliffe, P. J. (2009) Signalling cross talk of the HIF system: involvement of the FIH protein. *Curr Pharm Des* **15**, 3904-3907

57. Bertout, J. A., Patel, S. A., and Simon, M. C. (2008) The impact of O2 availability on human cancer. *Nat Rev Cancer* **8**, 967-975
58. Green, J., Crack, J. C., Thomson, A. J., and LeBrun, N. E. (2009) Bacterial sensors of oxygen. *Curr Opin Microbiol* **12**, 145-151
59. Khoroshilova, N., Popescu, C., Munck, E., Beinert, H., and Kiley, P. J. (1997) Iron-sulfur cluster disassembly in the FNR protein of Escherichia coli by O-2: [4Fe-4S] to [2Fe-2S] conversion with loss of biological activity. *P Natl Acad Sci USA* **94**, 6087-6092
60. Kiley, P. J., and Beinert, H. (1998) Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster. *Fems Microbiol Rev* **22**, 341-352
61. Gillesgonzalez, M. A., Gonzalez, G., and Perutz, M. F. (1995) Kinase-Activity of Oxygen Sensor Fixl Depends on the Spin-State of Its Heme Iron. *Biochemistry-Us* **34**, 232-236
62. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T., and Sedgwick, B. (2002) Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. *Nature* **419**, 174-178
63. Falnes, P. O., Johansen, R. F., and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **419**, 178-182
64. Eichhorn, E., van der Ploeg, J. R., Kertesz, M. A., and Leisinger, T. (1997) Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from Escherichia coli. *J Biol Chem* **272**, 23031-23036
65. Cockman, M. E., Webb, J. D., Kramer, H. B., Kessler, B. M., and Ratcliffe, P. J. (2009) Proteomics-based identification of novel factor inhibiting hypoxia-inducible factor (FIH) substrates indicates widespread asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Mol Cell Proteomics* **8**, 535-546
66. Hewitson, K. S., McNeill, L. A., Riordan, M. V., Tian, Y. M., Bullock, A. N., Welford, R. W., Elkins, J. M., Oldham, N. J., Bhattacharya, S., Gleadle, J. M., Ratcliffe, P. J., Pugh, C. W., and Schofield, C. J. (2002) Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J Biol Chem* **277**, 26351-26355
67. Webby, C. J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M. L., Schmitz, C., Butler, D. S., Yates, J. R., 3rd, Delahunty, C. M., Hahn, P., Lengeling, A., Mann, M., Proudfoot, N. J., Schofield, C. J., and Bottger, A. (2009) Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90-93
68. Brien, G. L., Gambero, G., O'Connell, D. J., Jerman, E., Turner, S. A., Egan, C. M., Dunne, E. J., Jurgens, M. C., Wynne, K., Piao, L., Lohan, A. J., Ferguson, N., Shi, X., Sinha, K. M., Loftus, B. J., Cagney, G., and Bracken, A. P. (2012) Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. *Nat Struct Mol Biol* **19**, 1273-1281
69. Sinha, K. M., Yasuda, H., Coombes, M. M., Dent, S. Y., and de Crombrugghe, B. (2010) Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *EMBO J* **29**, 68-79
70. Ge, W., Wolf, A., Feng, T., Ho, C. H., Sekirnik, R., Zayer, A., Granatino, N., Cockman, M. E., Loenarz, C., Loik, N. D., Hardy, A. P., Claridge, T. D., Hamed, R. B., Chowdhury, R., Gong, L., Robinson, C. V., Trudgian, D. C., Jiang, M., Mackeen, M. M., McCullagh, J. S., Gordiyenko, Y., Thalhammer, A., Yamamoto, A., Yang, M., Liu-Yi, P., Zhang, Z., Schmidt-Zachmann, M., Kessler, B. M., Ratcliffe, P. J., Preston, G. M., Coleman, M. L., and Schofield, C. J. (2012) Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol* **8**, 960-962
71. Brosius, J., and Chen, R. (1976) The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett* **68**, 105-109
72. Tao, Y., Wu, M., Zhou, X., Yin, W., Hu, B., de Crombrugghe, B., Sinha, K. M., and Zang, J. (2013) Structural insights into histone demethylase NO66 in interaction with osteoblast specific transcription factor Osterix and gene repression. *J Biol Chem*

73. Dann, C. E., 3rd, Bruick, R. K., and Deisenhofer, J. (2002) Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci USA* **99**, 15351-15356
74. Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1α. *J Biol Chem* **278**, 1802-1806
75. Del Rizzo, P. A., Krishnan, S., and Trievel, R. C. (2012) Crystal structure and functional analysis of JMJD5 indicate an alternate specificity and function. *Mol Cell Biol* **32**, 4044-4052
76. Ng, S. S., Kavanagh, K. L., McDonough, M. A., Butler, D., Pilka, E. S., Lienard, B. M., Bray, J. E., Savitsky, P., Gileadi, O., von Delft, F., Rose, N. R., Offer, J., Scheinost, J. C., Borowski, T., Sundstrom, M., Schofield, C. J., and Oppermann, U. (2007) Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87-91
77. Klose, R. J., Kallin, E. M., and Zhang, Y. (2006) JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**, 715-727
78. Sievers, A., Beringer, M., Rodnina, M. V., and Wolfenden, R. (2004) The ribosome as an entropy trap. *Proc Natl Acad Sci U S A* **101**, 7897-7901
79. Schmeing, T. M., and Ramakrishnan, V. (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234-1242
80. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905-920
81. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. *Nature* **407**, 327-339
82. Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M., and Cate, J. H. (2005) Structures of the bacterial ribosome at 3.5 A resolution. *Science* **310**, 827-834
83. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* **292**, 883-896
84. Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* **334**, 941-948
85. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A., and Ban, N. (2010) Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1. *Science*
86. Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011) The structure of the eukaryotic ribosome at 3.0 A resolution. *Science* **334**, 1524-1529
87. Simonovic, M., and Steitz, T. A. (2009) A structural view on the mechanism of the ribosome-catalyzed peptide bond formation. *Biochim Biophys Acta* **1789**, 612-623
88. Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) Functional insights from the structure of the 30S ribosomal subunit and its interactions with antibiotics. *Nature* **407**, 340-348
89. Moazed, D., and Noller, H. F. (1986) Transfer RNA shields specific nucleotides in 16S ribosomal RNA from attack by chemical probes. *Cell* **47**, 985-994
90. Ramakrishnan, V. (2008) What we have learned from ribosome structures. *Biochem Soc T* **36**, 567-574
91. Battle, D. J., and Doudna, J. A. (2002) Specificity of RNA-RNA helix recognition. *P Natl Acad Sci USA* **99**, 11676-11681
92. Bock, L. V., Blau, C., Schroder, G. F., Davydov, II, Fischer, N., Stark, H., Rodnina, M. V., Vaiana, A. C., and Grubmuller, H. (2013) Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat Struct Mol Biol* **20**, 1390-1396
93. Scolnick, E., Tompkins, R., Caskey, T., and Nirenberg, M. (1968) Release factors differing in specificity for terminator codons. *Proc Natl Acad Sci U S A* **61**, 768-774

94. Youngman, E. M., He, S. L., Nikstad, L. J., and Green, R. (2007) Stop codon recognition by release factors induces structural rearrangement of the ribosomal decoding center that is productive for peptide release. *Mol Cell* **28**, 533-543
95. Mora, L., Heurgue-Hamard, V., Champ, S., Ehrenberg, M., Kisselev, L. L., and Buckingham, R. H. (2003) The essential role of the invariant GGQ motif in the function and stability in vivo of bacterial release factors RF1 and RF2. *Mol Microbiol* **47**, 267-275
96. Frolova, L. Y., Tsivkovskii, R. Y., Sivolobova, G. F., Oparina, N. Y., Serpinsky, O. I., Blinov, V. M., Tatkov, S. I., and Kisselev, L. L. (1999) Mutations in the highly conserved GGQ motif of class 1 polypeptide release factors abolish ability of human eRF1 to trigger peptidyl-tRNA hydrolysis. *RNA* **5**, 1014-1020
97. Laurberg, M., Asahara, H., Korostelev, A., Zhu, J., Trakhanov, S., and Noller, H. F. (2008) Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852-857
98. Weixlbaumer, A., Jin, H., Neubauer, C., Voorhees, R. M., Petry, S., Kelley, A. C., and Ramakrishnan, V. (2008) Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953-956
99. Korostelev, A., Asahara, H., Lancaster, L., Laurberg, M., Hirschi, A., Zhu, J., Trakhanov, S., Scott, W. G., and Noller, H. F. (2008) Crystal structure of a translation termination complex formed with release factor RF2. *Proc Natl Acad Sci U S A* **105**, 19684-19689
100. Jin, H., Kelley, A. C., Loakes, D., and Ramakrishnan, V. (2010) Structure of the 70S ribosome bound to release factor 2 and a substrate analog provides insights into catalysis of peptide release. *Proc Natl Acad Sci U S A* **107**, 8593-8598
101. Woodson, S. A. (2008) RNA folding and ribosome assembly. *Curr Opin Chem Biol* **12**, 667-673
102. Decatur, W. A., and Fournier, M. J. (2002) rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-351
103. Shajani, Z., Sykes, M. T., and Williamson, J. R. (2011) Assembly of bacterial ribosomes. *Annu Rev Biochem* **80**, 501-526
104. Green, R., and Noller, H. F. (1999) Reconstitution of functional 50S ribosomes from in vitro transcripts of Bacillus stearothermophilus 23S rRNA. *Biochemistry-Us* **38**, 1772-1779
105. Krzyzosiak, W., Denman, R., Nurse, K., Hellmann, W., Boublik, M., Gehrke, C. W., Agris, P. F., and Ofengand, J. (1987) In vitro synthesis of 16S ribosomal RNA containing single base changes and assembly into a functional 30S ribosome. *Biochemistry-Us* **26**, 2353-2364
106. Bugl, H., Fauman, E. B., Staker, B. L., Zheng, F., Kushner, S. R., Saper, M. A., Bardwell, J. C., and Jakob, U. (2000) RNA methylation under heat shock control. *Mol Cell* **6**, 349-360
107. Widerak, M., Kern, R., Malki, A., and Richarme, G. (2005) U2552 methylation at the ribosomal A-site is a negative modulator of translational accuracy. *Gene* **347**, 109-114
108. Pokkunuri, I., and Champney, W. S. (2007) Characteristics of a 50S ribosomal subunit precursor particle as a substrate for ermE methyltransferase activity and erythromycin binding in Staphylococcus aureus. *RNA Biol* **4**, 147-153
109. Gutgsell, N. S., Deutscher, M. P., and Ofengand, J. (2005) The pseudouridine synthase RluD is required for normal ribosome assembly and function in Escherichia coli. *RNA* **11**, 1141-1152
110. Poehlsgaard, J., and Douthwaite, S. (2005) The bacterial ribosome as a target for antibiotics. *Nat Rev Microbiol* **3**, 870-881
111. Williams, D. J., Boots, J. L., and Hall, K. B. (2001) Thermodynamics of 2'-ribose substitutions in UUCG tetraloops. *RNA* **7**, 44-53
112. Wilson, D. N., and Nierhaus, K. H. (2007) The weird and wonderful world of bacterial ribosome regulation. *Crit Rev Biochem Mol Biol* **42**, 187-219
113. Arnold, R. J., and Reilly, J. P. (1999) Observation of Escherichia coli ribosomal proteins and their posttranslational modifications by mass spectrometry. *Anal Biochem* **269**, 105-112
114. Stock, A., Clarke, S., Clarke, C., and Stock, J. (1987) N-terminal methylation of proteins: structure, function and specificity. *FEBS Lett* **220**, 8-14
115. Sherman, F., Stewart, J. W., and Tsunasawa, S. (1985) Methionine or Not Methionine at the Beginning of a Protein. *Bioessays* **3**, 27-31

116. Wittmannliebold, B., and Pannenbecker, R. (1976) Primary Structure of Protein L33 from Large Subunit of Escherichia-Coli Ribosome. *FEBS Lett* **68**, 115-118
117. Henry, G. D., Winstanley, M. A., Dalgarno, D. C., Scott, G. M. M., Levine, B. A., and Trayer, I. P. (1985) Characterization of the Actin-Binding Site on the Alkali Light Chain of Myosin. *Biochimica Et Biophysica Acta* **830**, 233-243
118. Colson, C. (1977) Genetics of ribosomal protein methylation in Escherichia coli. I. A mutant deficient in methylation of protein L11. *Mol Gen Genet* **154**, 167-173
119. Cameron, D. M., Gregory, S. T., Thompson, J., Suh, M. J., Limbach, P. A., and Dahlberg, A. E. (2004) Thermus thermophilus L11 methyltransferase, PrmA, is dispensable for growth and preferentially modifies free ribosomal protein L11 prior to ribosome assembly. *J Bacteriol* **186**, 5819-5825
120. Demirci, H., Gregory, S. T., Dahlberg, A. E., and Jogl, G. (2008) Multiple-site trimethylation of ribosomal protein L11 by the PrmA methyltransferase. *Structure* **16**, 1059-1066
121. Amaro, A. M., and Jerez, C. A. (1984) Methylation of Ribosomal-Proteins in Bacteria - Evidence of Conserved Modification of the Eubacterial 50s Subunit. *J Bacteriol* **158**, 84-93
122. Vanet, A., Plumbridge, J. A., Guerin, M. F., and Alix, J. H. (1994) Ribosomal protein methylation in Escherichia coli: the gene prmA, encoding the ribosomal protein L11 methyltransferase, is dispensable. *Mol Microbiol* **14**, 947-958
123. Okamoto, Y., and Yount, R. G. (1985) Identification of an active site peptide of skeletal myosin after photoaffinity labeling with N-(4-azido-2-nitrophenyl)-2-aminoethyl diphosphate. *Proc Natl Acad Sci U S A* **82**, 1575-1579
124. Lhoest, J., and Colson, C. (1977) Genetics of ribosomal protein methylation in Escherichia coli. II. A mutant lacking a new type of methylated amino acid, N5-methylglutamine, in protein L3. *Mol Gen Genet* **154**, 175-180
125. Chang, F. N. (1978) Temperature-dependent variation in the extent of methylation of ribosomal proteins L7 and L12 in Escherichia coli. *J Bacteriol* **135**, 1165-1166
126. Polevoda, B., and Sherman, F. (2003) N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J Mol Biol* **325**, 595-622
127. Janda, I., Kitakawa, M., and Isono, K. (1985) Gene Rpmf for Ribosomal-Protein L32 and Gene Rimj for a Ribosomal-Protein Acetylating Enzyme Are Located near Pyrc (23.4 Min) in Escherichia-Coli. *Mol Gen Genet* **201**, 433-436
128. Yoshikawa, A., Isono, S., Sheback, A., and Isono, K. (1987) Cloning and Nucleotide Sequencing of the Genes Rimi and Rimj Which Encode Enzymes Acetylating Ribosomal-Proteins S18 and S5 of Escherichia-Coli-K12. *Mol Gen Genet* **209**, 481-488
129. Vetting, M. W., de Carvalho, L. P. S., Roderick, S. L., and Blanchard, J. S. (2005) A novel dimeric structure of the RimL N-alpha-acetyltransferase from Salmonella typhimurium. *J Biol Chem* **280**, 22108-22114
130. Tanaka, S., Matsushita, Y., Yoshikawa, A., and Isono, K. (1989) Cloning and Molecular Characterization of the Gene Riml Which Encodes an Enzyme Acetylating Ribosomal Protein-L12 of Escherichia-Coli-K12. *Mol Gen Genet* **217**, 289-293
131. Poot, R. A., Jeeninga, R. E., Pleij, C. W. A., and vanDuin, J. (1997) Acetylation of ribosomal protein S5 affected by defects in the central pseudoknot in 16S ribosomal RNA? *FEBS Lett* **401**, 175-179
132. Roy-Chaudhuri, B., Kirthi, N., Kelley, T., and Culver, G. M. (2008) Suppression of a cold-sensitive mutation in ribosomal protein S5 reveals a role for RimJ in ribosome biogenesis. *Mol Microbiol* **68**, 1547-1559
133. Vetting, M. W., Bareich, D. C., Yu, M., and Blanchard, J. S. (2008) Crystal structure of RimI from Salmonella typhimurium LT2, the GNAT responsible for N(alpha)-acetylation of ribosomal protein S18. *Protein Sci* **17**, 1781-1790
134. Terhorst, C., Moller, W., Laursen, R., and Wittmann.B. (1973) Primary Structure of an Acidic Protein from 50-S Ribosomes of Escherichia-Coli Which Is Involved in Gtp Hydrolysis Dependent on Elongation Factors G and T. *Eur J Biochem* **34**, 138-152

135. Miao, L., Fang, H., Li, Y., and Chen, H. (2007) Studies of the in vitro Nalpha-acetyltransferase activities of E. coli RimL protein. *Biochem Biophys Res Commun* **357**, 641-647
136. Ramagopa.S, and Subraman.Ar. (1974) Alteration in Acetylation Level of Ribosomal-Protein L12 during Growth-Cycle of Escherichia-Coli. *P Natl Acad Sci USA* **71**, 2136-2140
137. Subramanian, A. R. (1975) Copies of Proteins L7 and L12 and Heterogeneity of Large Subunit of Escherichia-Coli Ribosome. *J Mol Biol* **95**, 1-&
138. Pettersson, I., Hardy, S. J. S., and Liljas, A. (1976) Ribosomal-Protein L8 Is a Complex of L7-L12 and L10. *FEBS Lett* **64**, 135-138
139. Gordiyenko, Y., Deroo, S., Zhou, M., Videler, H., and Robinson, C. V. (2008) Acetylation of L12 increases interactions in the Escherichia coli ribosomal stalk complex. *J Mol Biol* **380**, 404-414
140. Gordiyenko, Y., Videler, H., Zhou, M., McKay, A. R., Fucini, P., Biegel, E., Muller, V., and Robinson, C. V. (2010) Mass spectrometry defines the stoichiometry of ribosomal stalk complexes across the phylogenetic tree. *Mol Cell Proteomics* **9**, 1774-1783
141. Mohr, D., Wintermeyer, W., and Rodnina, M. V. (2002) GTPase activation of elongation factors Tu and G on the ribosome. *Biochemistry-Us* **41**, 12520-12528
142. Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J. M., Tonevitsky, A. G., Stark, H., Rodnina, M. V., and Wahl, M. C. (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**, 991-1004
143. Lockwood, A. H., Maitra, U., Brot, N., and Weissbac.H. (1974) Role of Ribosomal-Proteins L7 and L12 in Polypeptide-Chain Initiation in Escherichia-Coli. *J Biol Chem* **249**, 1213-1218
144. Kowalak, J. A., and Walsh, K. A. (1996) beta-methylthio-aspartic acid: Identification of a novel posttranslational modification in ribosomal protein S12 from Escherichia coli. *Protein Sci* **5**, 1625-1632
145. Anton, B. P., Saleh, L., Benner, J. S., Raleigh, E. A., Kasif, S., and Roberts, R. J. (2008) RimO, a MiaB-like enzyme, methylthiolates the universally conserved Asp88 residue of ribosomal protein S12 in Escherichia coli. *Proc Natl Acad Sci U S A* **105**, 1826-1831
146. Arragain, S., Garcia-Serres, R., Blondin, G., Douki, T., Clemancey, M., Latour, J. M., Forouhar, F., Neely, H., Montelione, G. T., Hunt, J. F., Mulliez, E., Fontecave, M., and Atta, M. (2010) Structural and functional characterization of RimO from Thermotoga maritima, a radical S-adenosylmethionine methylthiotransferase. *J Biol Chem* **285**, 5792-5801
147. Cukras, A. R., Southworth, D. R., Brunelle, J. L., Culver, G. M., and Green, R. (2003) Ribosomal proteins S12 and S13 function as control elements for translocation of the mRNA : tRNA complex. *Mol Cell* **12**, 321-328
148. Sharma, D., Cukras, A. R., Rogers, E. J., Southworth, D. R., and Green, R. (2007) Mutational analysis of S12 protein and implications for the accuracy of decoding by the ribosome. *J Mol Biol* **374**, 1065-1076
149. Mikulik, K., Bobek, J., Zikova, A., Smetakova, M., and Bezouskova, S. (2011) Phosphorylation of ribosomal proteins influences subunit association and translation of poly (U) in Streptomyces coelicolor. *Mol Biosyst* **7**, 817-823
150. Macek, B., Gnad, F., Soufi, B., Kumar, C., Olsen, J. V., Mijakovic, I., and Mann, M. (2008) Phosphoproteome analysis of E. coli reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Mol Cell Proteomics* **7**, 299-307
151. Liu, X., and Reilly, J. P. (2009) Correlating the chemical modification of Escherichia coli ribosomal proteins with crystal structure data. *J Proteome Res* **8**, 4466-4478
152. David, C. L., Keener, J., and Aswad, D. W. (1999) Isoaspartate in ribosomal protein S11 of Escherichia coli. *J Bacteriol* **181**, 2872-2877
153. L'Italien, J. J., and Laursen, R. A. (1979) Location of the site of methylation in elongation factor Tu. *FEBS Lett* **107**, 359-362
154. Young, C. C., and Bernlohr, R. W. (1991) Elongation factor Tu is methylated in response to nutrient deprivation in Escherichia coli. *J Bacteriol* **173**, 3096-3100

155. Van Noort, J. M., Kraal, B., Sinjorgo, K. M., Persoon, N. L., Johanns, E. S., and Bosch, L. (1986) Methylation in vivo of elongation factor EF-Tu at lysine-56 decreases the rate of tRNA-dependent GTP hydrolysis. *Eur J Biochem* **160**, 557-561
156. Barbier, M., Owings, J. P., Martinez-Ramos, I., Damron, F. H., Gomila, R., Blazquez, J., Goldberg, J. B., and Alberti, S. (2013) Lysine trimethylation of EF-Tu mimics platelet-activating factor to initiate Pseudomonas aeruginosa pneumonia. *MBio* **4**, e00207-00213
157. Trobro, S., and Aqvist, J. (2007) A model for how ribosomal release factors induce peptidyl-tRNA cleavage in termination of protein synthesis. *Mol Cell* **27**, 758-766
158. Mora, L., Heurgue-Hamard, V., de Zamaroczy, M., Kervestin, S., and Buckingham, R. H. (2007) Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. *J Biol Chem* **282**, 35638-35645
159. Glick, B. R., and Ganoza, M. C. (1975) Identification of a soluble protein that stimulates peptide bond synthesis. *Proc Natl Acad Sci U S A* **72**, 4257-4260
160. Ganoza, M. C., and Aoki, H. (2000) Peptide bond synthesis: function of the efp gene product. *Biol Chem* **381**, 553-559
161. Park, M. H., Wolff, E. C., and Folk, J. E. (1993) Hypusine: its post-translational formation in eukaryotic initiation factor 5A and its potential role in cellular regulation. *Biofactors* **4**, 95-104
162. Park, M. H., Nishimura, K., Zanelli, C. F., and Valentini, S. R. (2010) Functional significance of eIF5A and its hypusine modification in eukaryotes. *Amino Acids* **38**, 491-500
163. Peil, L., Starosta, A. L., Virumae, K., Atkinson, G. C., Tenson, T., Remme, J., and Wilson, D. N. (2012) Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. *Nat Chem Biol* **8**, 695-697
164. Navarre, W. W., Zou, S. B., Roy, H., Xie, J. L., Savchenko, A., Singer, A., Edvokimova, E., Prost, L. R., Kumar, R., Ibba, M., and Fang, F. C. (2010) PoxA, yjeK, and elongation factor P coordinately modulate virulence and drug resistance in Salmonella enterica. *Mol Cell* **39**, 209-221
165. Ude, S., Lassak, J., Starosta, A. L., Kraxenberger, T., Wilson, D. N., and Jung, K. (2013) Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82-85
166. Blaha, G., Stanley, R. E., and Steitz, T. A. (2009) Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. *Science* **325**, 966-970

2. Mass spectrometric and functional studies on *ycfD*

Contents

2.1. Prior mass spectrometric approaches to studying ribosomal proteins 54

2.1.1. Aims 55

2.2. Development LC-MS method for the study of ribosomal proteins 55

2.3. *In vivo* hydroxylation of L16 is *ycfD*-dependent 66

2.4. Effect of O_2 and environmental conditions on ribosomal protein modifications 71

2.5. *YcfD* as a putative oxygen sensor 77

2.6. Phenotypic studies on the effect of *ycfD* deletion 78

2.6.1. *YcfD* deletion affects growth rate at 20°C and in minimal media 79

2.6.2. *YcfD* deletion affects growth in minimal media 80

2.6.3. *YcfD* does not affect ribosomal assembly 81

2.6.4. *YcfD* does not affect translational accuracy 83

2.6.5. *YcfD* affects global translation rate 84

2.6.6. *YcfD* deletion decreases growth rate on single carbon sources short-chain carboxylic acids, TCA cycle intermediate analogues and amino acids 86

2.7. YcfD and TCA cycle metabolites 89

2.8. Conclusions 92

2.9. References 93

2.1. Prior mass spectrometric approaches to studying ribosomal proteins

Previous approaches to the spectrometric study of ribosomal proteins included: (i) a Matrix-Assisted Laser Desorption Ionisation-Mass spectrometry (MALDI-MS) without chromatographic separation of proteins (1); and (ii) an LC-MS investigations consisting of a separation and mass spectrometry step (2,3). Chromatographic separation of proteins can enable rapid identification of proteins based on retention time and mass, whereas the MALDI-MS approaches do not enable such characterisation; furthermore, the MALDI-MS signal is dependent on the crystallisation matrix used and co-solvent (1) and may require repeated optimisations of conditions for individual samples or at least individual species. MALDI-MS was used to determine masses of ribosomal proteins to one decimal place, although some masses deviated from theoretical predictions (based on primary sequences, and allowing for post-translational modifications) by ±2.9 Da (1).

LC-MS approaches used previously included a 2-dimensional liquid chromatographic (2D-LC) system: a strong cation exchange (SCX) column was connected to a C4 reverse-phase (RP) column, which was connected to a Micro Q-Tof mass spectrometer. This approach achieved a mass error of < ±0.5 Da and > 95% protein coverage (3). An alternative strategy involved separating the samples on a 2D-LC system including an array of 20 C4 reverse phase 'trap' columns; the contents of the trap columns were individually eluted onto another C4 reverse-phase column connected to a quadrupole time-of-flight (QTOF) mass spectrometer containing an ESI source (4). This setup allowed a proteomic characterisation of proteins analysed by whole-protein MS.

Another study utilised an electrospray ionisation Fourier transform ion cyclotron resonance mass spectrometer (ESI-FT-ICR-MS) connected to a reverse-phase C4 column. The FT-ICR-MS spectrometer achieves broad-band mass resolution of 50,000, enabling isotopic resolution of ribosomal protein masses to 0.025 Da accuracy. The method nonetheless proved unable to achieve complete coverage and a number of mass assignments were spurious (5).

2.1.1. Aims

The finding that the 2OG oxygenase ycfD catalyses the hydroxylation of a peptide fragment of the *E. coli* ribosomal protein L16 raised questions about the roles of post-translational modifications to the ribosome. It was unknown whether oxygen-dependent modifications occur on the ribosome in cell and whether they vary with changes in oxygen concentrations or in response to other external stimuli. Furthermore, the effect of the 2OG oxygenase ycfD on ribosomal proteins has not been investigated; it was unclear whether the observation of its activity *in vitro* also applies in cells and whether L16 is the only ribosomal protein substrate of ycfD.

This chapter describes the work on applying an LC-MS method for the study of ribosomal protein modifications in *E. coli* cells subjected to different external conditions. Initially, work characterising *ycfD* as a 2OG-dependent ribosomal protein hydroxylases is described; followed by work on the effects of atmospheric oxygen on ribosomal protein modifications. The phenotypic effects of *ycfD* knock-out on *E. coli* , including growth rate in rich and minimal media, growth at different temperatures, as well as the effects on global translation rate and on translation accuracy, were assessed using a variety of approaches.

2.2. Development LC-MS method for the study of ribosomal proteins

In order to achieve a balance between mass accuracy and adequate protein coverage, with the constraint of available in-house technology, it was decided to pursue the goal of MS studies of ribosomal proteins using 1D-LC separation coupled to a time-of-flight mass spectrometer, which is capable of achieving adequate mass accuracy (< 0.5 Da). The MALDI-MS approach suffers from its inability to separate the proteins, as well as lower mass accuracy; the FT-ICR-MS has superior mass accuracy but is technically very sophisticated to enable sufficient throughput to study a multitude of conditions.

In order to study the effects of ycfD on ribosomal proteins, a robust protocol for purification of ribosomal proteins was first established (see work flow-chart Figure 2.1). Ribosomes from *E. coli* WT strain

BW2113 were isolated by ultracentrifugation through a sucrose cushion and ribosomal proteins extracted using a modified method from Hardy *et al.* (6). The extracted proteins were initially visualised on a SDS-PAGE gel (Figure 2.1), which demonstrated a relative enrichment of proteins at 10-35 kDa, consistent with masses of 70S ribosomal proteins, two intense bands between 35-40 kDa (mass of the ribosome-associated factor Ef-Tu is ~43 kDa), and three intense bands above 70 kDa likely corresponding to ribosome-associated factors (the mass of Ef-G is ~77 kDa).

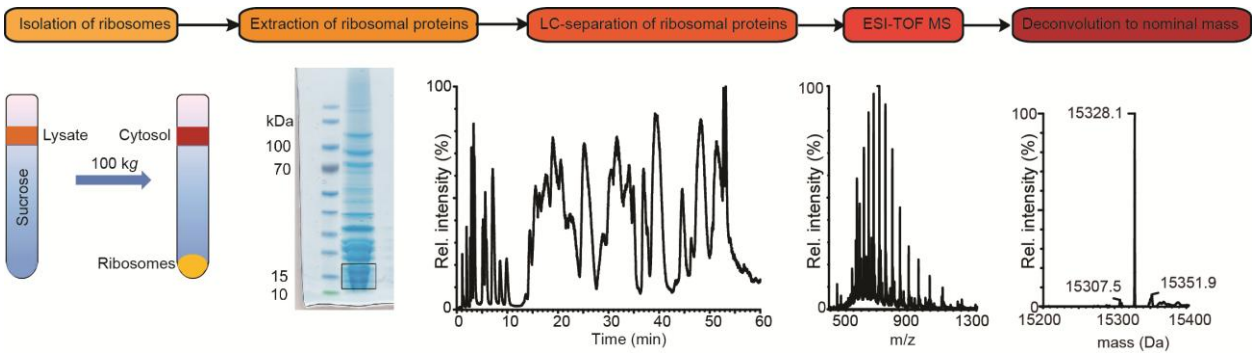


Figure 2.1: Proposed work-flow for the mass-spectrometric analysis of ribosomal protein masses. Ribosomes are isolated from cell lysates by ultracentrifugation; extraction of ribosomal proteins by the acetic acid/acetone precipitation method of Hardy *et al*. (6); extracted proteins can be visualised on an SDS-PAGE gel. High enrichment of proteins at 10-20 kDa is highlighted, which likely corresponds to 70S ribosomal proteins; liquid-chromatographic separation of ribosomal proteins using H_2O/acetonitrile solvent system with an appropriate co-solvent; MS analysis of eluted peaks reveals m/z envelopes corresponding to individual proteins/mixtures of proteins eluting at identical elution time; m/z envelopes are deconvoluted using automated deconvolution algorithm (MaxEnt1) or by manual peak selection, to reveal the intact mass of an eluted protein. The mass can then be compared to the predicted mass based on the primary protein sequence deduced from the genomic sequence.

The strategy employed to study ribosomal protein masses was based on previous reports of mass spectrometry studies on ribosomal proteins. The water/acetonitrile solvent system was chosen on the basis of literature review which suggested that this was the solvent system of choice for the separation of ribosomal proteins (7-12). An elution gradient was initially selected to reproduce a published procedure (9) (Figure 2.2) in which an unprecedented degree of separation of 70S ribosomal proteins was achieved by RP-HPLC C_{18}-silica column (6.5 μm particle size, pore size 300 Å, 4.1 × 250 mm), without the need for separation of the small (30S) and the large (50S) ribosomal subunits; HPLC peaks were identified using *N*-terminal sequencing. The reported chromatographic method was adapted for an analytical Vydac 214TP C_4 reverse-phase column (5 μm particle size, 4.6 × 250 mm, pore size 300 Å, Grace); C4 column

chemistry was chosen over C18 due to the superior protein separation achieved with a shorter alkyl chain. Initially, protein separation was monitored by UV-detection at 214 nm (Agilent 1260 Infinity).

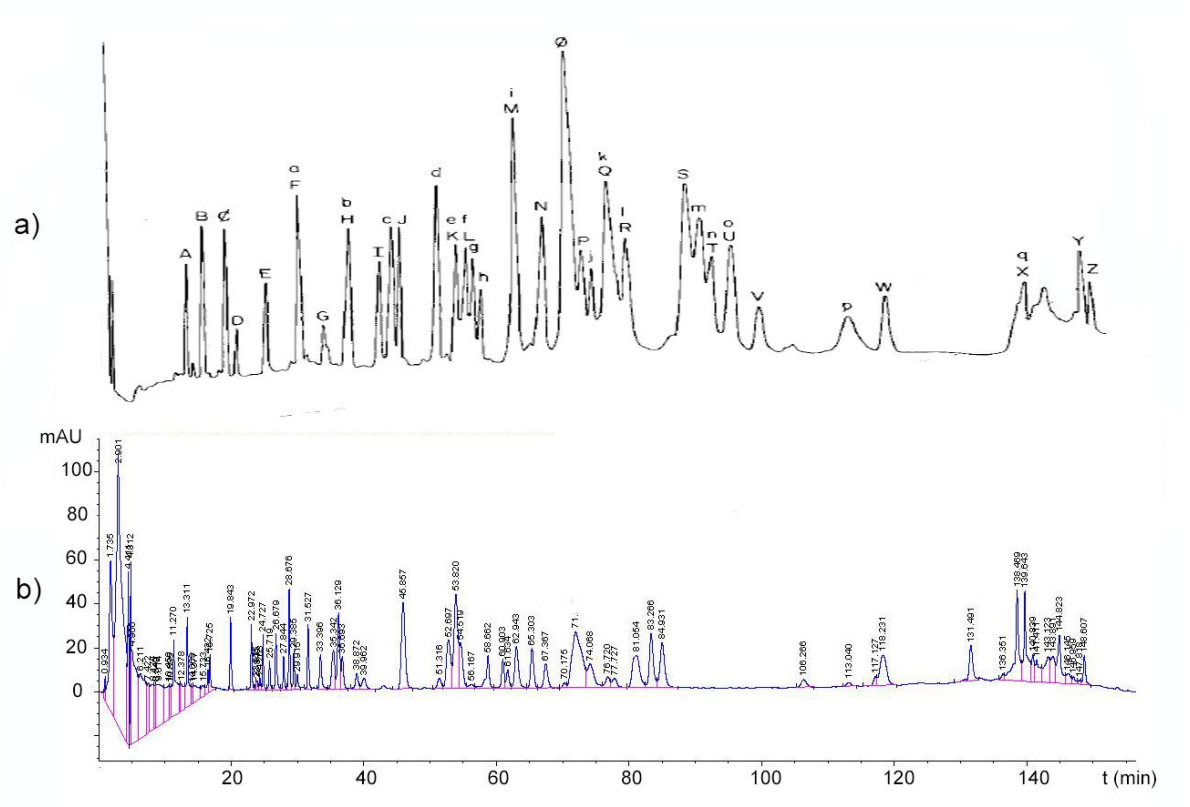


Figure 2.2: Comparison of HPLC traces of 70S ribosomal proteins between the published and the developed procedure. RP-HPLC separation of 70S ribosomal proteins using H_2O/acetonitrile system with CF_3COOH_{aq} as the co-solvent was based on the published method of Kerlevage et al. (9), which used a C_{18}-silica column (6.5 µm particle size, pore size 300Å, 4.1 × 250 mm). The method was adapted for the 214TP C_4 reverse-phase column (5 µm particle size, 4.6 × 250 mm, pore size 300 Å, Grace, b). A solution of *E. coli* ribosomal proteins (1 mg/mL in 0.1% CF_3COOH_{aq}) was injected onto the RP-HPLC column using an automated sampling device; flow-rate was 0.7 mL/min; elution gradient is detailed in Materials and Methods; protein peaks were detected by monitoring absorbance at 214 nm using the Agilent 1260 Infinity UV detector.

The chromatographic profile of *E. coli* BW25113 ribosomal proteins separated with this method showed high similarity to the published chromatogram (Figure 2.2a). The chromatograms differed in the region of elution times 0-11 min, where 3 peaks (a peak at 2 min followed by two peaks of lower intensity at 4.3-4.8 min) were observed using the C4 column, whereas no peaks were observed in the literature report. Two peaks eluted at 11.3 and 13.3 min, likely corresponding to peaks A and B from Figure 2.2. Twin peaks of low intensity eluting at 16 min co-eluted with peak C and peak at 19.8 min could correspond to peak D. 32 peaks were observed between 22-85 min, whilst only 23 were observed in the published

procedure, of which peak O had significantly higher intensity compared to other peaks, suggesting simultaneous elution of multiple proteins; given the high similarity of elution gradients, this confirmed that C4 chemistry may be better suited for resolving protein peaks. The chromatogram showed closer resemblance after elution time of 106 min. Notably, peaks at 118 min likely correspond to peaks P and W, the peak at 131 min may correspond to peaks Q and X and the sharp, intense peaks at 140 min likely correspond to peaks Y and Z.

Although a good separation of 70S proteins was achieved, the usefulness of the method was limited due to the use of the co-solvent CF_3COOH_{aq}, which improves HPLC separation; however, CF_3COOH_{aq} is relatively incompatible with mass spectrometry analysis due to the non-volatility of CF_3COOH_{aq}. Applying the same elution gradient concomitant with replacement of CF_3COOH_{aq} with $HCOOH_{aq}$ as the co-solvent resulted in severe peak broadening: only 19 peaks were resolved and none were observed after 40 min (data not shown). In an attempt to achieve better separation, the steepness of the elution gradient was decreased, which resulted in separation of 24 major peaks (Figure 2.3).

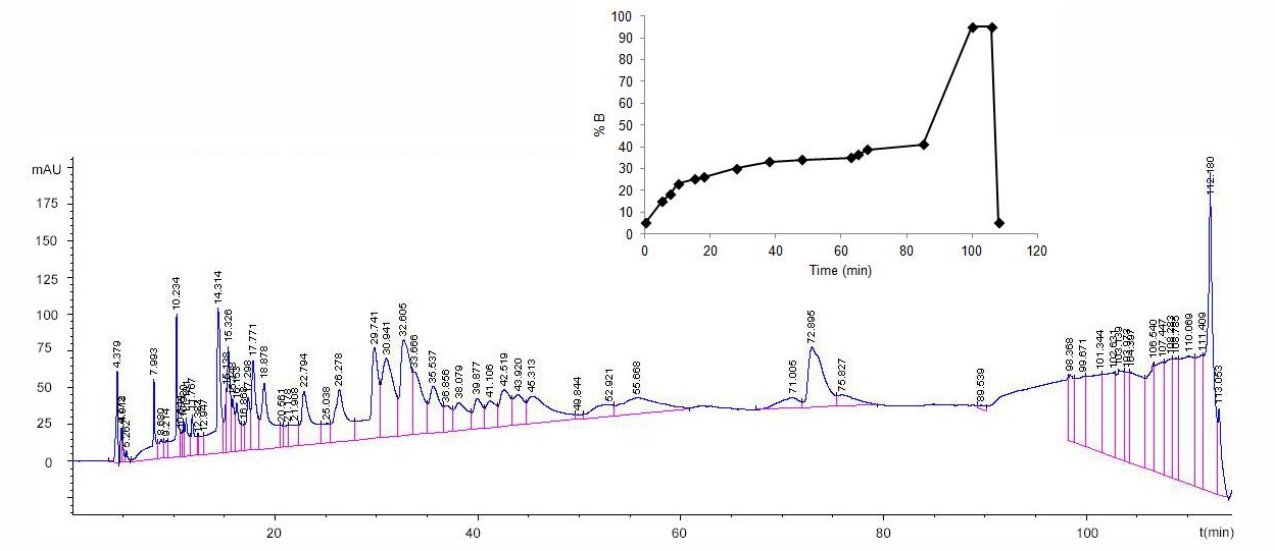


Figure 2.3: UV trace of the *E. coli* 70S ribosomal proteins separated using a Vydac 214 TP RP-HPLC C_4 column (5 µm particle size, 4.6 × 250 mm, pore size 300 Å, Grace) in a H_2O/acetonitrile solvent system with $HCOOH_{aq}$ (0.1%) used as a co-solvent. The elution gradient used in Figure 2.2 resulted in severe peak broadening and loss of signal; the steepness of the elution gradient was decreased (inset), which resulted in the separation of 24 major peaks. A solution of *E. coli* ribosomal proteins (1 mg/mL in 0.1% CF_3COOH_{aq}) was injected onto the RP-HPLC column using an automated sampling device; flow-rate was 0.7 mL/min; elution gradient is detailed in Materials and Methods; protein peaks were detected by monitoring absorbance at 214 nm using the Agilent 1260 Infinity UV detector.

This chromatographic method was then adapted for the RP-HPLC-Quad-MS ZMD mass analyzer. Mass-signal based detection resulted in detection of 25 distinct MS peaks (Figure 2.4). Since three proteins eluted during the first 20 min of the elution method, the steepness of the gradient was increased to reach 18% B within 2.5 min, the gradient was then reduced to span 18-35% B over 50 min. 28 peaks were thus resolved in total, corresponding to an average elution of two ribosomal proteins per peak.

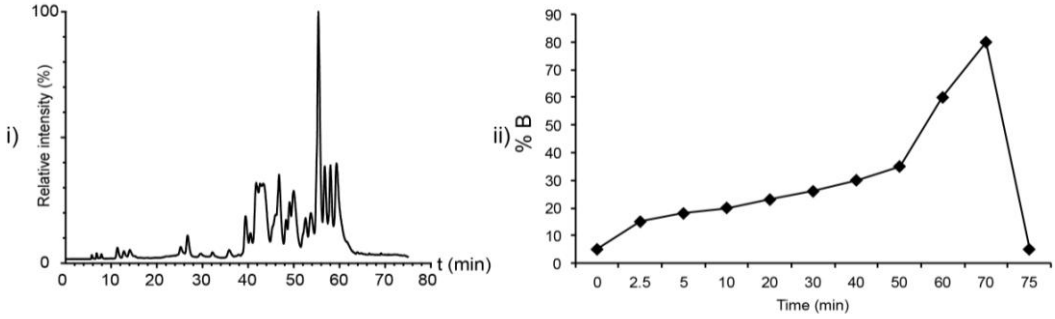


Figure 2.4: RP-HPLC-Quad-MS trace of the separation of *E. coli* 70S ribosomal proteins. Acquired with a Vydac 214 TP RP-HPLC C_4 column (5 µm particle size, 4.6 × 250 mm, pore size 300 Å, Grace) using the elution gradient shown in panel ii). Chromatographic separation of proteins was used for a preliminary mass-based identification of ribosomal proteins.

Mass-based identification of ribosomal proteins was attempted, with analysis of individual peaks by MaxEnt1 deconvolution of charged state envelopes. The calculated masses were compared to the theoretical predictions based on primary protein sequences, derived from the Uniprot database (www.uniprot.org); protein identities were assigned based on similarity to theoretical predictions. The following possible modifications, and combinations thereof, were taken into account: *N*-terminal methionine (NTM) cleavage (-131 Da), methylation (+14 Da), acetylation (+42 Da), hydroxylation (+16 Da), and β-methylthiolation (+46 Da). Assignments were further validated by comparison with previous MS analyses which was performed on a related strain (K-12) of *E. coli* (1). Within experimental error (± 2 Da), masses of 39 proteins were consistent with theoretical prediction of ribosomal protein masses, with PTMs taken into account. However, no MS signals were identified with masses corresponding to ribosomal proteins S1, S2, S5-7, S9-11, S14, L7/l12, L8, L9, L18, L29, L31 and L35. It was envisaged that these proteins were either lost from the ribosomal complex during the purification procedure, their concentration was below the limit of detection, or their ionization properties were incompatible with quadrupole mass spectrometry. Protein assignments were ambiguous due to a large

mass error, e.g. the mass of S3 was 6 Da lower than expected based on sequence mass after NTM removal, the mass of L3 was 8 Da higher than expected. In accordance with previous reports (1,13,14), mass differences of S12 and S18 (47 and 42 Da, respectively) suggested β-methylthiolation and acetylation, respectively. The masses of L16 and L33 were 46 and 15 Da larger than expected (15327 Da and 6256 Da. The mass of L16 was previously determined at 15327 Da (1); the predicted mass of L16 is 15281 Da; within experimental error, the difference of 46 Da could correspond to a methylation and two oxidative modifications or two methylations and an oxidative modification. N-terminal methylation and oxidative modification at R81 have been reported before (15). The mass of L33 15 Da larger than expected; within experimental error it could correspond to either methylation (14 Da) or hydroxylation (16 Da); methylation has been reported (1).

Table 2.1: Assignment of HPLC chromatographic peaks of *E. coli* BW25113 30S ribosomal proteins from quadrupole-MS deconvolution. MaxEnt1 deconvolution was used to determine nominal masses of 70S ribosomal proteins at 1 Da resolution.

Protein	Sequence mass (Da)	Experimental mass (Da)	Δ (Da)	Methionine lost?	Δ NML (Da)	PTM	Elution time (min)
S1	61158	n.d.	n.a.	n.a.	n.a.		n.a.
S2	26744	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S3	25983	25846	-137	Yes	-6		48.5
S4	23469	23339	-130	Yes	1		46.2
S5	17603	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S6	15704	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S7	20019	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S8	14127	13992	-135	Yes	-4		52.8
S9	14856	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S10	11736	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S11	13845	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S12	13737	13653	-84	Yes	47	modification	12.8
S13	13099	12968	-131	Yes	0		52.8
S14	11580	11449	-131	Yes	0		26.1
S15	10269	10137	-131	Yes	0		40.7
S16	9191	9193	2	No			48.5
S17	9704	9572	-132	Yes	-1		39.8
S18	8986	8897	-89	Yes	42	modification	38.1
S19	10430	10299	-131	Yes	0		32.4
S20	9684	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S21	8500	8369	-131	Yes	1		15.0
S22	5095.8	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.

Table 2.2: Assignment of HPLC chromatographic peaks of *E. coli* BW25113 50S ribosomal proteins from quadrupole-MS deconvolution. MaxEnt1 deconvolution was used to determine nominal masses of 70S ribosomal proteins at 1 Da resolution.

Protein	Sequence mass (Da)	Experimental mass (Da)	Δ (Da)	Methionine lost?	Δ NML (Da)	PTM	Elution time (min)
L1	24730	24600	-130	Yes	1		55.4
L2	29860	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L3	22244	22252	8	No		modification	46.9
L4	22087	22086	-1	No			59.6
L5	20302	20172	-130	Yes	1		55.4
L6	18904	18773	-131	Yes	0		55.4
L7	12295	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L9	15769	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L10	17712	17579	-133		-2		58.0
L11	14875	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L12	12295	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L13	16019	16016	-3	No			42.1
L14	13541	13541	0	No			42.1
L15	14980	14980	0	No			55.4
L16	15281	15327	46	No		modification	54.2
L17	14365	14364	-1	No			43.6
L18	12770	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L19	13133	13002	-131	Yes	0		43.6
L20	13497	13365	-132	Yes	-1		56.8
L21	11564	11565	1	No			46.2
L22	12226	12225	-1	No			50.4
L23	11199	11199	0	No			50.4
L24	11316	11184	-132	Yes	-1.0		26.9
L25	10693	10693	0	No			39.8
L26/S20	9684	9553	-131	Yes	0		29.6
L27	9124	8992	-132	Yes	-1		11.3
L28	9006	8875	-131	Yes	0		25.3
L29	7273	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L30	6542	6411	-131	Yes	0		36.2
L31	7871	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L32	6446	6314	-132	Yes	-1		3.6
L33	6372	6256	-116	Yes	15	modification	7.7
L34	5380	5380	0	No			3.6
L35	7289	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L36	4364	4365	1	No			7.4

The 70% coverage (38/55 ribosomal proteins) achieved with the HPLC-Quad-MS system was deemed inadequate for the purpose of studying the effect of ycfD on ribosomal proteins. In particular, the mass of L9, which was observed to interact with ycfD (see Chapter 1), could not be detected.

In order to simultaneously address the problems of poor separation and sensitivity, the method was therefore adapted for an ultra-performance liquid chromatography (UPLC) system interfaced with a time-of-flight (TOF) mass spectrometer with an electrospray ionization (ESI) source. ESI-TOF-MS has superior sensitivity to quadrupole-MS and has the ability to resolve complex mixtures of proteins, as demonstrated on a study of the soluble proteome of chick skeletal muscle, in which a number of highly expressed proteins could be identified based on whole-protein mass and post-translational modifications assigned based on comparison with predicted masses (16). UPLC chromatography is superior to HPLC due to smaller particle size, which leads to better separation (resolution is inversely proportional to the square root of particle size), therefore replacing the HPLC column (5 μm particle size) with a UPLC column (1.7 μm particle size) should afford ~1.7× better resolution. During the initial attempt, which used the same elution profile at reduced flow rate (0.3 mL/min), elution of >20 proteins was observed during the first 20 min of the method (between 26-34% B, Figure 2.5A), suggesting that the increase in %B was too rapid to enable separation of ribosomal proteins. The method was altered so as to reach 25% B over 50 min, followed by a rapid increase to 60% B, followed by a washing step at 95% B (Figure 2.5B). Most proteins eluted after 60 min, during the washing step, suggesting that 25% B was insufficient for the elution of ribosomal proteins. Finally, the method was adjusted so as to rapidly reach 15% B (over 2.5 min), followed by a slow gradient of 15-35% B over 52.5 min. Gratifyingly, this method achieved sequential elution of protein peaks over the entire length of the chromatographic method (Figure 2.5C). Despite peak-broadening, due to slower elution of individual proteins over the shallow gradient, separation was achieved to a maximal co-elution of three proteins.

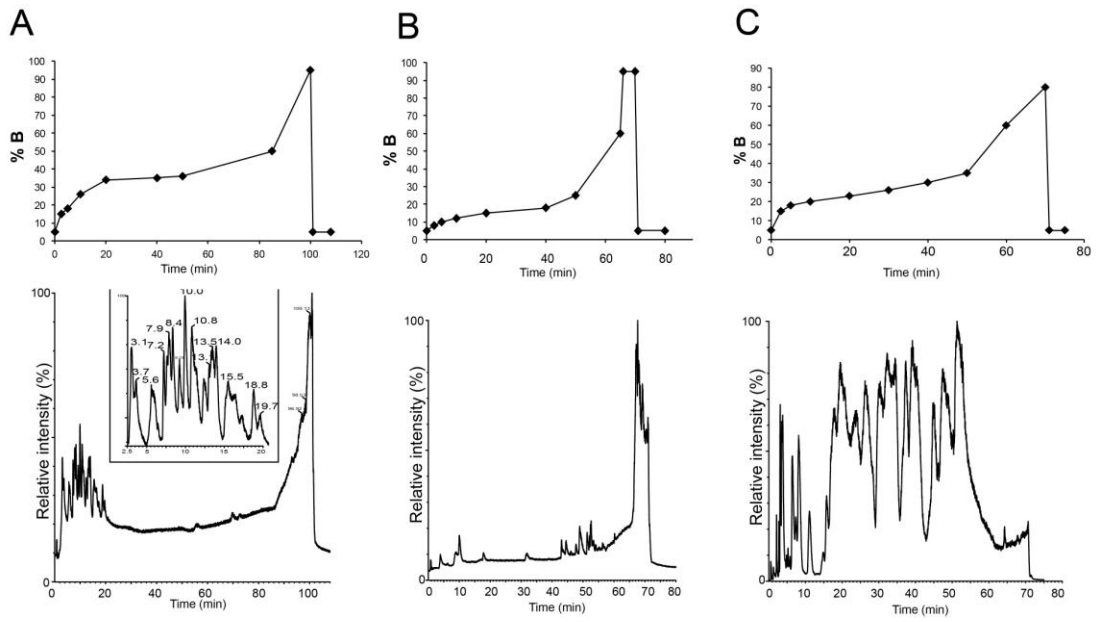


Figure 2.5: Development of a UPLC chromatographic elution method using H_2O/acetonitrile with $HCOOH_{aq}$ as a co-solvent for the separation of *E. coli* BW25113 70S ribosomal proteins. Method A used a steep gradient 18-34% B over 15 min resulting in elution of proteins during the first 20 min of the method; method B used a shallow gradient 5-25% B over 50 min, leading to the elution of the majority of proteins during the washing step; method C combined the findings to use a fast initial increase to 15% B followed by a shallow gradient to 35% B over 52.5 min.

Mass-based identification of 52 out of 55 ribosomal proteins (95% coverage) was achieved using this separation (Table 2.3-4). The elution of short basic proteins (L32-L36) was observed in the early part of the chromatogram (1-3 min), whereas acidic proteins (L7/L12) eluted last (52 min). Consistent with previous reports (1,3) the masses of all proteins apart from S16, L3, L4, L9, L14-L18, L21-L23, L25, L29, L34, L36 were consistent with NTM cleavage (-131 Da). Masses of S5, S18 and L7 were -88 Da from predicted mass, consistent with NTM cleavage (-131 Da) and acetylation (+42 Da). S11 and L33 were modified by -116 Da, consistent with NTM cleavage and single methylation; L3 was modified by +15 Da, consistent with methylation. The mass of S12 was -85 Da from predicted mass, consistent with NTM cleavage and β-methylthiolation (+46 Da). The mass of L11 was -5 Da from predicted value; this mass is consistent with NTM cleavage and nona-methylation (126 Da). The mass of L16 was 46 Da from predicted value. The mass was consistent with a methylation (14 Da) and two hydroxylations (32 Da), but may also correspond to two methylations and a hydroxylation (+44 Da). All of the observed modifications were consistent with previous studies (1,3).

Table 2.3: Assignment of HPLC chromatographic peaks of *E. coli* BW25113 30S ribosomal proteins from ESI-TOF-MS deconvolution. MaxEnt1 deconvolution (0.1 Da) was used to determine the masses. Sodium formate was used as the external calibrant.

Protein	Sequence mass (Da)	Experimental mass (Da)	Elution time (min)	Δ (Da)	Methionine lost?	Δ NML (Da)	PTM
S1	61158.0	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S2	26743.6	26613.8	44.6	-129.8	Yes	1.4	
S3	25983.2	25853.4	28.8	-129.8	Yes	1.4	
S4	23469.0	23339.0	22.4	-130.0	Yes	1.2	
S5	17603.3	17515.1	34.0	-88.2	Yes	43	Ac
S6	15703.5	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
S7	20019.0	19888.7	33.5	-130.3	Yes	0.9	
S8	14126.5	13996.0	34.9	-130.5	Yes	0.7	
S9	14856.2	14725.7	28.8	-130.5	Yes	0.7	
S10	11735.5	n.d.	n.a.	n.a.	n.a.	n.a.	
S11	13844.9	13729.2	17.5	-115.7	Yes	15.5	+Me
S12	13737.0	13652.5	3.5	-84.5	Yes	46.7	β-MeSH
S13	13099.3	12968.6	31.1	-130.7	Yes	0.5	
S14	11580.4	11450.2	5.9	-130.2	Yes	1	
S15	10268.7	10138.1	15.6	-130.6	Yes	0.6	
S16	9190.5	9190.8	22.4	0.3	No	0.3	
S17	9704.4	9573.6	16.3	-130.8	Yes	0.4	
S18	8986.4	8898.1	13.6	-88.3	Yes	42.9	Ac
S19	10430.2	10300.0	9.9	-130.2	Yes	1	
S20	9684.3	9553.9	5.2	-130.4	Yes	0.2	
S21	8499.9	8369.1	3.3	-130.8	Yes	0.4	
S22	5095.8	n.d.	n.a.	n.a.	n.a.	n.a.	

Table 2.4: Assignment of HPLC chromatographic peaks of *E. coli* BW25113 50S ribosomal proteins from ESI-TOF-MS deconvolution. MaxEnt1 deconvolution (0.1 Da) was used to determine the masses. Sodium formate was used as the external calibrant.

Protein	Sequence mass (Da)	Experimental mass (Da)	Elution time (min)	Δ (Da)	Methionine lost?	Δ NML (Da)	PTM
L1	24729.6	24599.6	39.2	-130.0	Yes	1.2	
L2	29860.4	29729.4	15.6	-131.0	Yes	0.2	
L3	22243.5	22258.5	25.2	15.0	No		Me
L4	22086.5	22087.5	48.2	1.0	No		
L5	20301.5	20171.2	39.4	-130.3	Yes	0.9	
L6	18903.7	18773.5	31.1	-130.2	Yes	1.0	
L7	12295.2	12205.9	53.2	-89.3	Yes	41.9	Ac, Me
L9	15769.0	15769.6	39.2	0.6	No		
L10	17711.5	17581.1	48.2	-130.4	Yes	0.8	
L11	14875.3	14870.8	37.2	-4.5	Yes	126.7	9 Me
L12	12295.2	12164.0	52.7	-131.2	Yes	0.0	Me
L13	16018.5	16019.5	19.0	1.0	No	1.0	
L14	13541.0	13541.6	20.7	0.6	No		
L15	14980.4	14981.1	31.7	0.7	No		
L16	15281.2	15328.1	30.5	46.9	No		Me, 2OH
L17	14364.5	14365.1	19.4	0.6	No		
L18	12769.6	12770.2	25.2	0.6	No		
L19	13133.2	13002.7	17.5	-130.5	Yes	0.7	
L20	13496.9	13366.4	33.5	-130.5	Yes	0.7	
L21	11564.3	11564.7	19.0	0.4	No		
L22	12226.2	12227.0	22.4	0.8	No		
L23	11199.1	11199.3	25.2	0.2	No		
L24	11316.2	11185.6	7.1	-130.6	Yes	0.6	
L25	10693.4	10694.2	14.5	0.8	No		
L26/S20	9684.3	9553.9	5.2	-130.4	Yes	0.2	
L27	9124.4	8993.7	3.3	-130.7	Yes	0.5	
L28	9006.4	8875.8	5.6	-130.6	Yes	0.6	
L29	7273.4	7273.5	28.8	0.1	No		
L30	6541.7	6411.0	8.6	-130.7	Yes	0.5	
L31	7871.0	n.d.	n.a.	n.a.	n.a.	n.a.	n.a.
L32	6446.3	6315.7	1.9	-130.6	Yes	0.6	
L33	6371.5	6254.9	2.6	-116.6	Yes	14.6	Me
L34	5380.3	5380.9	1.1	0.6	No		
L35	7288.9	7158.0	2.9	-130.9	Yes	0.3	
L36	4364.3	4363.6	1.9	-0.7	No		

2.3. *In vivo* hydroxylation of L16 is *ycfD*-dependent

Having established an LC-MS method with adequate coverage and mass accuracy to study modifications to *E. coli* ribosomal proteins, it was possible to address the function of ycfD with respect to catalysis of post-translational modifications to ribosomal proteins. The *ycfD* knock-out (KO) strain (JW1114), commercially available from the Keio collection was used; its parent strain BW25113 was used as a control (17).

Ribosomal proteins were initially purified and isolated from *ycfD* WT and KO strains, and studied by LC-MS (Figure 2.6). The UPLC chromatograms of *ycfD* WT and KO strains appeared identical (Figure 2.6A), suggesting that the composition of the 70S ribosome is unchanged between the two strains and that the variation in abundance of ribosomal proteins is minimal. Deconvolution (using MaxEnt1 to 0.1 Da accuracy) of each chromatographic peak from the two samples demonstrated that out of 52 ribosomal proteins analysed, only a single protein mass differed, corresponding to the ribosomal protein L16 (Table 2.5). In the WT strain, the protein was found at 15328.1 Da, whereas in ycfD KO strain the mass was 15312.1 Da, i.e. 16 Da lower, consistent with loss of a single hydroxyl group (Table 2.5, Figure 2.6C).

Masses of other proteins were identical between the two strain within ±0.4 Da. Importantly, the masses corresponding to proteins L9, L10, L13 and L22, identified as potential ycfD interactors (see Chapter 1), were identical within experimental error; masses of L9, L13 and L22 corresponded to nominal masses (observed at 15769.6 Da, 16019.4 Da and 12227.0 Da, respectively, predicted at 15769.0 Da, 16018.5 Da and 12226.2 Da), whereas the mass of L10 (17581.1, predicted (17711.5 Da) suggested NTM cleavage, but no other PTMs. It was therefore concluded that ycfD catalyses L16 hydroxylation in cells. Of the 52 observed ribosomal proteins, no protein apart from L16 is a hydroxylation substrate for ycfD.

Table 2.5: Deconvoluted masses of 70S ribosomal proteins from *ycfD* WT (BW25113) and KO (JW1114) strains determined by the UPLC-ESI-TOF-MS spectrometry. Protein masses were deconvoluted to 0.1 Da using MaxEnt1 automatic deconvolution software. Masses were confirmed by manual deconvolution. Out of 52 ribosomal proteins analysed, only a single protein mass differed, corresponding to the ribosomal protein L16 highlighted in red.

BW25113			JW1114			Δ (Da)
Elution time	Mass (Da)	Protein	Elution time	Mass (Da)	Protein	
1.1	5380.9	L34	1.1	5381	L34	-0.1
1.9	4363.6	L36	2.0	4363.2	L36	0.4
	6315.7	L32		6315.8	L32	-0.1
2.6	6254.9	L33	2.6	6255	L33	-0.1
2.9	7158	L35	2.9	7158	L35	0.0
3.3	8369.1	S21	3.3	8369.1	S21	0.0
	8993.7	L27		8993.7	L27	0.0
3.5	13652.5	S12	3.5	13652.5	S12	0.0
5.2	9553.9	L26/S20	5.2	9553.9	L26/S20	0.0
5.6	8875.8	L28	5.6	8875.8	L28	0.0
5.9	11450.2	S14	5.9	11450.2	S14	0.0
7.1	11185.6	L24	7.1	11185.6	L24	0.0
8.6	6411	L30	8.6	6410.8	L30	0.2
9.9	10300	S19	9.9	10300.2	S19	-0.2
13.6	8898.1	S18	13.7	8898	S18	0.1
14.5	10694.2	L25	14.5	10694.2	L25	0.0
15.6	10138.1	S15	15.6	10138.1	S15	0.0
	14865.4	L11		14865.5	L11	-0.1
	29729.4	L2		29730.6	L2	-1.2
16.3	9573.6	S17	16.3	9573.7	S17	-0.1
17.6	13002.7	L19	17.6	13002.7	L19	0.0
	13729.2	S11		13729.3	S11	-0.1
19.0	11564.7	L21	19.0	11564.6	L21	0.1
	14365.1	L17		14365.2	L17	-0.1
	16019.4	L13	19.0	16019.5	L13	+0.1
20.6	13541.6	L14	20.6	13541.6	L14	0.0
22.4	9190.8	S16	21.5	9191.1	S16	-0.3
	12227.0	L22		12227.0	L22	0.0
	23339	S4		23339.1	S4	-0.1
25.2	11199.3	L23	25.4	11199.5	L23	-0.2
	12770.2	L18		12770.3	L18	-0.1
	22258.5	L3		22258.5	L3	0.0
28.8	7273.5	L29	28.9	7272.9	L29	0.6
	14725.7	S9		14725.8	S9	-0.1
	25853.4	S3		25853.6	S3	-0.2
30.5	15328.1	L16	30.6	15312.1	L16	16.0
31.7	12968.6	S13	31.6	12968.6	S13	0.0
	14981.1	L15		14981.2	L15	-0.1
	18773.5	L6		18773.6	L6	-0.1
34.0	13366.4	L20	34.1	13366.4	L20	0.0
	17515.1	S5		17515.3	S5	-0.2
	19888.7	S7		19888.9	S7	-0.2
34.9	13669	S8	34.9	13669.1	S8	-0.1
36.9	14870.8	L11	36.9	14871	L11	-0.2
37.7	17586.0	L10	37.7	17586.2	L10	-0.2
39.2	15769.6	L9	39.5	15769.6	L9	0.0
	20171.2	L5		20171.4	L5	-0.2
	24599.6	L1		24599.7	L1	-0.1
44.6	26613.8	S21		26614.1	S21	-0.3
48.2	17581.1	L10	48.3	17581.2	L10	-0.1
	22087.5	L4		22087.5	L4	0.0
52.7	12164	L12	52.8	12164.3	L12	-0.3
53.2	12205.9	L7	53.2	12206.1	L7	-0.2

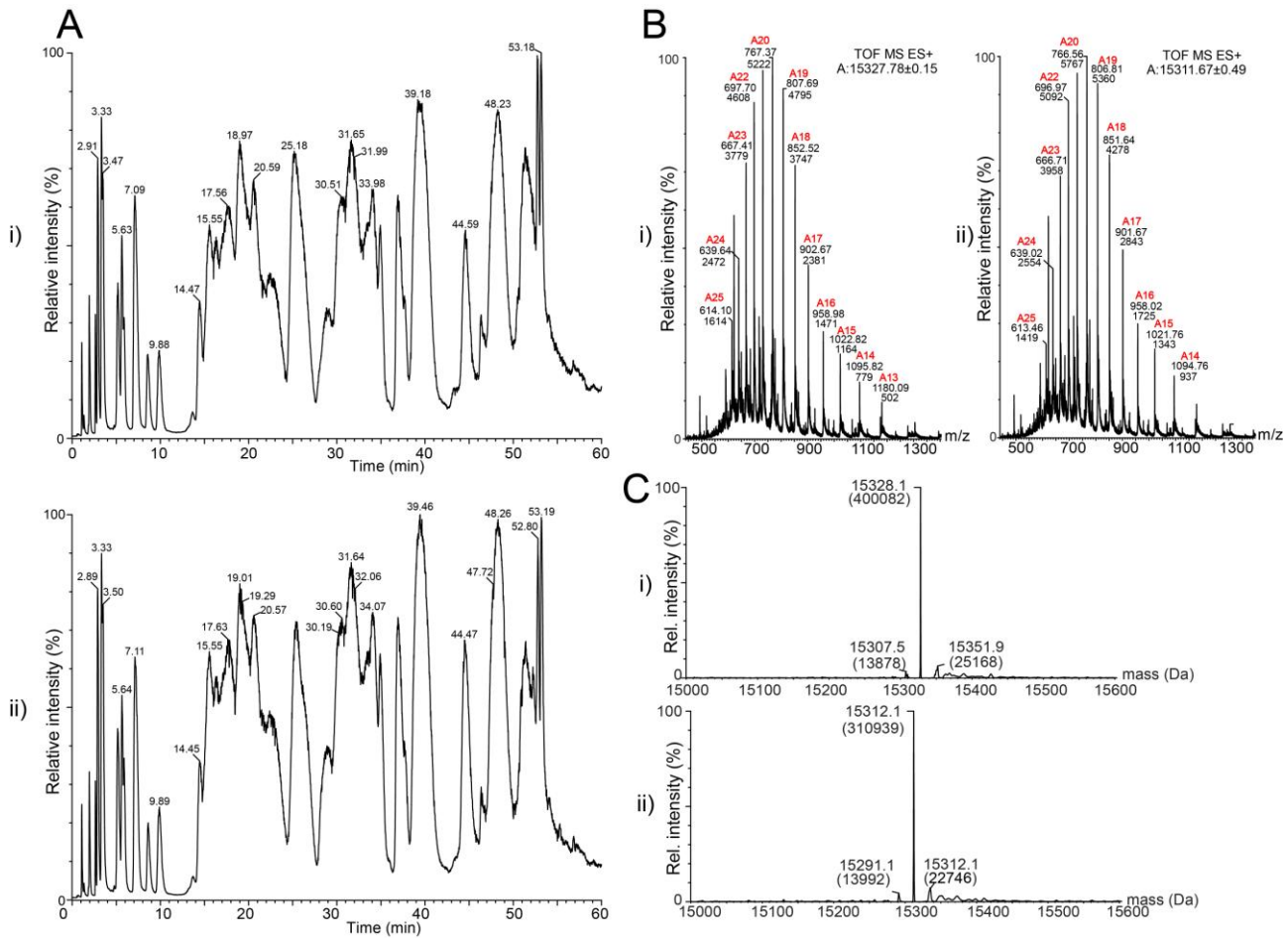


Figure 2.6: Comparison of ribosomal composition of BW25113 and JW1114 highlights different hydroxylation state of ribosomal protein L16. A) UPLC chromatograms of BW25113 and JW1114 reveal a high similarity in ribosomal protein composition. B) Charged state envelopes corresponding to the chromatographic peak eluting at 30.5±0.1 min. Manual deconvolution after subtraction, smoothing and centering of peaks reveals species of 15327.1±0.2 Da and 15311.7±0.5 Da respectively. C) Automatic MaxEnt1 deconvolution (to 0.1 Da) of L16 charged state envelopes demonstrates a mass difference of 16 Da for the protein eluting at 30.5±0.1 min between BW25113 and JW1114, corresponding to loss of a single hydroxyl group on ribosomal protein L16 for JW1114.

In order to study the effects of ycfD catalytic activity on L16 hydroxylation, ycfD WT and a facial triad variant H125A_D127A (predicted to lack catalytic activity) were ectopically over-expressed in JW1114 from the pRSet5D vector (18) containing an *N*-terminal GFP coding sequence (kindly provided by Dr. Alexander Wolf). Over-expression of the WT variant led to >95% recovery of L16 hydroxylation, whereas L16 in the strain over-expressing the catalytically inactive mutant was <5% hydroxlyated, consistent with the requirement of catalytic activity of ycfD for hydroxylation of L16 (Figure 2.7A).

The 70S ribosomal proteins from BW25113, as well as JW1114 and JW1114 expressing WT and a catalytically-inactive variant of ycfD, were resolved by SDS-PAGE, and bands at ~15±5 kDa excised and subjected to proteolytic digestion with trypsin. LC-MS analysis of trypsin-digested ribosomal proteins revealed that the mass variation observed at whole-protein level is also observed in the tryptic fragments, and is localised on the peptide harbouring R81 (^{67}VFPDKPITEKPLAVR^{81}, Figure 2.7B). The LC-MS analysis was performed by Dr Mukram Mackeen. It can be concluded that L16 hydroxylation depends on the catalytic activity of ycfD (Figure 2.7C).

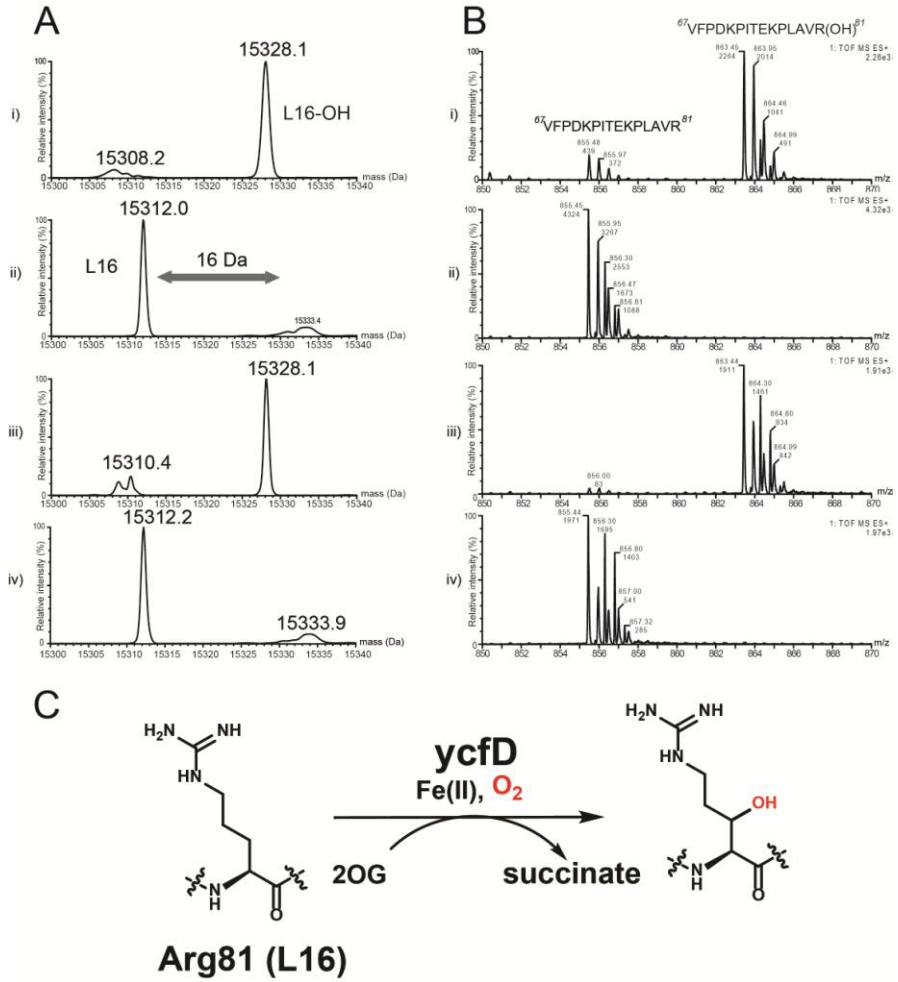


Figure 2.7: Deconvoluted ESI-MS spectra for deletion and rescue strains, confirm ycfD dependence of L16 hydroxylation A) Deconvoluted ESI-MS spectra of *E. coli* ribosomal protein L16 and corresponding tryptic fragments(^{67}VFPDKPITEKPLAVR^{81}): i) WT *E. coli* strain BW25113; ii) ycfD KO strain JW1114; iii) ectopically over-expressed WT ycfD in ycfD-null background (JW1114_WT); iv) ectopically over-expressed H125A_D127A ycfD in ycfD-null background (JW1114_AxA). Samples i) and iii) show a +16 Da shift compared to samples ii) and iv), suggesting the presence of a hydroxyl group on L16 in these two samples. B) ESI-MS spectra of L16 tryptic fragment $(M+2H)^{2+}$ charged state, corresponding to samples a)-d) showing variation of modification the 15-mer fragment containing R81. C) Detail of proposed ycfD mediated hydroxylation of R81 on L16.

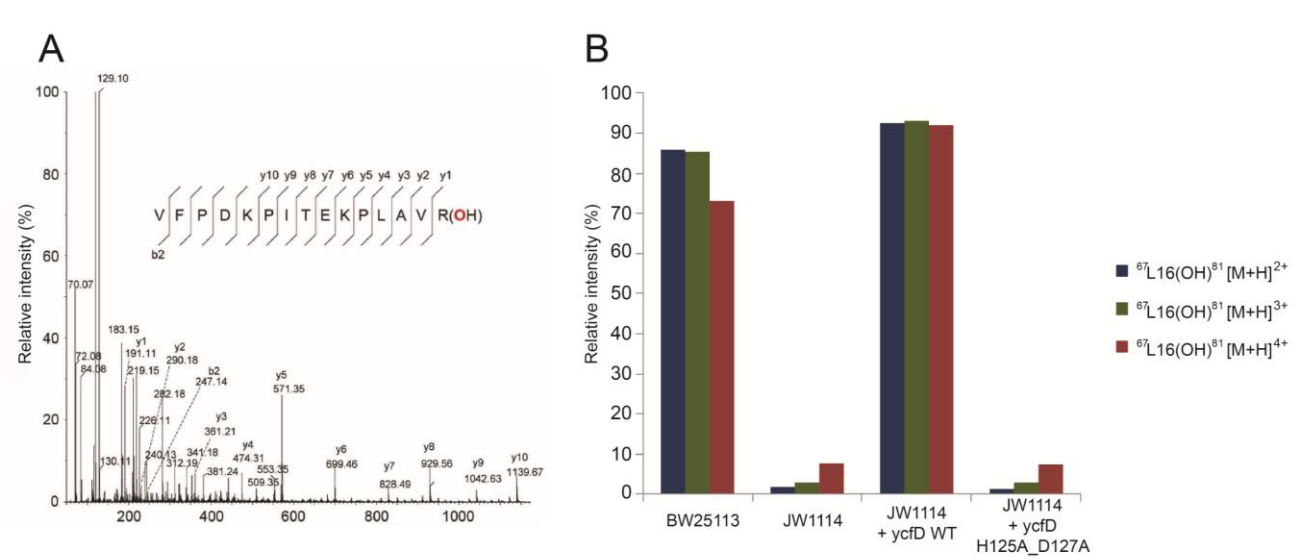


Figure 2.8: A) MS-MS analysis demonstrates the hydroxylation at R81 of ribosomal protein L16 in cells. B) Quantification of R81 hydroxylation in cells. Peak intensities of three charged states (+2, +3, +4) of the tryptic fragment VFPDKPITEKPLAVR(OH) were used to quantify the extent of L16 hydroxylation from a) WT *E. coli* strain BW25113; b) ycfD KO strain JW1114; c) JW1114_WT; d) JW1114_AxA. Samples a) and c) show 80% and 92% hydroxylation, respectively, whereas b) and d) show < 5% hydroxylation. The ESI-MS spectra of L16 tryptic fragment $(M+2H)^{2+}$ charged states were shown in Figure 2.7. The MS-MS spectrum was kindly provided by Dr Wei Ge; the MS-MS quantification was performed by Dr Mukram Mackeen.

Tandem mass spectrometry (MS/MS) analysis of the tryptic fragment with m/z 863.45 demonstrated that the hydroxylation occurs on R81 (Figure 2.8A). LC-MS analysis of tryptic peptides allowed quantification of the extent of hydroxylation, which was used for comparison with whole-protein data. Masses corresponding to three charged states (+2, +3, +4) of the tryptic fragment VFPDKPITEKPLAVR, and its hydroxylated form, were extracted from the LC chromatogram. In a comparison of peaks intensities, L16 in WT *E. coli* showed an average of ~80% hydroxylation, whereas in a *ycfD* KO strain the hydroxylation was reduced to ~5%. Ectopic over-expression of ycfD WT in the JW1114 strain resulted in ~90% hydroxylation of L16, whereas L16 in JW1114 with over-expression of ycfD H125A_D127A variant was only ~5% hydroxylated (Figure 2.8B).

Overall, the results presented above demonstrate that ycfD post-translationally hydroxylates L16 in cells, that it does not modify other ribosomal proteins, and that its function as a L16 hydroxylase is non-redundant and depends on the catalytic activity of ycfD.

2.4. Effect of O_2 and environmental conditions on ribosomal protein modifications

The motivation for studying ribosomal protein modifications originated from the identification of ycfD as a prokaryotic homologue of the human oxygen-sensing 2OG oxygenase FIH and the elucidation of its role *in vitro* as the hydroxylase of ribosomal protein L16. As discussed in the Chapter 1, no 2OG oxygenase has thus far been implicated in oxygen sensing in prokaryotes; known oxygen sensors converge on gene transcription, not translation (19,20). YcfD-catalysed hydroxylation of a ribosomal protein therefore presents a potential oxygen sensing interface that would affect translation. *E. coli* is a facultative anaerobe, its natural habitat in the intestine (21) is hypo-oxygenated (22); tissue oxygenation correlates with protection from bacterial invasion, including *E. coli* (23). *E. coli* appears to be highly responsive to oxygen levels in the intenstinal environment, to which it actively contributes by scavenging oxygen, thereby creating a more anoxic atmosphere; it is able to adapt to the latter by reverting to anaerobic metabolism (24). Complex oxygen sensing mechanisms, which include the *fnr* sensor that switches on the fumarate and nitrate reductases, exist in *E. coli* and confer relative growth advantage in anaerobic conditions (19).

In the *E. coli* growth environment with high variability in nutrient levels (including O_2), the translational machinery may represent a potential check-point for regulating the rate of protein synthesis and thereby affecting the growth rate, with the caveat that the life-time of a ribosome is long on the time-scale of environmental change (determined between 5-10 days in regenerating rat livers, the value depending on the proliferative state of the cells (25)), which may prevent the transmission of signal to the translation machinery in real-time. Ribosomal protein L16 is a key component of the peptidyl transferase centre (PTC) of the bacterial ribosome (26); deletion of the *rplP* gene is lethal (27); it is possible that a modification of L16 may contribute to the rate of protein synthesis (discussed later). The potential role of L16 hydroxylation as an oxygen sensing modification affecting translation was therefore addressed. The effect of atmospheric O_2 on ribosomal protein modifications, including hydroxylation of L16, was investigated by culturing *E. coli* BW25113 in a hypoxic chamber under an O_2 atmosphere of controlled

concentration. 2TY growth medium was used in the presence and absence of glucose to facilitate glycolytic metabolism under hypoxic growth, whereas all other growth parameters, including temperature (37°C), remained unchanged. Hypoxic cells were harvested at OD_{600} 0.8 and their ribosomal proteins purified as described above.

An overlay of UPLC chromatograms of BW25113 grown in the presence and absence of O_2 (21% and 0.1%, respectively) reveals that between 0-50 min, the elution profile remained predominantly unchanged (Figure 2.9A,

Table 2.6). The differences between ribosomes extracted from cells grown at 21% and 0.1% included the disappearance of minor peaks at 14.5 min, 43.5 min and 47 min. However, fronting of the subsequent peak was observed in all cases, suggesting that the differences occur due to poorer chromatographic separation, and may be artefactual. A major difference occured at 53 min (Figure 2.9B): the intensity of a twin peak, corresponding to L12 and L7 (52.8 and 53.2 min, respectively), decreased from 100% to 65% between normoxic and hypoxic growth.

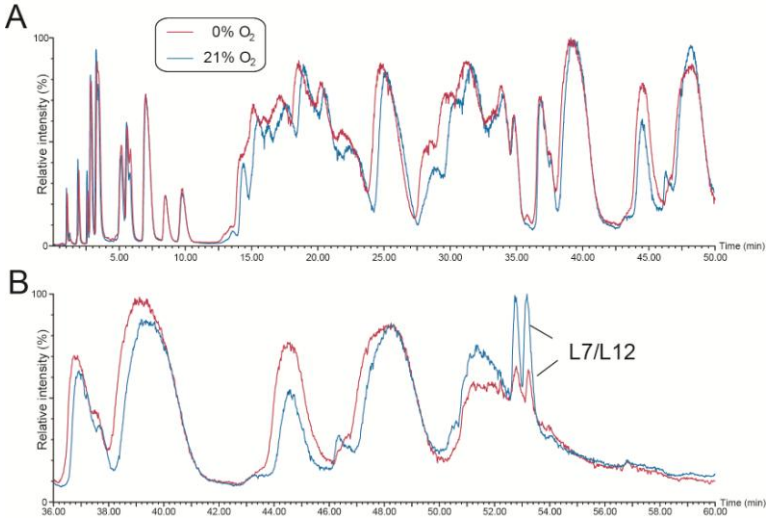


Figure 2.9: UPLC chromatograms of *E. coli* BW25113 grown under a) 21% and b) 0.1% O_2 shows minor differences in intensities of peaks between 0-50 min and a large decrease in L7/L12 concentration during aerobic growth. A) UPLC chromatogram in the elution region 0-50 min reveals differences including the disappearance of minor peaks at 14.5 min, 43.5 min and 47 min for the anoxic sample. Fronting of subsequent peaks suggests that the differences occur due to poorer chromatographic separation. B) UPLC chromatogram of the elution region 36-60 min suggests a major difference in peak intensities at 53 min; the intensity of a twin peak, corresponding to L12 and L7 (52.8 and 53.2 min, respectively), decreases from 100% to 65% between normoxic and anoxic growth. Due to the relative quantification of peak intensities performed by MassLynx, the chromatograms were split into 0-50 min and 36-60 min regions, to enable better comparison of peaks eluting before 50 min.

Table 2.6: Deconvoluted ESI-MS masses of BW25113 70S ribosomal proteins under normoxic (21% O_2) and anoxic (0.1% O_2) growth. MaxEnt1 was used to deconvolute m/z charged states of eluted proteins to 1 Da resolution. Myoglobin was used for calibration post-acquisition.

21% O_2			0% O_2		
Elution time (min)	**Mass (Da)**	**Protein**	**Elution time (min)**	**Mass (Da)**	**Proteins**
1.1	5381	L34	**1.1**	5381	L34
1.9	4364	L36	**1.9**	4363	L36
	6316	L32		6315	L32
2.6	6255	L33+Me	**2.6**	6255	L33+Me
2.9	7158	L35	**2.8**	7158	L35
3.3	8369	S21	**3.3**	8369	S21
	8994	L27		8994	L27
3.5	13653	S12+b-MeS	**3.4**	13652	S12+b-MeS
5.2	9554	L26/S20	**5.1**	9554	L26/S20
5.6	8876	L28	**5.6**	8876	L28
5.9	11450	S14	**5.9**	11450	S14
7.1	11186	L24	**6.8**	11186	L24
8.6	6411	L30	**8.3**	6411	L30
9.9	10300	S19	**9.6**	10300	S19
13.6	8898	S18	**13.8**	8898	S18
14.5	10694	L25	**14.0**	10694	L25
15.6	29731	L2		29731	L2
	10138	S15	**14.5**	10138	S15
16.3	9574	S17	**15.3**	9574	S17
17.5	13003	L19	**15.9**	13003	L19
	13729	S11	**17.1**	13726	S11
19.0	11565	L21	**17.7**	11565	L21
	16020	L13	**17.9**	16020	L13
19.4	14366	L17	**18.0**	14366	L17
20.7	13542	L14	**19.6**	13542	L14
22.4	9191	S16	**21.0**	9191	S16
	12227	L22		12227	L22
	23339	S4		23339	S4
25.2	11200	L23	**24.4**	11200	L23
	22259	L3		12771	L18
28.8	12927	?		22259	L3
			28.0	7275	L29
	14726	S9		14726	S9
	25854	S3		25854	S3
30.5	15328	L16+OH	**28.9**	15312	L16
31.1	12969	S13	**30.3**	12969	S13
	18774	L6		18774	L6
31.7	14981	L15	**30.5**	14981	L15
33.5	13367	L20	**32.2**	13367	L20
	19889	S7		19889	S7
34.0	17516	S5			
34.9	13996	S8	**34.7**	13996	S8
36.9	14871	L11+9Me	**36.5**	14871	L11+9Me
	17586	L10		17586	L10
39.2	15770	L9	**38.9**	15770	L9
39.4	20172	L5		20172	L5
	24600	L1		24600	L1
44.6	26614	S21	**44.5**	26614	S21
48.2	17582	L10	**48.0**	17582	L10
	22088	L4		22088	L4
52.7	12164/12178	L12 +Me	**52.7**	12164	L12
53.2	12206/12220	L7 +Me	**53.2**	12206	L7

Analysis of ribosomal protein masses revealed two main differences: hydroxylation of L16 was abolished to <5% in cells grown at 0.1% O_2 (while the intensity of the chromatographic peak did not change, suggesting that the concentration of L16 was approximately identical under two conditions), and the intensity of the ribosomal protein L7 and L12 peaks decreased by ~65%, suggesting that hypoxic cells contain relatively less L7/L12 than normoxic cells. Reduction in hydroxylation is consistent with an oxygenase-mediated modification that depends on atmospheric O_2, rather than oxygen atoms from water. This is consistent with incorporation of ^{18}O into the ycfD substrate peptide, when ycfD was exposed to $^{18}O_2$ (28).

Another important difference observed was a decrease in the relative abundance of L7/L12, which elute at 53 min (Figure 2.9B). L7/L12 proteins (encoded by the gene *rplL*), identical except for *N*-terminal acetylation present in L7, form the ribosomal stalk complex in combination with L10, and are important for the interaction of a ribosome with the incoming elongation factors (29). L7/L12 are critical for the catalytic activity of the ribosome (30). Ribosomes depleted of L7/L12 can bind elongation factors EF-Tu and EF-G but cannot activate GTP hydrolysis. Furthermore, the rate of GTP hydrolysis *in vitro* depends on the concentration of stalk proteins in the reaction mixture (31). Thus, depletion of L7/L12 may reflect a mechanism for decreasing an overall rate of translation. A crystal structure of the $L12_6$.L10 complex from the thermophilic bacterium *T. thermophilus* has been solved and used to model the stalk onto the 50S subunit. The model demonstrated the spatial proximity of the stalk complex to the incoming ribosome-associated factors, including Ef-Tu, thereby providing a rational for the importance of the stalk complex for the rate of Ef-Tu GTP hydroxlysis (31). The observation that L7/L12 concentration decreases under anaerobic growth therefore suggests a mechanism by which translation rate may be controlled in response to external stress conditions.

Deconvolution of L7/L12 protein signals from normoxic and hypoxic *E. coli* cells suggested that methylation of L7/L12 proteins decreases from ~22% in aerobic conditions, grown to mid-log phase (OD_{600} of 0.8) to < 5% in hypoxia (grown to mid-log phase, Figure 2.10a,c). MS-MS analysis of trypsin-

digested ribosomal proteins separated by SDS-PAGE suggested that L7/L12 methylation occurs on K82 (Figure 2.11). It has previously been shown that L7/L12 methylation is temperature-dependent; methylation at K82 increased with decreasing growth temperature (from 4% to 65% K82 methylation as the temperature decreased from 40°C to 27°C (32).

It was decided to investigate the methylation change of L7/L12 under different growth conditions. Growth phase and temperature variation were investigated by UPLC-ESI-MS methodology. The growth rate profile of the *E. coli* strain BW25113 under normoxic conditions at 37°C was determined by monitoring OD_{600}, then methylation was studied in bacteria harvested at different growth phases (early-log, mid-log and stationary).

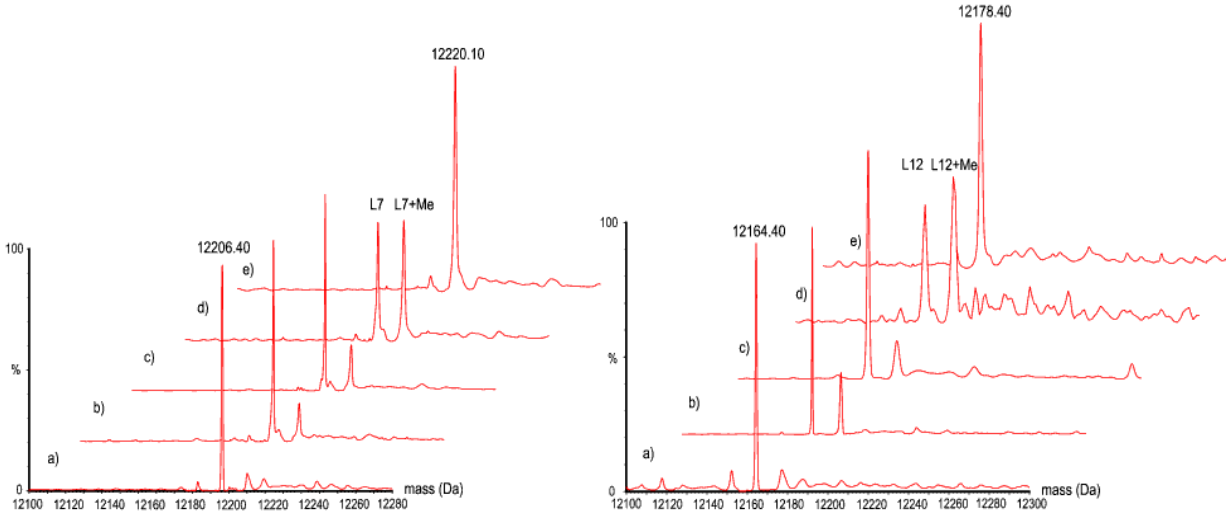


Figure 2.10: Deconvoluted ESI-MS spectra of ribosomal proteins L7 (left) and L12 (right), from *E. coli* (strain BW25113) grown under various conditions. a) BW25113 grown under anaerobic conditions; b) BW25113 grown aerobically, harvested in early-log phase (OD_{600}~0.3); c) BW25113 grown aerobically, harvested in mid-log phase (OD_{600}~0.6); d) BW25113 grown aerobically, harvested in stationary phase (OD_{600}~1.0); e) BW25113 grown aerobically to mid log-phase, then cold-shocked at 18°C and grown to stationary phase at that temperature. The calculated mass of L12 (*N*-terminal Met cleaved) is 12164 Da, and its acetylated analogue L7 is 12206 Da. An increase of 14 Da is observed for both proteins, consistent with methylation.

Comparison of the methylation status from three growth phases of *E. coli* suggested that the modification is highly dynamic: methylation increased from ~15% at OD_{600} 0.3 (early log phase) to ~20% at mid-log (OD_{600} 0.6) and ~55% at OD_{600} 1.0 (stationary phase). The same methylation pattern was observed for L7 and L12 independently (Figure 2.10 b-d). Consistent with previous reports (32), cold-shock resulted in

92% methylation of L7/L12 (Figure 2.10e). The precise role of L7/L12 methylation is unclear, but it has been proposed to function as a 'sensor' to changes in environmental conditions and to contribute to reducing the rate of translation in harsh environmental conditions, such as low temperature (18°C) (32). The enzyme that catalyzes the methylation remains unknown. Preliminary attempts at co-immunoprecipitation with FLAG-tagged L7 failed (data not shown).

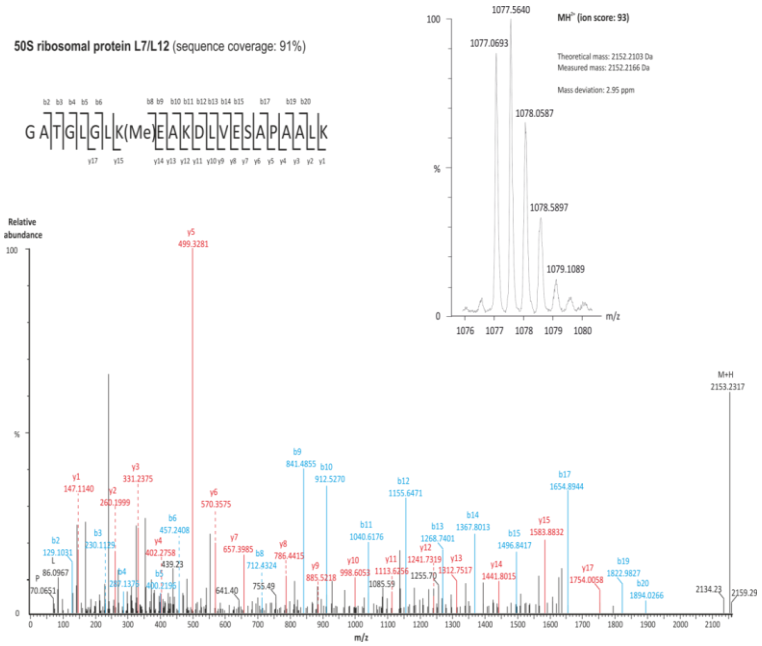


Figure 2.11: MS/MS chromatogram of the tryptic peptide of ribosomal protein L7/L12 (GATGLGLK(me)EAKDLVESAPAALK), showing methylation at residue K82. Inset: m/z spectrum of the intact tryptic fragment (MH^{2+} state). The figure was kindly provided by Dr Mukram Mackeen.

L16 hydroxylation was analysed under the conditions that led to a change in L7/L12 methylation (Figure 2.12). In mid-log phase, L16 hydroxylation was >95% (Figure 2.12b). Intriguingly, the cells harvested at early-log and stationary growth phases showed the presence of a peak at 15309 Da which varied in intensity between 14% (early-log) and 16% (stationary, Figure 2.12 a,c). The peak was also observed in *E. coli* subjected to cold-shock (10% intensity relative to the L16-OH peak, Figure 2.12d). Due to the low intensity of the 15309/15310 Da peak it was not possible to characterise the peak by manual deconvolution and thereby determine the accuracy of the MS signal. The presence of a small amount (15%) of unhydroxylated L16 (expected mass 15312 Da) or dehydroxy-arginyl L16 (predicted mass 15310 Da) therefore cannot be excluded.

This is the first demonstration that the ribosomal machinery is subject to modifications that depend on environmental changes in oxygen concentration - environmental growth conditions of the organism appear to be directly relayed onto the structure of the ribosome itself. Although a number of ribosomal proteins are post-translationally modified (see Table 2.5), only hydroxylation of L16 and methylation of L7/L12, as well as their abundance, vary with oxygen concentration; suggesting that they may perform an oxygen sensing role. Detailed kinetic experiments on isolated modified ribosomes would allow a dissection of the role of modifications. However, since the methyltransferase which catalyzes L7/L12 modification is unknown, only the effects of ycfD could at the present be studied.

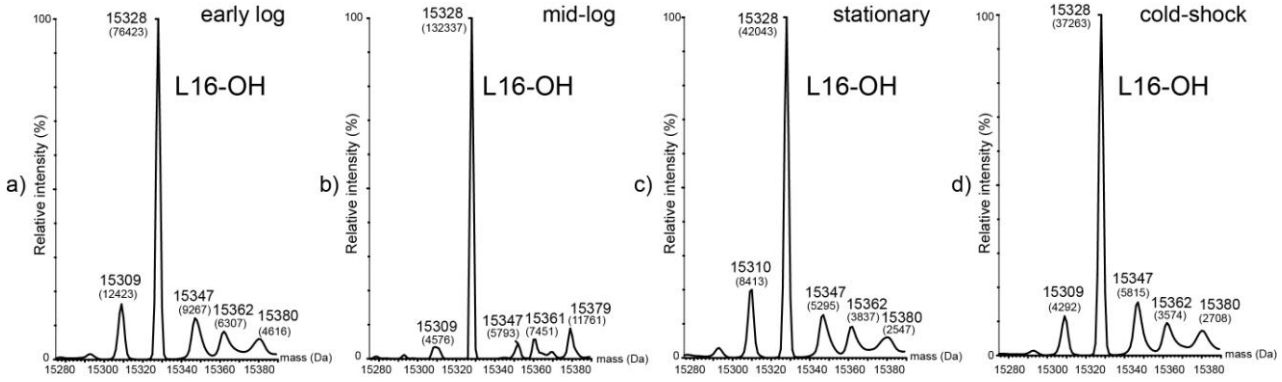


Figure 2.12: Deconvoluted ESI-MS spectra of ribosomal protein L16 from *E. coli* (strain BW25113) grown under various conditions: a) BW25113 grown in normoxia (21% O_2) at 37°C, harvested in early-log phase (OD_{600} ~0.3); b) BW25113 grown in normoxia at 37°C, harvested in mid-log phase (OD_{600} ~0.6); c) BW25113 grown in normoxia at 37°C, harvested in stationary phase (OD_{600} ~1.0); d) BW25113 grown in normoxia at 37°C to mid log-phase, then exposed to cold-shock (18°C) overnight. The calculated mass of L16 is 15281.2 Da. MaxEnt1 deconvolution was used to determine masses to 1 Da resolution.

2.5. *YcfD* as a putative oxygen sensor

Next, the variation of L16 hydroxylation was investigated as a function of changes in the levels of atmospheric O_2. The *E. coli* BW25113 strain was grown to mid-log phase (OD_{600} 0.8) at 37°C in a hypoxic chamber pre-equilibrated at selected concentrations of O_2 varying from 0-20%. L16 hydroxylation increased with increasing pO_2 (Figure 2.13). At 0% O_2, L16 was <5% hydroxylated; hydroxylation increased to 30% at 5% O_2, 45% at 10% O_2, and 80% at 20% O_2 (Figure 2.13A).

The results suggest that, at least in mid-log phase, the catalytic activity of ycfD appears to be limited by oxygen concentration. The increase in the percentage of hydroxylated L16 detected by WPMS was approximately linear with pO_2 (R^2=0.992, Figure 2.13B), consistent with an oxygen sensing role in cells. The results are preliminary and would require biological repeats to reach statistical significance.

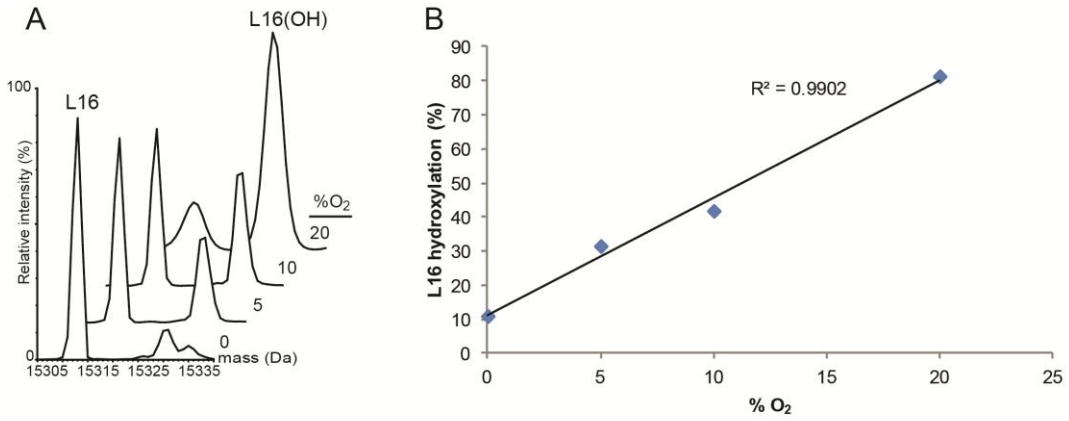


Figure 2.13: A) Deconvoluted ESI-MS spectra of ribosomal protein L16 from *E. coli* (strain BW25113) grown under increasing O_2 concentration (0-20%). BW25113 was grown in unbaffled Erlenmeyer growth flasks (1 L) containing 500 mL 2TY media supplemented with 2% glucose in a hypoxic chamber incubated at indicated concentrations of O_2 at 37°C. Stirring was achieved with a suitable stirrer bar rotating at a rate that induced vortex formation in the culture. When cells reached OD_{600} 0.8 they were harvested by centrifugation **B) L16 hydroxylation increases linearly with pO_2.** L16 peak intensities measured by ESI-TOF-MS were used to calculate %OH for each concentration of O_2 and were plotted against pO_2. Linear regression function suggested that at L16 hydroxylation varies linearly with pO_2, at least under the conditions tested.

2.6. Phenotypic studies on the effect of *ycfD* deletion

The demonstration that ycfD-catalysed L16 hydroxylation varies with growth conditions, including O_2 concentration, suggests it may serve a sensory role. For this to be the case, lack of L16 hydroxylation should confer a growth phenotype under selected conditions. A number of ribosomal protein/ribosome-associated factor modifications have been found to affect cellular processes: (i) deletion of the *E. coli* L3 methyltransferase PrmB confers cold-sensitivity, which is manifested as abnormally slow growth at 22°C, and affects ribosome biogenesis, observed as a reduction in intensity of 30S and 50S peaks, and an increase in intensity of intermediate peaks in ribosome profiles (33); (ii) acetylation of the *E. coli* S5, catalysed by the acetyltransferase RimJ, is regulated by the ribosome tertiary structure (34) and is a marker of defective ribosomal assembly. It correlates with a cold-sensitivity phenotype and reduced translation accuracy (35); (iii) release factors 1-2 (RF1-2) are methylated on Q240, the key residue for the

catalytic function of RFs, by methyltransferase PrmC. Deletion of PrmC has no effect on cell growth in rich media but shows considerably reduced growth on poor carbon sources (36); (iv) deletion of the sarcin-ricin loop, which results in the loss of L16 from the ribosome, results in a growth phenotype at 42°C (37).

Deletion of ribosomal protein/ribosome-associated factor modifying enzymes therefore affects at least five observable parameters:

i) Growth rate at low/high temperatures (22°C/42°C)

ii) Growth in nutrient-poor media

iii) Ribosome assembly

iv) Rate of translation

v) Accuracy of translation

The role of L16 hydroxylation in the above processes was investigated by comparison of the *E. coli* strains BW25113 and JW1114.

2.6.1. *YcfD* deletion affects growth rate at 20°C and in minimal media

The effect of *ycfD* deletion on growth rate was measured under three temperatures: at 20°C, 37°C (standard *E. coli* growth temperature) and 42°C. BW25113 and JW1114 grew similarly at 37°C, which was therefore used as a control temperature (Figure 2.14). When growth temperature was increased to 42°C, an initial lag in growth of the *ycfD*-deleted strain was observed at 6h. At 10 h, the growth difference was no longer observable, and by 16 h the dilution series of BW25113 and JW1114 appeared identical. In contrast, when the strains were grown at 20°C, a severe growth defect of JW1114 was observed. The growth rate of both strains was significantly lower compared to 37°C, and colony growth was observable only after 48 h. Despite the identical initial concentration of bacterial cultures, colony growth was observable for BW25113 only (Figure 2.14). The experiment was independently replicated twice with identical results. It was concluded, therefore, that ycfD affects growth at low temperatures, which may indicate defects in energy-consuming processes, such as translation.

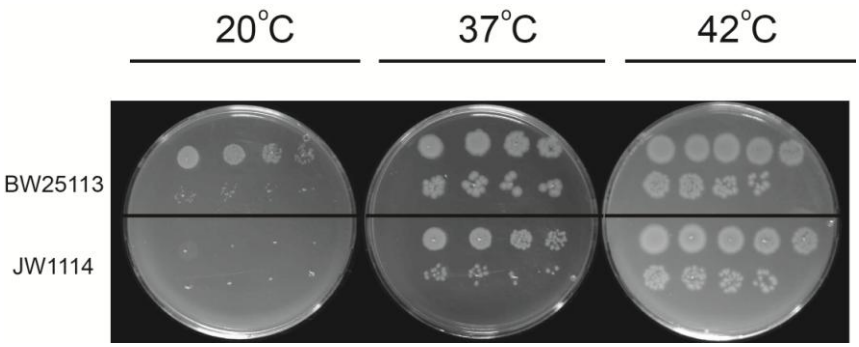


Figure 2.14: Colony growth on LB-agar plates reveals a growth phenotype of *ycfD*-deleted strain (JW1114) at low temperature (20°C). A serial 10-fold dilution of bacterial cultures was prepared from an inoculate of a bacterial lawn (grown within 2 days of performing the experiment) in fresh 2TY medium to initial OD_{600} 0.05. Diluted bacterial cultures were spotted on LB-agar plate (5 µL each). Plates were incubated in a 20°C, 37°C, and 42°C incubator, for indicated amounts of time. Growth at 37°C was used as a control and was performed alongside each separate experiment.

2.6.2. *YcfD* deletion affects growth in minimal media

The effect of ycfD on growth was then was assessed in minimal medium M9 supplemented with 2% glucose as the carbon source (Figure 2.15, experiment conducted by Dr. Alexander Wolf). No difference in growth rate was observed between WT and KO strains in 2TY (rich) medium, whereas growth in M9 minimal medium resulted in retardation of growth for both strains; WT demonstrated a retardation in reaching mid-log phase by 5 h compared to rich medium. The KO strain showed only a minimal extent of growth, OD_{600} increased from 0.3 to 0.5 (OD_{600} increased from 0.3 to 1.2 for WT grown under identical conditions). Furthermore, a 40% decrease in L16 hydroxylation was observed in M9 medium (Figure 2.15B), compared to the complete hydroxylation observed in 2TY media, suggesting that ycfD may sense nutrient deprivation, contributing to slower growth rate in minimal media.

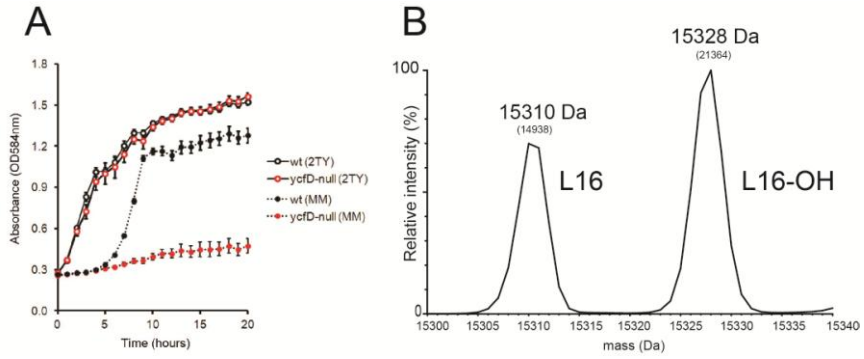


Figure 2.15: YcfD promotes bacterial cell growth under conditions of limited nutrient availability. Rich media (2TY) or minimal media (MM) were inoculated with a smear of bacterial lawn of Wild-type (wt, BW25113) or ycfD-deleted (JW1114) strains grown on LB-agar plates. OD_{600} of inoculates was adjusted to 0.05 cells; growth rate of 200 µL of bacterial culture was then monitored at 584 nm at 37°C. Figure 17A kindly provided by Dr. Alexander Wolf.

2.6.3. *YcfD* does not affect ribosomal assembly

Ribosome assembly can be monitored using density gradient centrifugation, which separates ribosomal subunits according to their sedimentation coefficient (from which the names of subunits, 30S, 50S and 70S are derived). The method uses a medium of high density, such as sucrose, glycerol or KCl_{aq}, the density of which increases from top to bottom of a centrifuge tube. Application of centrifugal force causes migration of particles with a sedimentation coefficient towards the region of equal density. At the end of a centrifugal run, the sample is forced out of the centrifugation tube by a medium of higher density than sucrose, and is passed through a UV-detector measuring absorbance at 260 nm. The resulting UV trace is characteristic of the ribosome assembly status. In eukaryotes, ribosomes can be separated into the 40S, 60S, 80S and polysomal particles, which are resolved in a ribosome profile. In bacteria, polysomes are not observed, and the ribosome profile consists of the small (30S) and the large (50S) subunit, as well as the fully assembled ribosome (70S, Figure 2.16). Ribosomal assembly defects may result in peaks of intermediate sedimentation coefficient.

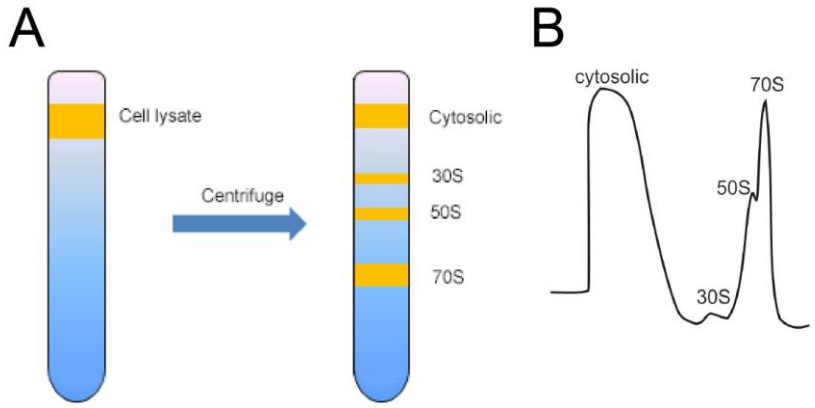


Figure 2.16: Sucrose density gradient centrifugation of ribosomal subunits. A) Cell lysate is loaded onto a layer of sucrose with increasing concentration towards the bottom. Samples are centrifuged at high speed (100,000 g) to separate particles with defined sedimentation coefficients. Layers with different densities, containing cellular components of different sedimentation coefficients, are then pushed out of the tube in reverse direction; the presence of ribosomes is analysed by measuring absorbance at 260 nm, indicative of RNA. A ribosome profile (B) is obtained, which consists of a cytosolic peak, the small subunit (30S), large subunit (50S) and assembled ribosome (70S).

In order to assess the effects of ycfD on ribosomal assembly by comparison of WT to KO strains, the method was first optimised for optimal concentration of the cellular lysate, and the optimal sucrose density concentration gradient (Figure 2.17). When the concentration of lysate loaded onto the sucrose

gradient was varied from 5-30 a.u., optimal UV signal was obtained at 10 a.u. load, which was used for subsequent studies (Figure 2.17A). Next, the concentration gradient of sucrose was varied: cytosolic and 30S peaks were resolved in 10-50% and 10-25% sucrose gradients, however the 50S peak appeared as a shoulder of the 70S peak when 10-50% sucrose gradient was applied, and a shallower gradient of 10-25% over 11 mL was subsequently used (Figure 2.17B).

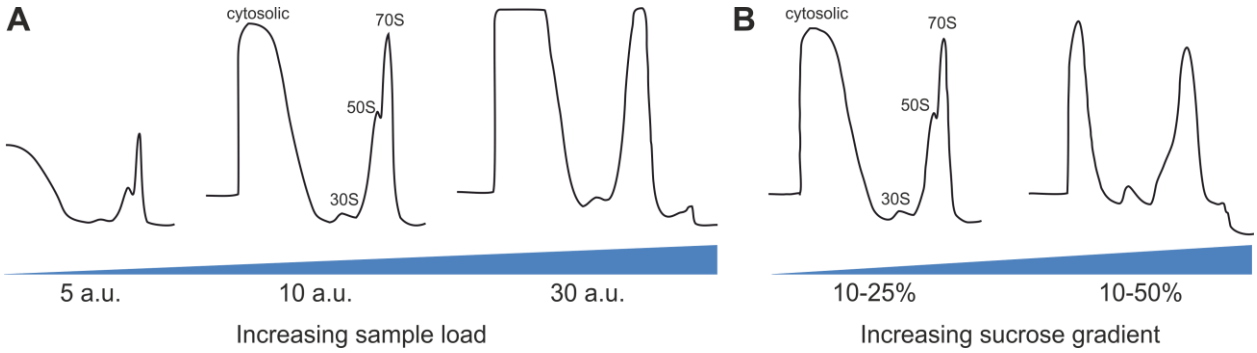


Figure 2.17: UV traces of *E. coli* ribosome profiles separated by density gradient for optimisation of conditions. The profiles were recorded by measuring the absorbance at 260 nm. A) Optimising the concentration of BW25113 lysate applied onto the sucrose gradient The sample concentration was measured as absorbance (in absorbance units, a.u.) at 260 nm using NanoDrop. B) Optimising the sucrose gradient concentration for optimal separation of 30S, 50S and 70S ribosome components. Two concentration ranges were tested: 10-25% sucrose and 10-50% sucrose.

Comparison of the ribosome profiles of BW25113 and JW1114 grown in 2TY media at 37°C revealed an identical profile for the two strains (Figure 2.18), suggesting that under the conditions tested, ycfD is not involved in ribosome biogenesis. When the strains were grown at 30°C in 2TY medium, the 70S peak, corresponding to assembled ribosome, of JW1114 was ~1.25× more intense than the cytosolic peak. This observation suggests that a relatively higher proportion of ribosomes is assembled into the 70S particle in ycfD-deleted strain at 30°C.

Chloramphenicol affects ribosomal assembly by stalling ribosome maturation at 45S stage, leading to accumulation of pre-ribosomal particles (38). Although only partial resolution of the 70S and pre-ribosomal particles was achieved (Figure 2.18), the comparison of BW25113 and JW1114 ribosome profiles revealed that their ribosomal profiles upon chloramphenicol treatment are affected to a similar extent, suggesting that the effects of chloramphenicol on ycfD KO, discussed in Chapter 3 do not operate through ribosome assembly or biogenesis.

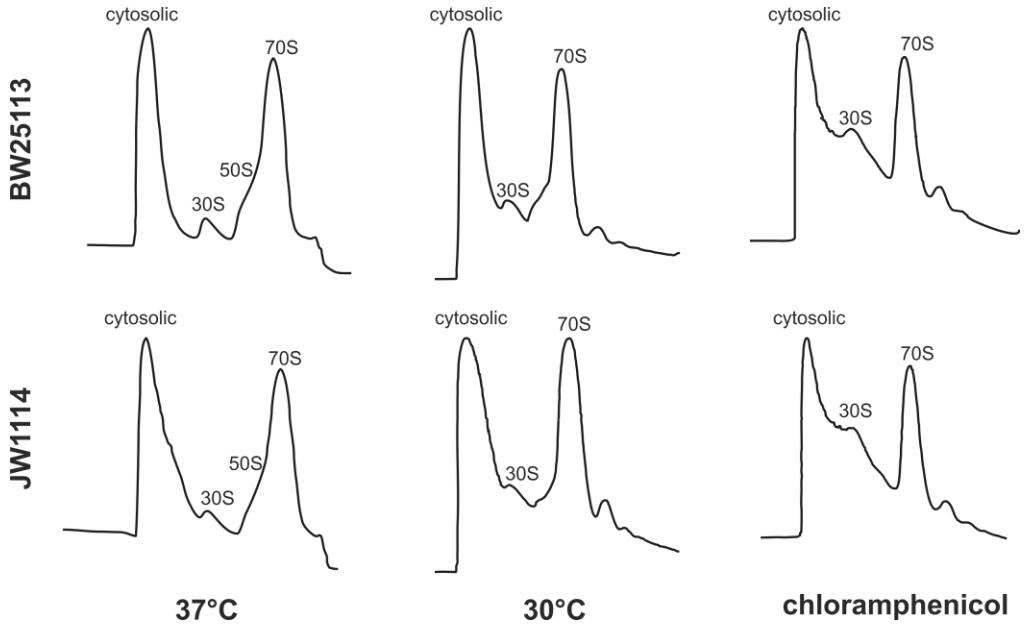


Figure 2.18: Comparison of ribosome profiles of BW25113 and JW1114 grown to mid-log phase (OD_{600} 0.7) at 37°C, 30°C, and after treatment with 7 μg/mL chloramphenicol at 37°C for 3 h. The figure shows UV traces of ribosomal assemblies separated by sucrose density gradient (10-50%) measured as absorbance at 260 nm.

2.6.4. *YcfD* does not affect translational accuracy

Effects on translational accuracy by gene deletions or ribosomal antibiotics can be measured using dual luciferase reporter (DLR) systems (39), which employ a construct containing two luciferase genes (typically from *Renilla* and firefly separated by a spacer containing a stop codon). Upon translation, the ribosome reads through the first luciferase gene, producing an active enzyme which converts luciferase substrates (coelenterazine and luciferin for *Renilla* and firefly luciferases, respectively) into luminescent derivatives; the intensity of luminescence can be measured spectrophotometrically. The ratio of intensities of firefly to *Renilla* substrate luminescence is then correlated with relative levels of translation of the two genes; stop-codon readthrough is calculated from the ratio of firefly:*Renilla* (FLuc/RLuc) substrate conversion.

The processivity of translation was measured indirectly by assessing the activity ratios RLuc:Fluc luciferases separated by a sense codon (CGA, Figure 2.19B). The activity ratios for BW25113 and JW1114 were equal within experimental error, suggesting that *ycfD* deletion does not affect ribosomal

processivity. The efficiency of translation termination was tested in rich media in stationary phase, for three stop codons (TAG, TAA and TGA) in two types of growth media, 2TY and LB (Figure 2.19C-D). Within experimental error, the luciferase activity ratios were equal for BW25113 and JW1114 for all stop codons tested, in both types of media. Overall, the results suggested that L16 hydroxylation by ycfD does not play an important role in control of translational accuracy.

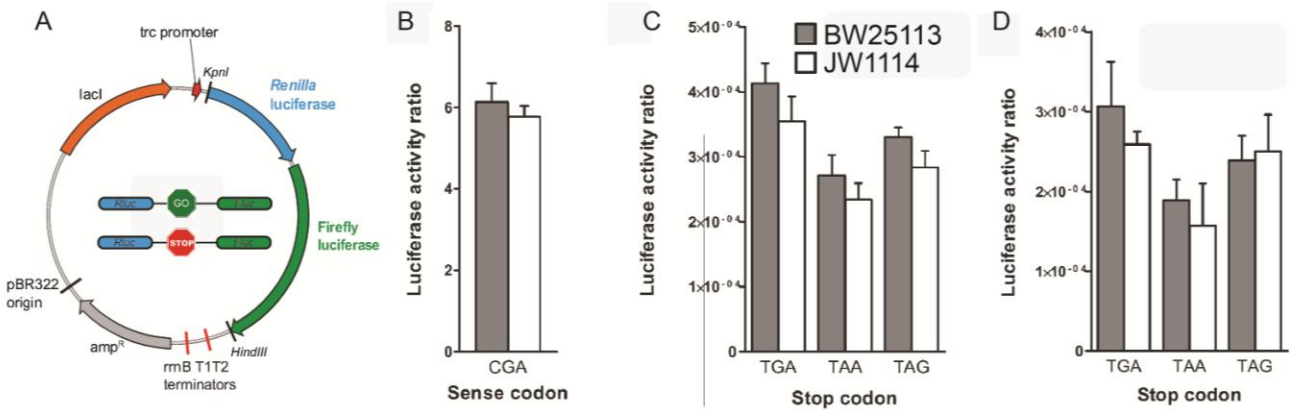


Figure 2.19: Stop-codon readthrough assays on BW25114 and JW1114 using a vector containing *Renilla* and firefly luciferases separated by a sense codon or a stop codon. A) The vector map of the pTRC99a-derived iD6000 vector used to determine translation accuracy. The key vector features include *Renilla* and firefly luciferase genes separated by a variable insert containing either a sense (CGA) or stop codons (TGA, TAA, TAG). The vector contains a *trc* promoter, which promotes constitutive, IPTG-independent expression. The *lacI* repressor suppresses luciferase expression under non-inducing conditions. Ampicillin resistance gene provides a selection marker that is orthogonal to inherent kanamycin-resistance of the *ycfD* gene. B) The processivity of ribosomes from BW25113 and JW1114 was measured by luciferase activity ratios for expression of a sense-codon containing iD6000 construct. C/D) Translation termination accuracy was measured by determining the luciferase activity ratios for the iD6000 construct containing *Rluc* and *Fluc* genes separated by a stop codon (TGA, TAA, TAG). The ratios were determined in C) 2TY and D) LB-Miller medium. All experiments performed by Armin Thalhammer, who kindly provided the figure.

2.6.5. *YcfD* affects global translation rate

Due to the considerable energetic requirement of protein translation (~2/3 of the cellular ATP content is dedicated to protein synthesis (40)), growth and translation rates in bacteria are intimately linked (41,42). The effect of ycfD on global translation rate was therefore tested using a modified 'pulse-chase' experiment which measured the incorporation of ^{35}S methionine in newly-synthesised proteins (Figure 2.20) with Dr Mathew Coleman (Welcome Trust Centre for Human Genomics, University of Oxford). The global translation rates of *E. coli* BW25113 and JW1114 were studied in M9 minimal medium, in which *ycfD* KO demonstrated a growth defect. BW25113 and JW1114 cells were grown to mid-log

phase, then diluted to OD_{600} 0.1 before spiking the media with [^{35}S]-Met. At specified time-points, aliquots were removed from the culture, centrifuged and lysed by trichloroacetic acid precipitation. The rate of incorporation of [^{35}S]-Met was measured by scintillation counting.

The rate of global protein synthesis in JW1114, as measured by [^{35}S]-Met incorporation, was 1.1-1.4×10^3 dpm/min, whereas the rate of incorporation in BW25113 was 2.9-6.9×10^3 dpm/min. The global translation rate is therefore 4-6x lower in *ycfD*-deleted cells compared to the WT cells (Figure 2.21). It can be concluded that under the conditions tested, L16 hydroxylation increases global translation rate.

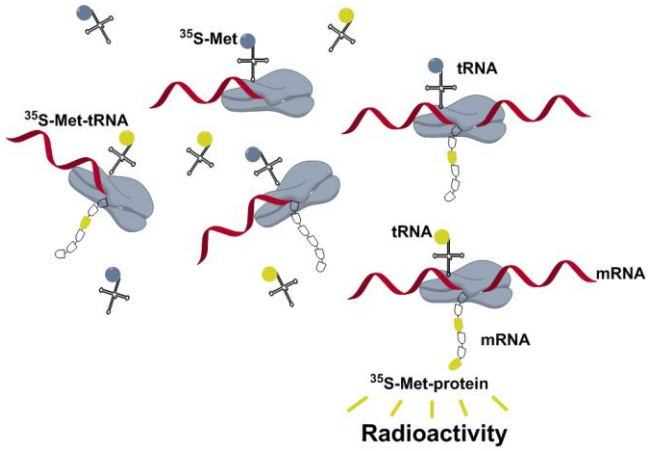


Figure 2.20: Bulk translation rate is measured by exposing a bacterial culture grown in defined minimal medium lacking exogenous tRNA to a dose of radio-labelled methionine (^{35}S-Met, yellow circles). Radioactive [^{35}S]-Met is incorporated into nascent polypeptide (elongated chains), leading to a measurable time-dependent increase in radioactivity, which is measured over a defined time period.

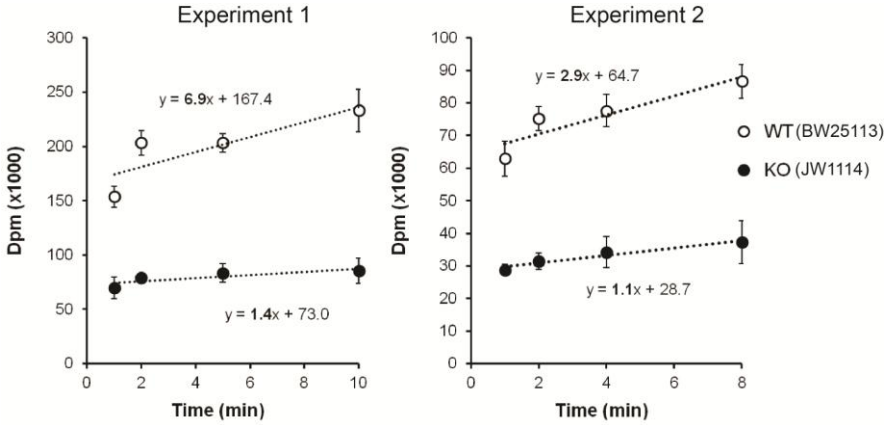
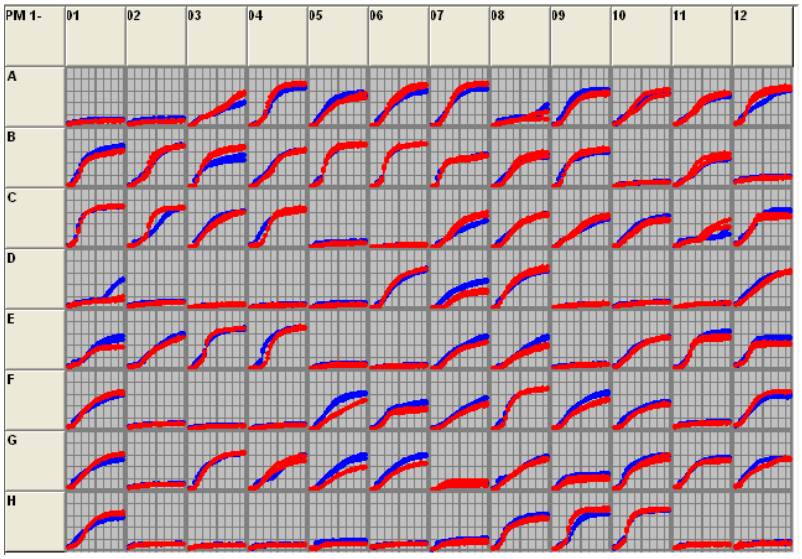


Figure 2.21: *E.coli* ycfD knockout strain displays reduced bulk translation rate. Wild-type (BW25113) and *ycfD*-deleted (JW1114) strains grown in minimal M9 media supplemented with glucose (0.2%) were labelled with [^{35}S]-methionine for the indicated times, before measuring radiolabel incorporation into trichloroacetic acid precipitates by scintillation counting. The results of two independent experiments are shown.

2.6.6. *YcfD* deletion decreases growth rate on single carbon sources short-chain carboxylic acids, TCA cycle intermediate analogues and amino acids

Phenotypic differences between *ycfD* WT and KO *E. coli* strains were further assessed by a phenotypic microarray study (BioLog PM1 Microplate™, conducted by the Animal Health and Veterinary Laboratory Agency, Weybridge, UK), which analysed growth differences in the presence of carbon sources and ribosomal antibiotics. The study monitored cellular respiration as a reporter for cellular growth on a single carbon source. Under normal growth conditions, cells undergo active respiration, which leads to chemical reduction of a patented tetrazolium dye (structure not disclosed), the product of which is strongly coloured. A weakly positive or negative phenotype leads to a slower or discontinued respiration, therefore little or no colour is observed in the well (Figure 2.22).



Carbon source	Δ (log)	Well
Glycerol	-0.39	B3
D-Fructose	-0.36	C7
Adenosine	0.30	E12
α-Hydroxy-Butyric Acid	0.31	E7
Propionic Acid	0.33	F7
L-Glutamine	0.35	E1
Glycolic Acid	0.38	F9
β-Methyl-D-Glucoside	0.41	E8
Bromo-Succinic Acid	0.41	F6
L-Alanyl-Glycine	0.45	G6
Mono Methyl Succinate	0.50	G9
Fumaric Acid	0.62	F5
L-Alanine	0.65	G5
α-Keto-Butyric Acid	0.84	D7

Figure 2.22: Phenotype MicroArray Single Panel Comparison Display showing Cellular Responses for 96 Kinetic Assays on Biolog PM1 Microplate™ Carbon Sources plate. Blue: BW25113; red: JW1114. A01: Negative Control, A02: L-Arabinose, A03: *N*-Acetyl-D-Glucosamine, A04: D-Saccharic Acid, A05: Succinic Acid, A06: D-Galactose, A07: L-Aspartic Acid, A08: L-Proline, A09: D-Alanine, A10: D-Trehalose, A11: D-Mannose, A12: Dulcitol, B01: D-Serine, B02: D-Sorbitol, B03: Glycerol, B04: L-Fucose, B05: D-Glucuronic Acid, B06: D-Gluconic Acid, B07: D,L-α-Glycerol- Phosphate, B08: D-Xylose, B09: L-Lactic Acid, B10: Formic Acid, B11: D-Mannitol, B12: L-Glutamic Acid, C01: D-Glucose-6-Phosphate, C02: D-Galactonic Acid-γ-Lactone, C03: D,L-Malic Acid, C04: D-Ribose, C05: Tween 20, C06: L-Rhamnose, C07: D-Fructose, C08: Acetic Acid, C09: α-D-Glucose, C10: Maltose, C11: D-Melibiose, C12: Thymidine, D01: L-Asparagine, D02: D-Aspartic Acid, D03: D-Glucosaminic Acid, D04: 1,2-Propanediol, D05: Tween 40, D06: α-Keto-Glutaric Acid, D07: α-Keto-Butyric Acid, D08: α-Methyl-D-Galactoside, D09: α-D-Lactose, D10: Lactulose, D11: Sucrose, D12: Uridine, E01: L-Glutamine, E02: m-Tartaric Acid, E03: D-Glucose-1-Phosphate, E04: D-Fructose-6-Phosphate, E05: Tween 80, E06: α-Hydroxy Glutaric Acid-γ-Lactone, E07: α-Hydroxy-butyric Acid, E08: β-Methyl-D-Glucoside, E09: Adonitol, E10: Maltotriose, E11: 2'-Deoxy-Adenosine, E12: Adenosine, F01: Glycyl-l-Aspartic Acid, F02: Citric Acid, F03: m-Inositol, F04: d-Threonine, F05: Fumaric Acid, F06: Bromo-Succinic Acid, F07: Propionic Acid, F08: Mucic Acid, F09: Glycolic Acid, F10: Glyoxylic Acid, F11: D-Cellobiose, F12: Inosine, G01: Glycyl-L-Glutamic Acid, G02: Tricarballylic Acid, G03: L-Serine, G04: L-Threonine, G05: L-Alanine, G06: L-Alanyl-Glycine, G07: Acetoacetic Acid, G08: *N*-Acetyl-β-d-Mannosamine, G09: Mono-Methyl Succinate, G10: Methyl Pyruvate, G11: D-Malic Acid, G12: L-Malic Acid, H01: Glycyl-L-Proline, H02: p-Hydroxy-Phenylacetic Acid, H03: m-Hydroxy-Phenylacetic Acid, H04: Tyramine, H05: D-Psicose, H06: L-Lyxose, H07: Glucuronamide, H08: Pyruvic Acid, H09: L-Galactonic Acid-γ-Lactone, H10: D-Galacturonic Acid, H11: β-Phenylethylamine, H12: Ethanolamine.

The area under the curve (AUC), corresponding to the growth rate of *E. coli* cells, was measured in triplicate for the growth of BW25113 and JW1114 in the presence of 96 carbon sources. Δlog(AUC) values in Figure 2.22 are reported for carbon sources that generated Δlog(AUC)>±0.3 and suggested that the *ycfD*-deleted strain is less able to grow on a selected carbon source (positive Δlog(AUC) value) or better able to grow on the selected carbon source (negative Δlog(AUC) value).

Three conditions were identified in which the metabolic activity (log(AUC)) of the ycfD KO strain was >0.3 greater compared to WT: glycerol, D-galactonic acid-γ-lactone and D-fructose (Figure 2.22). Melibiose, although affecting growth of WT more negatively than of KO, resulted in severe suppression of metabolic activity of both strains and was therefore not considered as significant.

Several conditions resulted in greater metabolic activity of WT compared to KO: L-asparagine led to a decreased initial metabolic activity of both strains, which was recovered by WT but not KO strain. α-Keto-butyric acid resulted in difference in log(AUC) of 0.84, suggesting a markedly greater metabolic activity of WT compared to KO; L-glutamine, fumaric acid, glycolic acid, L-alanine and L-alanyl-glycine affected the metabolic activity of ycfD KO more negatively than the WT by Δlog(AUC) of 0.35, 0.62, 0.38, 0.65 and 0.45, respectively. The identification of α-keto/α-hydroxy acids (α-keto-butyric acid, α-hydroxy-butyric acid and glycolic acid) and TCA cycle intermediates or analogues thereof (bromo-succinic acid, mono methyl succinate and fumaric acid) and metabolically important amino acids (glutamine and asparagine) as having a relative negative effect on the metabolic activity of *ycfD* KO compared to WT suggested that L16 hydroxylation by ycfD may be involved in the regulation of the TCA cycle intermediate metabolism (Figure 2.23).

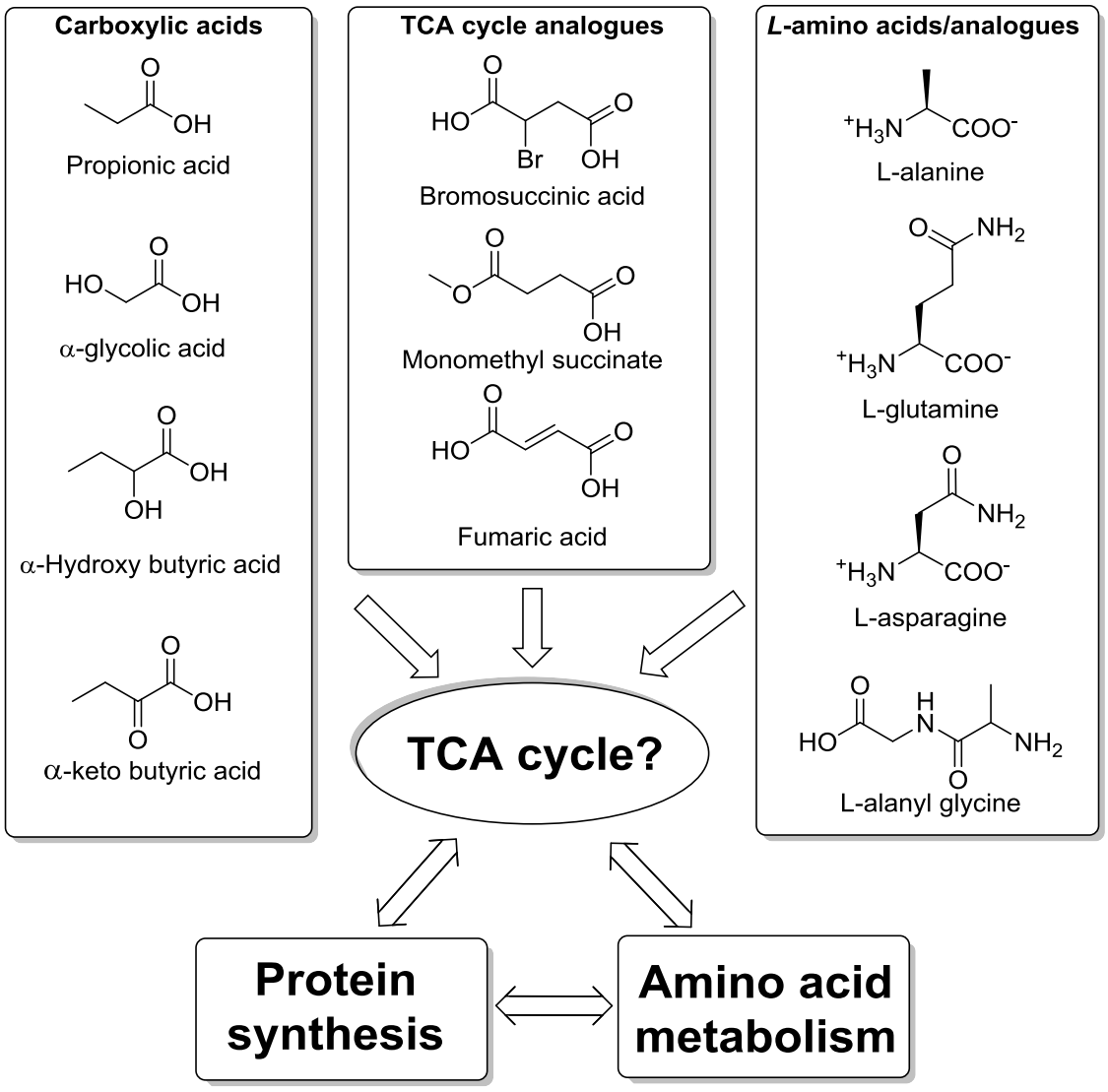


Figure 2.23: Carbon sources which lead to lower metabolic rate of *ycfD* knock-out relative to wild-type, identified by the Phenotype MicroArray Single Panel Comparison Display can be classified into i) short-chain carboxylic acids, ii) TCA cycle intermediate analogues and iii) l-amino acids/analogues. The decreased ability of *ycfD* KO to utilise these carbon sources may suggest alterations in metabolism, including deregulation of TCA cycle intermediates, which may lead to effects on protein synthesis and amino acid metabolism.

2.7. *YcfD* and TCA cycle metabolites

Aerobic organisms generate most of their chemical energy for survival from a series of chemical reactions that oxidise acetate derived from carbohydrates, fats and proteins, producing CO_2 and adenosine triphosphate (ATP) as an energy-rich molecule used as cellular fuel. The reactions are connected into a chemical series that begins and ends with production of citric acid, wherefrom the name 'tricarboxylic acid (TCA) cycle' stems. The cycle produces reducing equivalents in the form of nicotinamide adenine dinucleotide (NAD+) and flavin adenine dinucleotide (FAD), as well as ATP via oxidative phosphorylation. The TCA cycle is responsible for the generation of approximately 67% of all reducing equivalents per molecule of glucose and is therefore the energy generator of choice for aerobic organisms. The TCA cycle flux is dependent upon the supply of acetyl units and activation of three non-equilibrium reactions within the TCA cycle (citrate synthase to produce citrate, isocitrate dehydrogenase to produce 2OG and 2OG dehydrogenase to produce succinyl-coA). Many enzymes are responsible for correct cycling of the TCA cycle, converting one intermediate into the next. However, since many intermediates are substrates for enzymes not directly involved in the propagation of the TCA cycle, it is possible that a deregulation of the cycle can be caused by activity of a TCA-cycle consuming enzyme. 2OG-dependent oxygenases utilise 2OG as a co-substrate for oxidation of their prime substrates leading to oxidative decarboxylation of 2OG to succinate. High catalytic activity of 2OG oxygenases may therefore affect the ratio of 2OG:succinate, leading to a misbalance in the cycle.

In humans, mutations in isocitrate dehydrogenase (IDH1/2), succinate dehydrogenase (SDH) and fumarate hydratase (FH) increase intracellular levels of fumarate, and have been found to upregulate the levels of (*R*)-2-hydroxyglutarate, succinate and fumarate, respectively, which can act as competitive inhibitors of 2OG oxygenases, including hypoxia-inducible factor (HIF) hydroxylases, histone and DNA demethylases, thus activating oncogenic HIF pathways under normal oxygen tensions (43).

The effects of a 2OG oxygenase on the levels of TCA cycle intermediates in prokaryotes have not been addressed to date. YcfD catalyzes the hydroxylation of an abundant cellular protein (ribosomes constitute ~1/3 of the cellular proteome (41)) and is therefore likely constitutively active, thereby converting 2OG into succinate. *YcfD* deletion may therefore affect the levels of TCA cycle intermediates, which could be assessed in a time-dependent study of metabolite concentrations by LC-MS in *ycfD* WT and KO cells. The LC-MS experiment was performed by Khalid al-Qahtani.

The analysis of metabolite levels revealed strikingly similar profiles for BW25113 and JW1114 strains for all metabolites apart from glutamine. The levels of citrate, *cis*-aconitate and fumarate were stable throughout the duration of growth, and identical for two strains within experimental error. Surprisingly, the last sampled time-point (48 h) resulted in a marked increase in levels of citrate, *cis*-aconitate and fumarate for JW1114 but not BW25113. Because the levels of other metabolites analysed did not follow the same trend, it is unlikely that this is a result of an experimental error, although the biological significance is unclear. Levels of 2OG and succinate were identical for the two strains; both were observed to decrease during the exponential rate of bacterial growth (1-8 h) before an increase at 8 h. After overnight incubation the levels were found to be depleted (16-24 h), suggesting that nutrient supply from the growth medium has been exhausted. This is consistent with the trend observed for glutamate, which decreased linearly during the exponential growth phase of both BW25113 and JW1114, temporarily increased when bacteria reached maximal OD_{600} before decreasing again during overnight growth in stationary phase.

Intriguingly, glutamine levels increased during the exponential growth phase, before decreasing when bacteria reached the stationary phase. The levels of glutamine during the exponential phase were >1 s.d. larger for JW1114 than for BW25113. This is consistent with the phenotypic microarray result presented above, which suggested that JW1114 is less able to utilise glutamine as the sole carbon source than BW25113. The result suggests that glutamine metabolism is deregulated in JW1114.

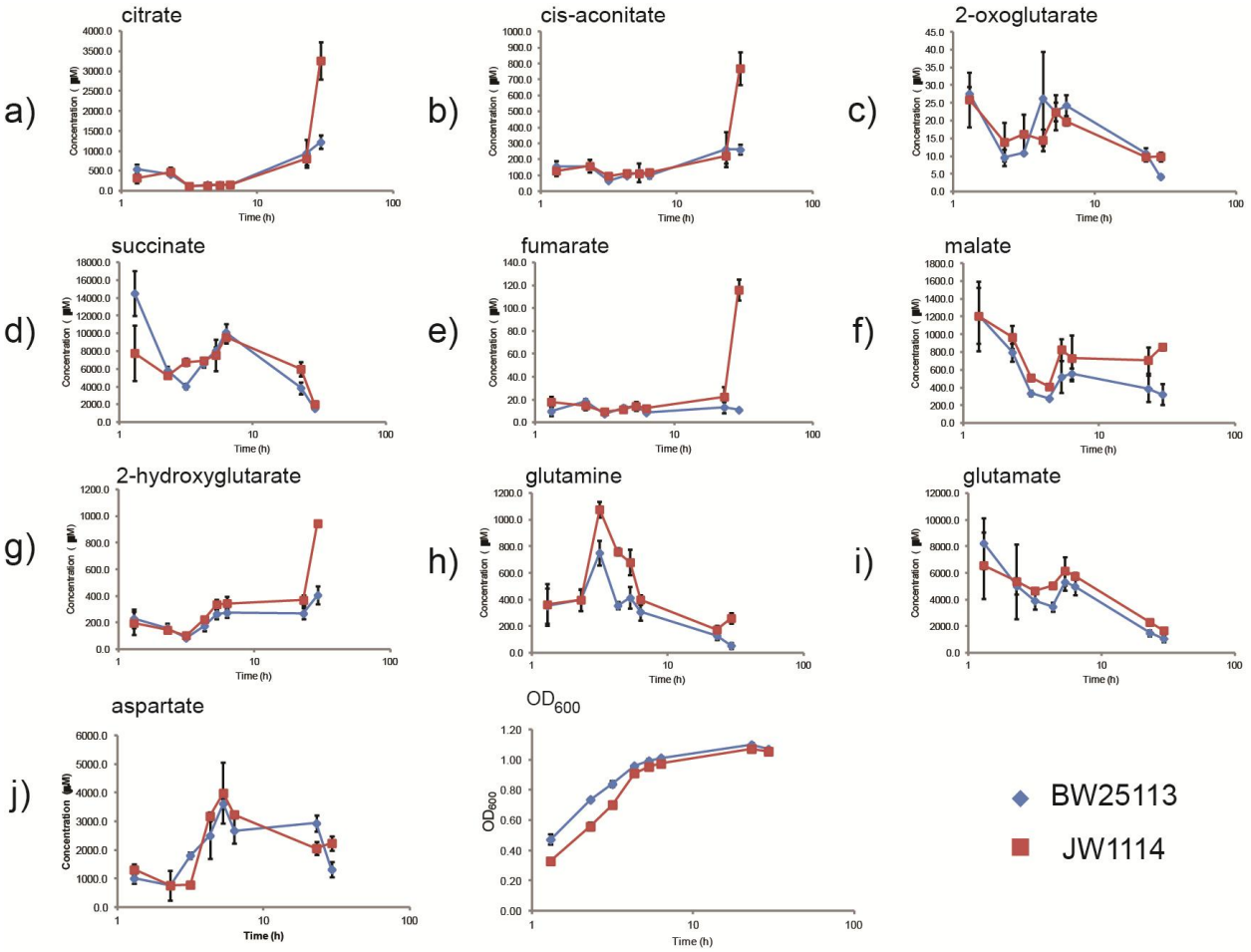


Figure 2.24: TCA cycle intermediate variation with time for *E. coli* WT and *ycfD* KO. BW25113 (blue diamonds) and JW1114 (red squares) were grown in 2TY medim. At indicated time points, an aliquot was removed, centrifuged anad the cell pellet frozen at -80°C. to lyse the cells, the cell pellet was resuspended in 80% MeOH/H_2O and samples sonicated in ice-bath for 10 min in 30 sec intervals. Cell debris was removed by centrifugation and the supernatant evaporated to yield a concentrated sample of metabolites, which were analysed by LC-MS using Quatro-ESI-MS method (LC-MS performed by Khalid al-Qahtani. A) Citrate, b) *cis*-aconitate, c) 2OG, d) succinate, e) fumarate, f) malate, g) 2-hydroxyglutarate (2-HG), h) glutamine, i) glutamate, j) aspartate. OD_{600} values of bacterial growth were measured as a function of time. Note that time is plotted on a log-scale. The data are preliminary.

2.8. Conclusions

Following from the identification of 50S ribosomal proteins as interacting partners of *E. coli* ycfD, and the demonstration that ycfD catalyzes the hydroxylation of L16 peptide fragment, the potential importance of oxygen-dependent modifications of ribosomal proteins became apparent. The need for a comprehensive analysis of *ycfD* effects on ribosomal proteins inspired the development of a UPLC-ESI-TOF-MS method (often referred to as whole-protein mass spectrometry, WPMS) by which the masses of individual ribosomal proteins, and to an extent their relative concentrations, could be measured.

In this chapter, it was demonstrated that

i) Out of 52 ribosomal proteins identified, only L16 is oxidatively modified by ycfD;
ii) The modification was located on R81, in agreement with the early reports on the primary structure of L16, as well as the first MALDI-MS study of ribosomal proteins, and the MS/MS data of ycfD-treated L16 fragment peptide;
iii) hydroxylation of L16 varies with pO_2, suggesting that ycfD has the potential to act as an oxygen sensor;
iv) *ycfD* deletion resulted in growth phenotype at low temperature, and in minimal medium;
v) *ycfD* deletion decreased the global rate of translation in minimal medium;
vi) *ycfD* deletion did not affect translational accuracy or ribosome assembly.

Collectively, the results indicate that ycfD catalyses a post-translational modification of a key ribosomal protein required for translation. The modification appears to be required for survival under extreme conditions of nutrient stress, and at low temperatures. YcfD has the capacity to act as an oxygen/nutrient sensor, which may act to fine-tune the rate of translation and potentially affect the growth rate. Given the living environment of *E. coli* which rapidly fluctuates both in oxygen concentration and nutrient availability, this modification may serve to fine-tune the rate of translation in response to fluctuations in nutrient levels.

2.9. References

1. Arnold, R. J., and Reilly, J. P. (1999) Observation of Escherichia coli ribosomal proteins and their posttranslational modifications by mass spectrometry. *Anal Biochem* **269**, 105-112
2. Running, W. E., Ravipaty, S., Karty, J. A., and Reilly, J. P. (2007) A top-down/bottom-up study of the ribosomal proteins of Caulobacter crescentus. *J Proteome Res* **6**, 337-347
3. Liu, X., and Reilly, J. P. (2009) Correlating the chemical modification of Escherichia coli ribosomal proteins with crystal structure data. *J Proteome Res* **8**, 4466-4478
4. Lauber, M. A., Running, W. E., and Reilly, J. P. (2009) B. subtilis Ribosomal Proteins: Structural Homology and Post-Translational Modifications. *J Proteome Res* **8**, 4193-4206
5. Strader, M. B., Verberkmoes, N. C., Tabb, D. L., Connelly, H. M., Barton, J. W., Bruce, B. D., Pelletier, D. A., Davison, B. H., Hettich, R. L., Larimer, F. W., and Hurst, G. B. (2004) Characterization of the 70S Ribosome from Rhodopseudomonas palustris using an integrated "top-down" and "bottom-up" mass spectrometric approach. *J Proteome Res* **3**, 965-978
6. Hardy, S. J., Kurland, C. G., Voynow, P., and Mora, G. (1969) The ribosomal proteins of Escherichia coli. I. Purification of the 30S ribosomal proteins. *Biochemistry* **8**, 2897-2905
7. Kerlavage, A. R., Cooperman, B. S., and Kahan, L. (1982) Reverse-Phase High-Performance Liquid-Chromatography of Ribosomal Small Subunit Proteins. *Fed Proc* **41**, 1385-1385
8. Kerlavage, A. R., Kahan, L., and Cooperman, B. S. (1982) Reverse-Phase High-Performance Liquid-Chromatography of Escherichia-Coli Ribosomal Small Subunit Proteins. *Anal Biochem* **123**, 342-348
9. Kerlavage, A. R., Hasan, T., and Cooperman, B. S. (1983) Reverse Phase High-Performance Liquid-Chromatography of Escherichia-Coli Ribosomal-Proteins - Standardization of 70-S, 50-S, and 30-S Protein Chromatograms - Functional-Activity of Purified Proteins. *J Biol Chem* **258**, 6313-6318
10. Kerlavage, A. R., Weitzmann, C. J., Hasan, T., and Cooperman, B. S. (1983) Reversed-Phase High-Performance Liquid-Chromatography of Escherichia-Coli Ribosomal-Proteins - Characteristics of the Separation of a Complex Protein Mixture. *J Chromatogr* **266**, 225-237
11. Kerlavage, A. R., Weitzmann, C. J., and Cooperman, B. S. (1984) Application of High-Performance Liquid-Chromatography to the Reconstitution of Ribosomal-Subunits. *J Chromatogr* **317**, 201-212
12. Nowotny, P., Eckardt, H., Nowotny, V., and Kamp, R. M. (1988) Preparation of Active Ribosomal-Proteins by Hplc. *Chromatographia* **25**, 409-412
13. Liu, X. H., and Reilly, J. P. (2009) Correlating the Chemical Modification of Escherichia coli Ribosomal Proteins with Crystal Structure Data. *J Proteome Res* **8**, 4466-4478
14. Nesterchuk, M. V., Sergiev, P. V., and Dontsova, O. A. (2011) Posttranslational Modifications of Ribosomal Proteins in Escherichia coli. *Acta Naturae* **3**, 22-33
15. Brosius, J., and Chen, R. (1976) The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett* **68**, 105-109
16. Hayter, J. R., Robertson, D. H. L., Gaskell, S. J., and Beynon, R. J. (2003) Proteome analysis of intact proteins in complex mixtures. *Mol Cell Proteomics* **2**, 85-95
17. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**
18. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M. C., and Leonhardt, H. (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* **7**, 282-289
19. Green, J., Crack, J. C., Thomson, A. J., and LeBrun, N. E. (2009) Bacterial sensors of oxygen. *Curr Opin Microbiol* **12**, 145-151

20. Kiley, P. J., and Beinert, H. (1998) Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster. *Fems Microbiol Rev* **22**, 341-352
21. Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., and Relman, D. A. (2005) Diversity of the human intestinal microbial flora. *Science* **308**, 1635-1638
22. Jones, S. A., Gibson, T., Maltby, R. C., Chowdhury, F. Z., Stewart, V., Cohen, P. S., and Conway, T. (2011) Anaerobic respiration of Escherichia coli in the mouse intestine. *Infect Immun* **79**, 4218-4226
23. Baylor, A. E., Diebel, L. N., Liberati, D. M., Dulchavsky, S. A., Diglio, C. A., and Brown, W. J. (2003) The effects of varying oxygen conditions and immunoglobulin A on barrier defense to bacterial invasion. *Am Surg* **69**, 231-237; discussion 237
24. Jones, S. A., Chowdhury, F. Z., Fabich, A. J., Anderson, A., Schreiner, D. M., House, A. L., Autieri, S. M., Leatham, M. P., Lins, J. J., Jorgensen, M., Cohen, P. S., and Conway, T. (2007) Respiration of Escherichia coli in the mouse intestine. *Infect Immun* **75**, 4891-4899
25. Nikolov, E. N., Dabeva, M. D., and Nikolov, T. K. (1983) Turnover of ribosomes in regenerating rat liver. *Int J Biochem* **15**, 1255-1260
26. Moore, V. G., Atchison, R. E., Thomas, G., Moran, M., and Noller, H. F. (1975) Identification of a ribosomal protein essential for peptidyl transferase activity. *Proc Natl Acad Sci USA* **72**, 844-848
27. Shoji, S., Dambacher, C. M., Shajani, Z., Williamson, J. R., and Schultz, P. G. (2011) Systematic Chromosomal Deletion of Bacterial Ribosomal Protein Genes. *J Mol Biol* **413**, 751-761
28. Ge, W., Wolf, A., Feng, T., Ho, C. H., Sekirnik, R., Zayer, A., Granatino, N., Cockman, M. E., Loenarz, C., Loik, N. D., Hardy, A. P., Claridge, T. D., Hamed, R. B., Chowdhury, R., Gong, L., Robinson, C. V., Trudgian, D. C., Jiang, M., Mackeen, M. M., McCullagh, J. S., Gordiyenko, Y., Thalhammer, A., Yamamoto, A., Yang, M., Liu-Yi, P., Zhang, Z., Schmidt-Zachmann, M., Kessler, B. M., Ratcliffe, P. J., Preston, G. M., Coleman, M. L., and Schofield, C. J. (2012) Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol* **8**, 960-962
29. Gordiyenko, Y., Videler, H., Zhou, M., McKay, A. R., Fucini, P., Biegel, E., Muller, V., and Robinson, C. V. (2010) Mass spectrometry defines the stoichiometry of ribosomal stalk complexes across the phylogenetic tree. *Mol Cell Proteomics* **9**, 1774-1783
30. Mohr, D., Wintermeyer, W., and Rodnina, M. V. (2002) GTPase activation of elongation factors Tu and G on the ribosome. *Biochemistry* **41**, 12520-12528
31. Diaconu, M., Kothe, U., Schlunzen, F., Fischer, N., Harms, J. M., Tonevitsky, A. G., Stark, H., Rodnina, M. V., and Wahl, M. C. (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**, 991-1004
32. Chang, F. N. (1978) Temperature-dependent variation in the extent of methylation of ribosomal proteins L7 and L12 in Escherichia coli. *J Bacteriol* **135**, 1165-1166
33. Lhoest, J., and Colson, C. (1977) Genetics of ribosomal protein methylation in Escherichia coli. II. A mutant lacking a new type of methylated amino acid, N5-methylglutamine, in protein L3. *Mol Gen Genet* **154**, 175-180
34. Poot, R. A., Jeeninga, R. E., Pleij, C. W. A., and vanDuin, J. (1997) Acetylation of ribosomal protein S5 affected by defects in the central pseudoknot in 16S ribosomal RNA? *FEBS Lett* **401**, 175-179
35. Roy-Chaudhuri, B., Kirthi, N., Kelley, T., and Culver, G. M. (2008) Suppression of a cold-sensitive mutation in ribosomal protein S5 reveals a role for RimJ in ribosome biogenesis. *Mol Microbiol* **68**, 1547-1559
36. Mora, L., Heurgue-Hamard, V., de Zamaroczy, M., Kervestin, S., and Buckingham, R. H. (2007) Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. *J Biol Chem* **282**, 35638-35645

37. Lancaster, L., Lambert, N. J., Maklan, E. J., Horan, L. H., and Noller, H. F. (2008) The sarcin-ricin loop of 23S rRNA is essential for assembly of the functional core of the 50S ribosomal subunit. *RNA* **14**, 1999-2012
38. Siibak, T., Peil, L., Donhofer, A., Tats, A., Remm, M., Wilson, D. N., Tenson, T., and Remme, J. (2011) Antibiotic-induced ribosomal assembly defects result from changes in the synthesis of ribosomal proteins. *Mol Microbiol* **80**, 54-67
39. Holberger, L. E., and Hayes, C. S. (2009) Ribosomal protein S12 and aminoglycoside antibiotics modulate A-site mRNA cleavage and transfer-messenger RNA activity in Escherichia coli. *J Biol Chem* **284**, 32188-32200
40. Russell, J. B., and Cook, G. M. (1995) Energetics of Bacterial-Growth - Balance of Anabolic and Catabolic Reactions. *Microbiol Rev* **59**, 48-62
41. Klumpp, S., Scott, M., Pedersen, S., and Hwa, T. (2013) Molecular crowding limits translation and cell growth. *P Natl Acad Sci USA* **110**, 16754-16759
42. Dennis, P. P., Ehrenberg, M., and Bremer, H. (2004) Control of rRNA synthesis in Escherichia coli: a systems biology approach. *Microbiol Mol Biol R* **68**, 639-+
43. Adam, J., Yang, M., Soga, T., and Pollard, P. J. (2013) Rare insights into cancer biology. *Oncogene*

3. *YcfD* mediates resistance to peptidyl transferase ribosomal antibiotics

Contents

3.1. YcfD and ribosomal antibiotics 97
3.1.1. Aims 98
3.2. The effect of *ycfD* deletion on antibiotic sensitivity 99
3.2.1. Sensitivity to antibiotics affecting the translation mechanism 100
3.2.2. Effect of chloramphenicol analogues on growth of *ycfD*-deleted strain 103
3.3. Increasing antibiotic sensitivity of BW25113 by inhibiting ycfD 104
3.3.1. Iron chelator DFO increases susceptibility of BW25113 to chloramphenicol 105
3.3.2. Co(II) treatment increases the susceptibility of BW25113 to chloramphenicol 105
3.3.3. Chloramphenicol sensitivity of BW25113 increases in hypoxia 106
3.4. Isogenic model system of *ycfD* activity 108
3.5. YcfD inhibitor screening 110
3.5.1. pH optimisation for ycfD activity 111
3.5.2. Optimising kinetic parameters for ycfD activity screening 112
3.5.3. *In vitro* MALDI-MS Inhibitor Screen 113
3.5.4. Differential scanning fluorimetry (DSF) screen of ycfD stabilisers 115
3.5.5. Determination of IC_{50} values for IOX1 analogues against ycfD 118
3.5.6. IOX1 treatment increases the antibiotic potency of chloramphenicol 120
3.6. Conclusions 123
3.7. References 126

3.1. *YcfD* and ribosomal antibiotics

The complex multi-unit structure of the ribosome suggests it could be regarded as a molecular machine which requires proper arrangement of all its component parts for functioning. A small spatial disruption in the optimal arrangement on the ribosome can result in loss of function. Ribosomal antibiotics often comprise natural products that have evolved to disrupt the catalytic cycle of the ribosome. They frequently achieve antibiotic effects by inserting into key binding sites for ribosomal components, preventing successful continuation of the protein synthesis cycle, akin to poking a screwdriver into the computer or throwing a spanner in the cogs of an engine (1). The binding sites of ribosomal antibiotics are highly clustered in the proximity of key functional sites of the ribosome (2), such as the peptidyl transferase centre (PTC), which is a target of a number of ribosomal antibiotics including chloramphenicol, clindamycin and erythromycin (3) (Figure 3.1).

Post-translational modifications (PTMs) to the ribosome, including of rRNA and ribosomal proteins, likewise cluster around functionally important sites (4) and it cannot be excluded that PTMs, at least in part, mediate resistance to antibiotics with adjacent binding sites. Methylation of 23S rRNA (nucleotide A2058 in *E. coli*) increases resistance to macrolides with proximal binding sites, including erythromycin (5,6); hydroxy-lysinylation of ribosome associated factor EF-P has been reported to increase resistance to gentamicin (7). The effects of ribosomal protein modifications on antibiotic resistance have not yet been described.

In Chapter 2 it was demonstrated that *E. coli* ycfD catalyses hydroxylation of R81 on ribosomal protein L16, in the proximity of the PTC on the ribosome, which affects global translation rate and survival under limited nutrient conditions and cold shock. It was therefore envisaged that L16 hydroxylation may also affect susceptibility/resistance to ribosomal antibiotics. Such an oxygenase-mediated antibiotic resistance mechanism could be exploited therapeutically in one of two ways: if oxygenase activity increases resistance to antibiotics, a small-molecule inhibitor could be designed to inhibit the action of the oxygenase, thereby increasing sensitivity to the antibiotic under normoxic conditions; oxygenase-dependent sensitivity to ribosomal antibiotics could be exploited therapeutically in hypoxic/anoxic tissues, such as solid tumours, where ribosomal proteins may be unhydroxylated.

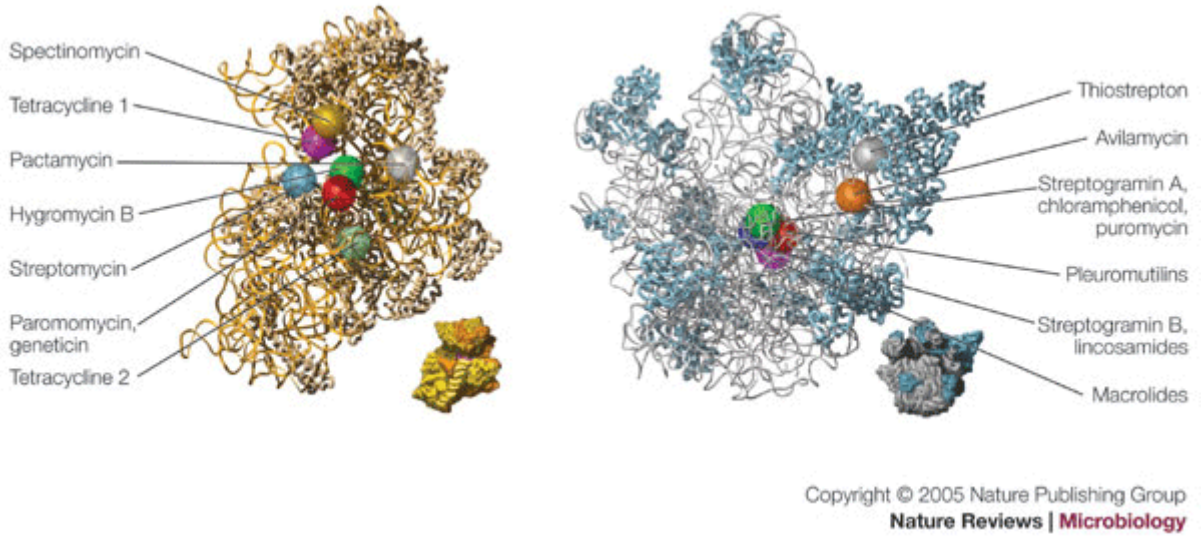


Figure 3.1: Ribosomal antibiotic binding sites cluster around functional regions of the ribosome. A view of the 30S (left) and the 50S (right) ribosomal subunits with binding sites of antibiotics indicated. In the 30S subunit, antibiotic binding sites are clustered in proximity of the decoding centre (hygromycin B, streptomycin, paromomycin, geneticin). 50S antibiotic binding sites are clustered around the peptidyl transferase centre (PTC; streptogramin, chloramphenicol, puromycin, pleromutilins, streptogramin B, lincosamides and macrolides). Reproduced with permission from (2).

3.1.1. Aims

This chapter describes work on the effects of *ycfD* deletion on susceptibility to antibiotics. A commercial library of antibiotics was screened against WT (BW25113) and KO (JW1114) strains; initial commercial screens were followed up with in-house validation of hits; antibiotic analogues were tested to exclude random effects. The ycfD-dependent sensitisation to antibiotics was then demonstrated by removing necessary co-factors for enzymatic activity (Fe(II) and O_2). A ycfD inhibitor was identified using a MALDI-MS based activity assay and a differential scanning fluorimetry (DSF)-based protein stabilisation assay. The most potent inhibitor was advanced to studies in cells to investigate whether antibiotic susceptibility could be enhanced by an inhibitor of ycfD in cells and its action was correlated with analysis of L16 hydroxylation in cells.

3.2. The effect of *ycfD* deletion on antibiotic sensitivity

Initially, a panel of antibiotics was tested for their effects on cellular metabolism of *ycfD* WT and KO strains using a commercial screening service (BioLog, conducted by the Animal Health and Veterinary Laboratories Agency, Weybridge, UK). The metabolic activity of two strains was measured by a tetrazolium reporter dye, as described in Chapter 2. Out of 24 tested antibiotics, which included ribosomal antibiotics binding to the 30S subunit (amikacin, neomycin, kanamycin, gentamicin, chlortetracycline, minocycline and capreomycin), the 50S subunit (lincomycin, chloramphenicol, erythromycin, demeclocycline), β-lactam antibiotics (amoxicillin, cloxacillin, nafcillin), cephalosporins (cefazolin, ceftriaxone, cephalothin), fluoroquinolones (lomefloxacin, enoxacin, ofloxacin), DNA strand break-inducing agents (bleomycin), a polymyxin (colistin) and DNA replication inhibitors (nalidixic acid), only four were found to reproducibly lead to sensitisation of ycfD KO strain: lincomycin, demeclocycline, erythromycin and chloramphenicol (Figure 3.2).

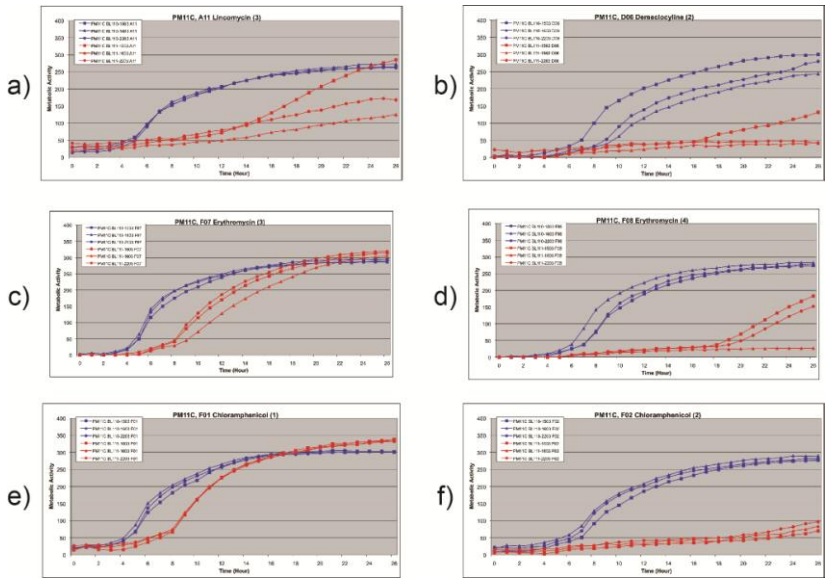


Figure 3.2: Selected metabolic activity curves of BW25113 (blue) and JW1114 (red) in the presence of antibiotics.
a) lincomycin, b) demeclocycline, c-d) erythromycin, e-f) chloramphenicol. Concentrations of antibiotics used by Biolog were not disclosed, however increasing concentrations resulted in an increased differentiation in metabolism (d/f).

Strikingly, all the four antibiotics that inhibited growth of ycfD KO, but not WT, were ribosomal antibiotics targeting the 50S subunit. Analysis of the 50S binding sites of chloramphenicol (8,9), erythromycin (8), and the lincomycin analogue clindamycin (8), for which crystal structures are

available, revealed that they bind in the immediate proximity of R81 on L16 (Figure 3.3), potentially suggesting a different interaction with unhydroxylated, compared to hydroxylated ribosomes.

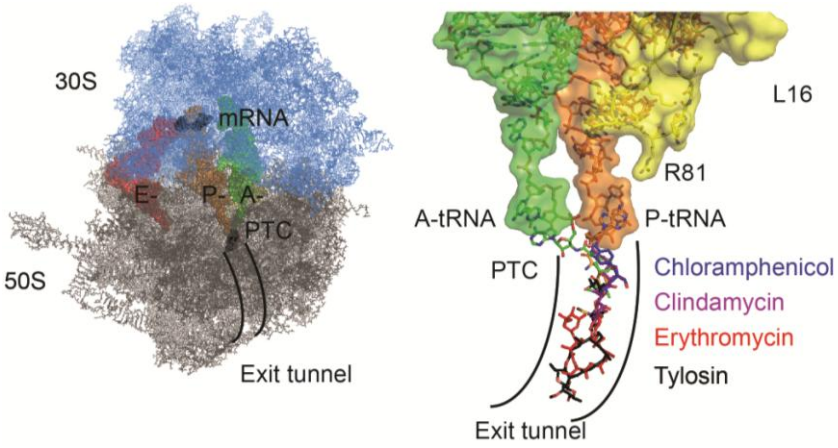


Figure 3.3: Analysis of the 50S binding sites of chloramphenicol, clindamycin and erythromycin, all of which bind at the PTC in the proximity of the ycfD hydroxylation substrate R81 on L16. The clindamycin propyl pyrrolidinyl moiety interferes with A-site aminoacyl-tRNA binding (8), and also overlaps with the nitrobenzene ring of chloramphenicol. This figure was compiled using PDB entries 3OFC,3OFZ, 3OFR, 1K9M, 2WDK, 2WDL (8-10).

3.2.1. Sensitivity to antibiotics affecting the translation mechanism

In order to validate this finding, a range of commercially available antibiotics was selected for further screening against the WT and KO strains. Antibiotics targeting different steps of the translation mechanism were selected, including streptomycin and paromomycin (30S decoding centre inhibitors (11)), minocycline and doxycycline (tetracyclines), spectinomycin (EF-G binding translocation inhibitor), hygromycin (translocation inhibitor), puromycin (inhibitor of the 50S A-site), chloramphenicol, anisomycin, clindamycin, lincomycin and erythromycin (peptidyl transferase inhibitors) and fusidic acid (EF-G.GPD stabilising) (1).

Growth differences were observed for JW1114 grown in presence of chloramphenicol (5 μg/mL) and lincomycin (400 μg/mL) (Figure 3.4). Paromomycin treatment (50 μg/mL) led to decreased growth of BW25114, but not JW1114. This observation can be rationalised by the knowledge that JW1114 contains a kanamycin resistance cassette, which confers resistance to other aminoglycosides, including paromomycin. Other antibiotics did not differentially affect growth of the two strains (Figure 3.4).

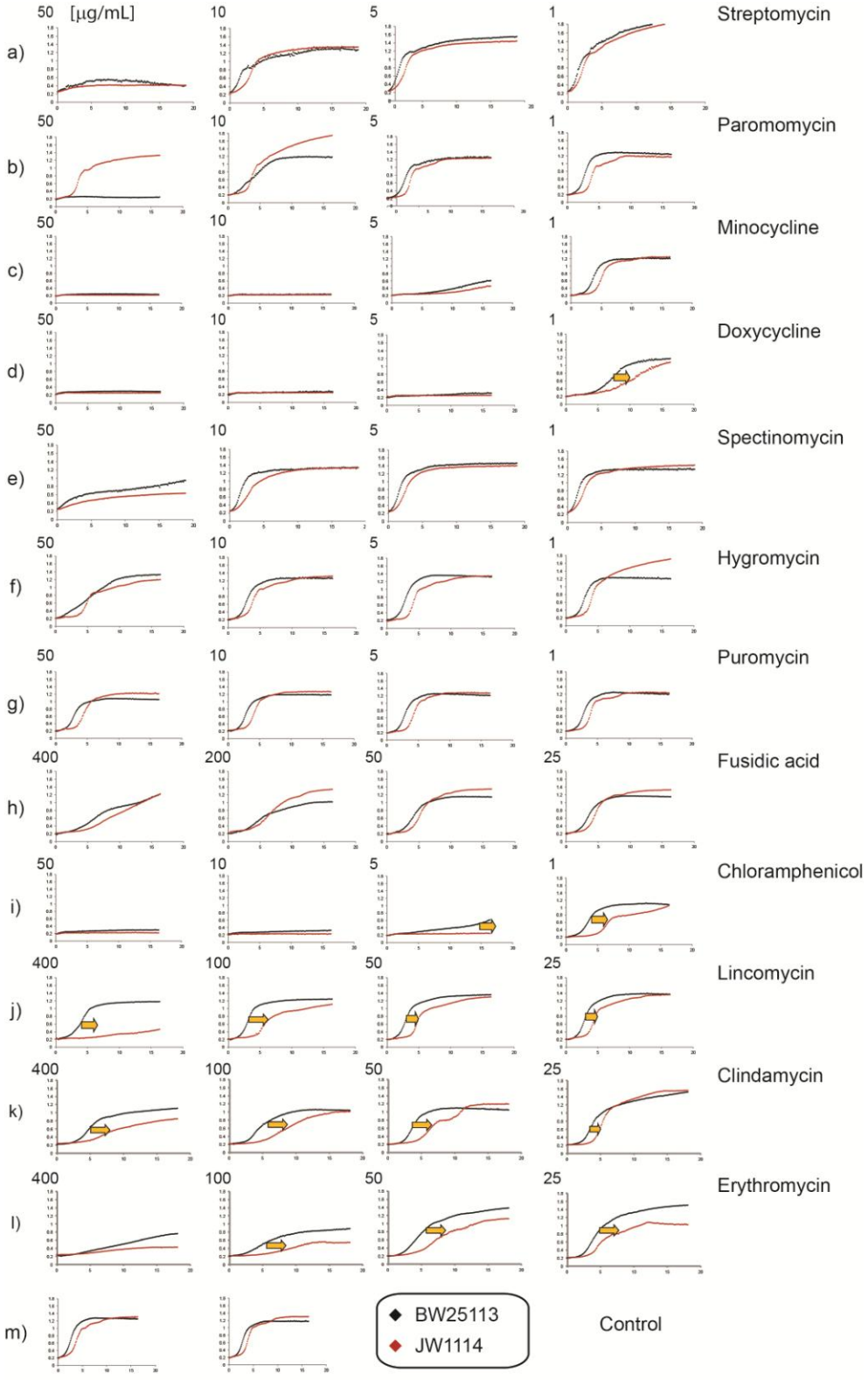


Figure 3.4: BW25113 and JW1114 growth is selectively inhibited by 50S antibiotics which bind at the peptidyl transferase centre (PTC). BW25113 and JW1114 were grown in 2TY medium (200 µL) at 37°C in the presence of serial dilution of selected antibiotics. OD_{600} was measured spectrophotometrically for 16 h after inoculation with BW25113 and JW1114 cultures diluted to OD_{600} 0.05 in 2TY medium.

The dose-dependent selective inhibition of JW1114 growth by chloramphenicol and lincomycin was reproducible. Chloramphenicol was tested at 6 concentrations, leading to a dose-dependent growth delay of JW1114 relative to BW25113: a lag of 1 h between BW25113 and JW1114 was observed, which increased to 2.5 h at 1 μg/mL chloramphenicol and 7 h at 3 μg/mL; additionally maximum optical density decreased from 1.2 to 1.0 a.u. At 5 μg/mL only BW25113 growth was observed (Figure 3.5A). Lincomycin treatment achieved a dose-dependent growth delay of JW1114: growth delay of JW1114 relative to BW25113 was 2 h at 100 μg/mL, 4.2 h at 200 μg/mL and 6.3 h at 300 μg/mL (Figure 3.5B).

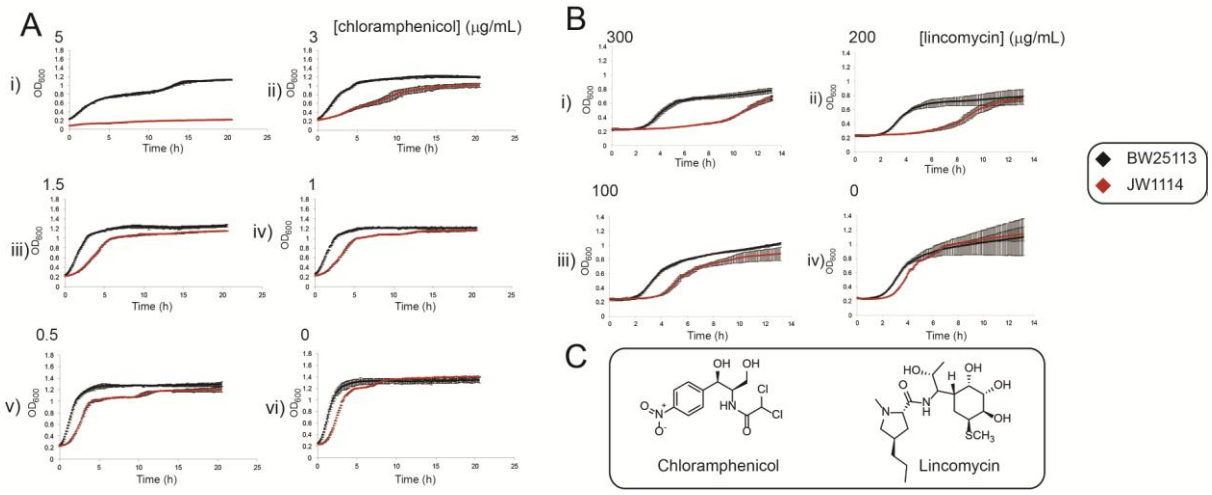


Figure 3.5: Growth of BW25113 and JW1114 in 2TY medium (200 μL) at 37°C in the presence of serial dilution of A) chloramphenicol and B) lincoymcin. OD_{600} was measured spectrophotometrically for 20 h and 13 h (chloramphenicol and lincomycin, respectively) after inoculation with BW25113 and JW1114 cultures diluted to OD_{600} 0.05 in 2TY medium. Each experiment was conducted in triplicate; error bars denote standard deviation from the mean.

It is known that chloramphenicol binding depends on the presence of the 50S ribosomal protein L16 (12). However, whilst L16 interacts with the 3' end of tRNA so as to fix the aminoacyl stem of the tRNA in the A-site (13), chloramphenicol interferes with the 3' end of aminoacyl tRNA binding. The results presented above suggest that chloramphenicol binding is enhanced in the absence of L16 hydroxylation, and that chloramphenicol and L16 R81 hydroxylation may have opposing effects on the tRNA accommodation in the A-site. This observation is consistent with a recent report which described the role of *E. coli* R81 on L16 as facilitating the translocation step of tRNA from A-site to P-site (14).

Lincomycin, which also showed dose-dependent selective inhibition of JW1114, is a close chemical analogue of clindamycin (they are equivalent apart from replacement of a hydroxyethyl moiety in lincomycin with a chloroethyl moiety in clindamycin). Consistent with their similar biological activity against the *ycfD* KO, the bound structures of chloramphenicol and clindamycin (and likely lincomycin) superimpose: the chloramphenicol nitrobenzene ring and the pyrrolidinyl propyl group of clindamycin overlap (9). Furthermore, the chlorines of chloramphenicol and clindamycin (replaced with a hydroxyl group in lincomycin) overlap (9). The clindamycin propyl pyrrolidinyl moiety interferes with A-site tRNA positioning in *E. coli* (8).

3.2.2. Effect of chloramphenicol analogues on growth of *ycfD*-deleted strain

To further support the hypothesis that PTC inhibitors cause differential growth inhibition, chloramphenicol analogues thiamphenicol and monoiodoamphenicol were tested. Thiamphenicol acts analogously to chloramphenicol, stabilising binding of fMet-tRNA^{fMet} to the P-site (15). The chloramphenicol analogue monoiodoamphenicol irreversibly modifies L16, but its affinity for binding to the ribosome is ten-fold lower than for chloramphenicol (16). Monoiodoamphenicol shows different modes of ribosome inhibition in *E. coli* and *Bacillus stearothermophilus*; *E. coli* ribosomes are irreversibly inhibited by monoiodoamphenicol, whereas *B. stearothermophilus* ribosomes are inhibited reversibly (16). Intriguingly, no ycfD homologue is found in the *B. stearothermophilus* genome, suggesting that L16 in *B. stearothermophilus* is not hydroxylated. Monoiodoamphenicol shows high selectivity for irreversible modification of *E. coli* L16 (17). *In vivo*, it cross-links to S6, L16, and L24 in total protein extracts; *in vitro*, it only cross-links to L16. However, L16 is not observed in ribosome extracts after monoiodoamphenicol treatment, consistent with the disruption of ribosome biogenesis by chloramphenicol, which was reported to lead to a decrease in L16 incorporation (17,18). Substitution of the chloramphenicol nitro group with hydroxylamine, nitroso, hydroxamic acid, methyl hydroxamate and *O*-acetyl hydroxamate groups leads to loss of biological activity against *E. coli* and other bacteria (19), therefore these analogues were not tested in the BW25113/JW1113 growth assay.

The chloramphenicol analogues monoiodoamphenicol and thiamphenicol were tested for selective inhibition of JW1114 growth (Figure 3.6). At 10 μg/mL thiamphenicol the mid-log phase of JW1114 was delayed by 2.6 h relative to BW25113, at 25 μg/mL it was delayed by 6.6 h, at 50 μg/mL no growth of JW1114 was observed in the presence of either antibiotic, whilst BW25113 was still viable (Figure 3.6). The 10-fold increase in minimum inhibitory concentration (MIC) required to achieve complete inhibition of JW1114 growth was consistent with the reported 10-fold decrease in affinity of binding to the ribosome compared to chloramphenicol (16). Collectively, the results demonstrated that peptidyl transferase inhibitors which bind in the immediate proximity of L16 R81 differentially inhibit the growth of JW1114 over BW25113.

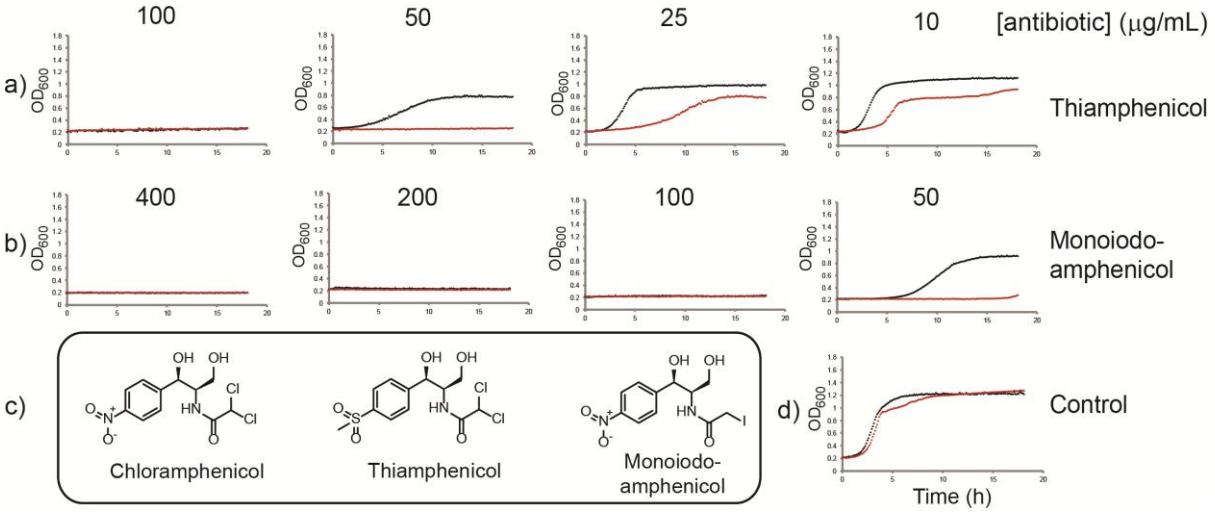


Figure 3.6: OD_{600} measurements of BW25113 (black) and JW1114 (red) growth rates in the presence of chloramphenicol analogues. a) thiamphenicol; b) monoiodoamphenicol; c) chemical structures of chloramphenicol analogues demonstrate that replacement of the nitro group by a sulfonyl group, and the dichloro moiety by a monoiodo moiety leads to a 10-fold increase in MIC for inhibition of JW1114 growth; d) control growth curves for BW25113 and JW1114. Growth of BW25113 and JW1114 was conducted in 2TY medium (200 μL) at 37°C in the presence of serial dilution of antibiotics. OD_{600} was measured spectrophotometrically for 18 h after inoculation with BW25113 and JW1114 cultures diluted to OD_{600} 0.05 in 2TY medium.

3.3. Increasing antibiotic sensitivity of BW25113 by inhibiting ycfD

YcfD is Fe(II), 2OG, and O_2 dependent (Chapter 2). In order to demonstrate the dependence of ycfD mediated antibiotic resistance on Fe(II) and O_2, the sensitivity of BW25113 to PTC inhibitors was measured in the presence of an iron-chelator desferrioxamine (DFO) and under anaerobic growth. The antibiotic with the lowest MIC required to achieve differential growth inhibition of JW1114, chloramphenicol, was chosen for these studies.

3.3.1. Iron chelator DFO increases susceptibility of BW25113 to chloramphenicol

First, removal of Fe(II) was attempted by treatment of BW25113 with iron chelator DFO; sensitisation to chloramphenicol was measured using the growth assay described above. Consistent with a Fe(II)-dependent 2OG-oxygenase-mediated chloramphenicol sensitisation mechanism, chelating Fe(II) from growth medium with 1 mM DFO resulted in a dose-dependent retardation of BW25113 growth in the presence of chloramphenicol (Figure 3.7). At 1 μg/mL chloramphenicol, 1 mM DFO resulted in 1 h lag in reaching mid-log phase relative to control treated with chloramphenicol alone; at 2 mg/mL chloramphenicol the DFO-treated sample did not enter log-phase and maximum optical density decreased from 0.6 to 0.4 (Figure 3.7).

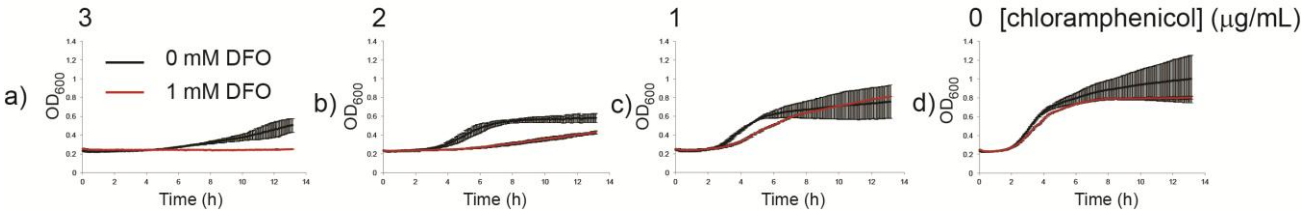


Figure 3.7: OD_{600} measurements of BW25113 treated with 1 mM DFO (red) or equivalent volume of water (black) grown in the presence of chloramphenicol at a) 3 μg/mL, b) 2 μg/mL, c) 1 μg/mL or d) no chlorampenicol. Growth of BW25113 was conducted in 2TY medium (200 μL) at 37°C. OD_{600} was measured spectrophotometrically for 13 h after inoculation with BW25113 culture diluted to OD_{600} 0.05 in 2TY medium. Each experiment was conducted in triplicate, with error bars denoting standard deviation of the mean.

3.3.2. Co(II) treatment increases the susceptibility of BW25113 to chloramphenicol

Treatment of BW25113 with 1 mM Co(II) and a serial dilution of chloramphenicol resulted in growth delay approximately equivalent to JW1114 treated with chloramphenicol alone. At 1.5 μg/mL chloramphenicol, the growth profiles of JW1114 treated with chloramphenicol alone and BW25113 treated with chloramphenicol and Co(II) overlapped. Below 1.5 μg/mL chloramphenicol, BW25113 treated with Co(II) showed a decreased lag in reaching mid-log phase compared to JW1114. Analysis of L16 hydroxylation revealed that at 1 mM Co(II), L16 hydroxylation was decreased to ~60% (Figure 3.8B), suggesting that the observed decrease in growth rate of Co(II)-treated BW25113 is mediated, at least in part, by L16 hydroxylation.

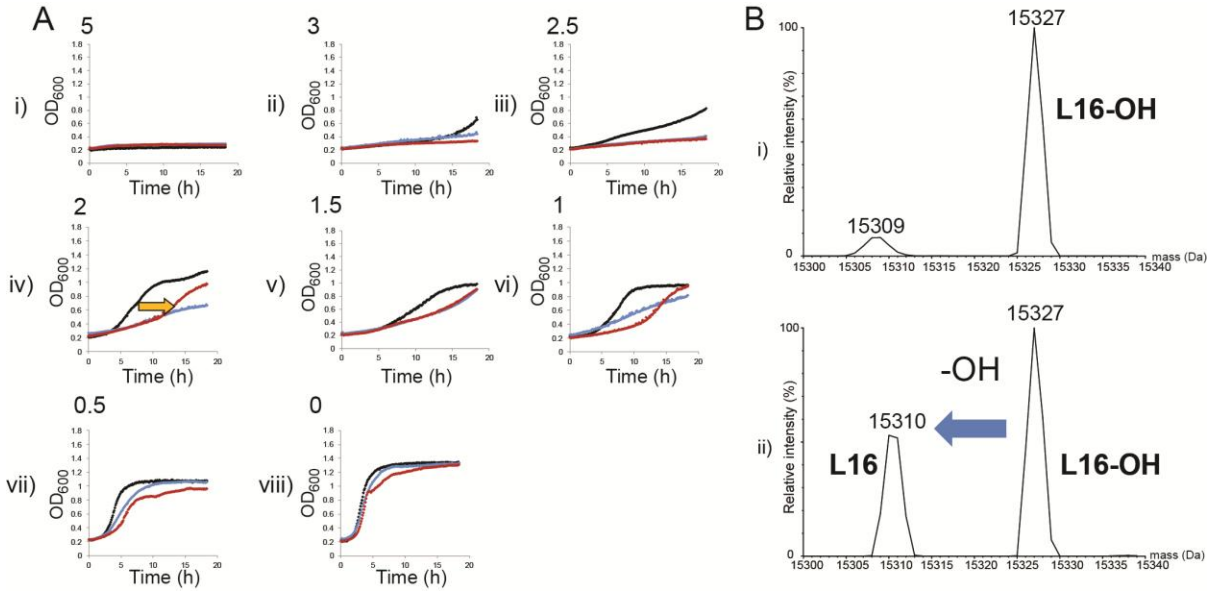


Figure 3.8: Co(II) treatment leads to decreased growth rate of BW25113 in the presence of chloramphenicol, mediated, in part, by inhibition of L16 hydroxylation. A) OD_{600} measurements of BW25113 (black), BW25113 treated with 1 mM Co(II) (blue) and JW1114 (red) grown in the presence of chloramphenicol at i) 5 μg/mL, ii) 3 μg/mL, iii) 2.5 μg/mL, iv) 2 μg/mL, v) 1.5 μg/mL, vi) 1 μg/mL, vii) 0.5 μg/mL, or viii) no chlorampenicol. Growth of BW25113 and JW1114 was conducted in 2TY medium (200 μL) at 37°C. OD_{600} was measured spectrophotometrically for 18 h after inoculation with BW25113 and JW1114 cultures diluted to OD_{600} 0.05 in 2TY medium. B) Deconvoluted ESI-MS spectra of ribosomal protein L16 from BW25113 treated with i) water or ii) 1 mM Co(II). A starter culture of BW25113 was sub-diluted 1:100 into 0.6 L of 2TY media containing 1 mM Co(II) and the bacteria were grown at 37°C overnight. Ribosomes were isolated and the whole-protein MS conducted as detailed in Materials and Methods.

3.3.3. Chloramphenicol sensitivity of BW25113 increases in hypoxia

Next, the effect of O_2 on chloramphenicol sensitivity was studied. Due to ease of handling the bacterial cultures, the growth assay was translated from liquid culture to an agar-plate growth assay; bacterial colonies of BW25113/JW1114 were grown on chloramphenicol-infused agar plates in 10-fold serial dilutions from starting cultures of equal optical density (0.05). Under aerobic growth, no growth difference was observed in the absence of chloramphenicol. At 1.5 μg/mL growth difference was observed at 10^5-fold dilution, which resulted in viable colonies of BW25113, but not JW1114 (Figure 3.9); at 3 μg/mL, no growth was observed for JW1114, whereas growth of BW25113 was observed up to 10^4-fold dilution. At 5 μg/mL, no growth was observed for either strain (Figure 3.9).

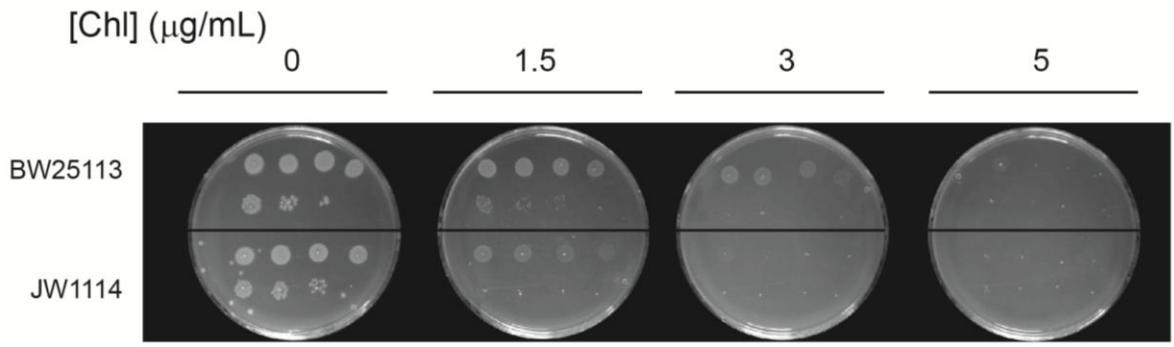


Figure 3.9: Differential sensitivity to chloramphenicol in aerobic conditions is observed on chloramphenicol-infused agar plates. BW25113 and JW1114 cultures were diluted to OD_{600} 0.05 in 2TY medium, then a serial 10-fold dilution was spotted on chloramphenicol-infused agar plates (5 μL per spot). LB-agar containing 0, 1.5, 3 or 5 μg/mL chloramphenicol was used. Plates were incubated overnight at 37°C. The plates were photographed using the UVI-Pro Chemi visualisation chamber and software.

Chloramphenicol sensitivity was then assessed under hypoxic growth conditions. As demonstrated in Chapter 2, L16 hydroxylation is ablated in *E. coli* grown under hypoxic conditions (0.1% O_2). Serial 10-fold dilutions of BW25113 and JW1114 cultures were grown on agar plates infused with 0-5 mg/mL chloramphenicol; a control plate without chloramphenicol was grown at 21% O_2 (Figure 3.10). In order to avoid growth retardation in hypoxia, growth medium was supplemented with 2% glucose, which enabled bacteria to grow at comparable rates in normoxia and hypoxia (no growth of either strain was observed in the absence of glucose; data not shown). No growth difference was observed at 21% and 0.1% O_2 in the absence of chloramphenicol (Figure 3.10); growth of both strains was inhibited by 1 μg/mL chloramphenicol, suggesting that chloramphenicol sensitivity of both strains had increased. This observation is consistent with an oxygenase-mediated effect of chloramphenicol sensitivity. The experiment was repeated for the concentration range 0-1 μg/mL chloramphenicol (Figure 3.10B). At 1 μg/mL growth of BW25113 was observed to 10^3-fold dilution, and JW1114 to 10^2-fold dilution; therefore the growth difference has been reduced. At 0.5 μg/mL the number of colonies was approximately equivalent.

Overall, growth in anaerobic conditions demonstrated that the MIC of chloramphenicol leading to differential growth has been reduced from 3 mg/mL to <1 mg/mL, and that the differential growth effect has decreased from 10^3-fold (observed at 1.5 mg/mL in normoxia) to ~10^1-fold in anoxia, further corroborating an oxygenase-mediated effect.

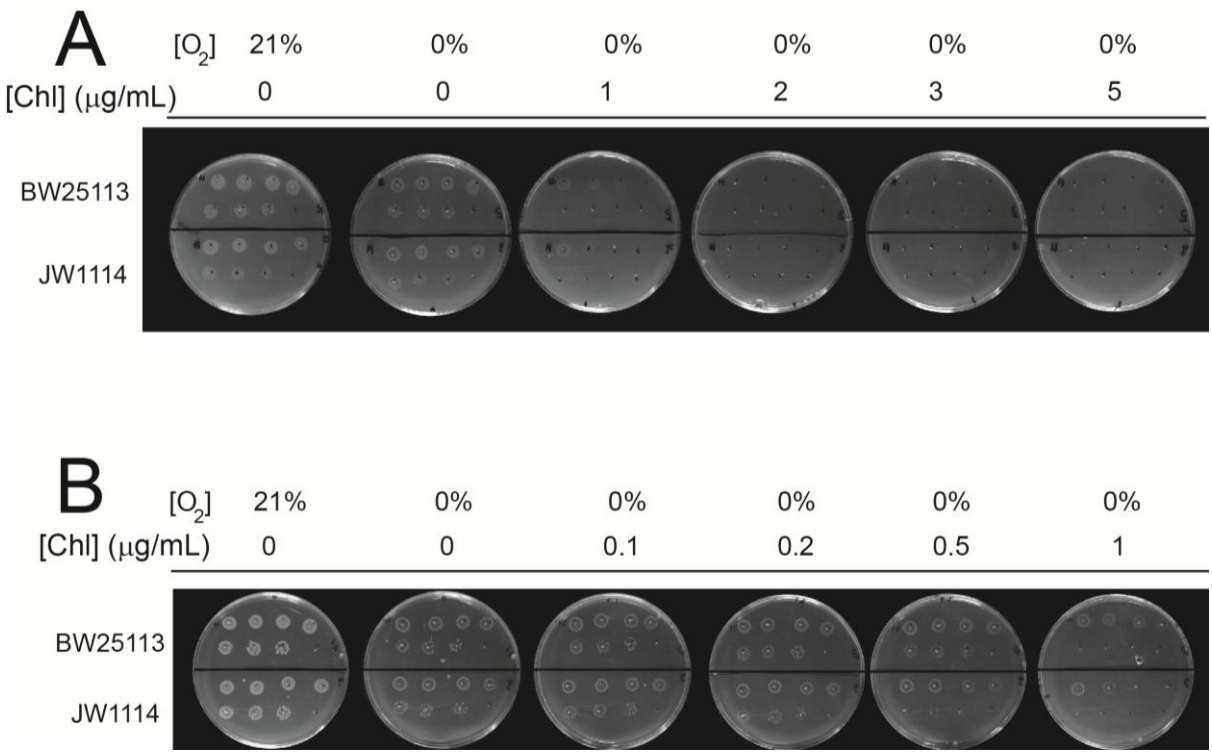


Figure 3.10: Growth of BW25113 and JW1114 in the presence of chloramphenicol in hypoxia (0.1% O_2). BW25113 and JW1114 cultures were diluted to OD_{600} 0.05 in 2TY medium, then a serial 10-fold dilution was spotted on chloramphenicol-infused agar plates (5 μL per spot). LB-agar plates containing A) 0, 1, 2, 3 or 5 μg/mL chloramphenicol or B) 0, 0.1, 0.2, 0.5 and 1 μg/mL chloramphenicol were used. Plates were incubated overnight at 37°C in a benchtop incubator (21% O_2) or InVivo hypoxia work-station equilibrated at 0.1% O_2, 37°C with an active humidistat to avoid water evaporation. The plates were photographed using the UVI-Pro Chemi visualisation chamber and software.

3.4. Isogenic model system of ycfD activity

In an attempt to demonstrate the dependence of chloramphenicol sensitivity on ycfD, an isogenic model was created by reconstituting the KO strain with empty vector (JW1114_EV), wild-type ycfD (JW1114_WT) or its catalytically inactive variant H125A_D127A (JW1114_AxA), in a pRSet5D vector (20) backbone. However, two potentially confounding factors limited the utility of this experiment. Despite equal optical densities of starter cultures used to induce colony growth, faster growth of JW1114_EV was observed relative to JW1114_WT and JW1114_AxA strains (Figure 3.11). As demonstrated by the co-immunoprecipitation experiment discussed in the Introduction, ycfD interacts strongly with the ribosomal protein L16, which is crucial for the peptidyl transferase activity of the ribosome (21); over-expression of ycfD may sequester L16 from the ribosome, thereby slowing the growth rate of over-expressing bacteria.

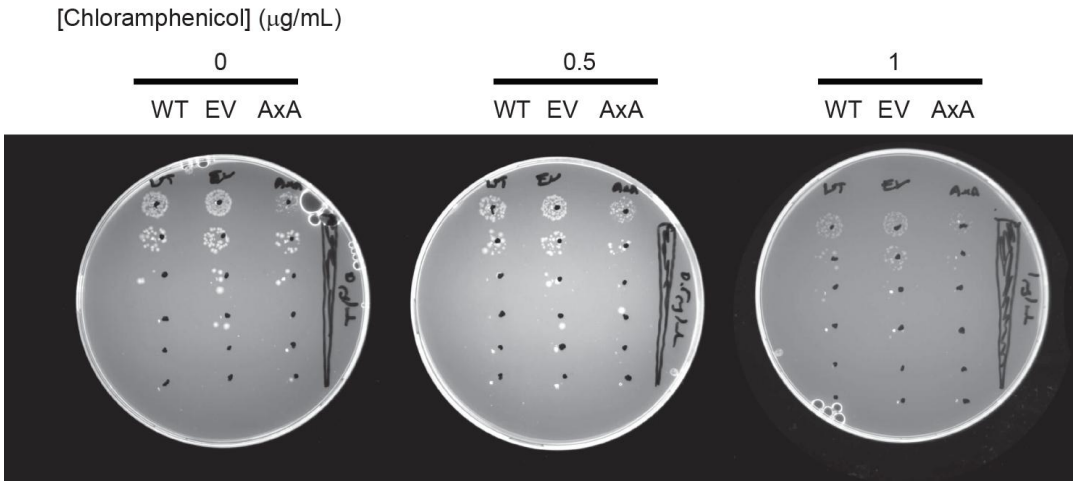


Figure 3.11: Growth of JW1114 reconstituted with empty vector (EV), wild-type ycfD (WT) or its catalytically inactive variant H125A_D127A (AxA) grown on chloramphenicol (0, 0.5 and 1.0 μg/mL)-infused LB agar plates in normoxia at 37°C. JW1114_EV/WT/AxA cultures were diluted to OD_{600} 0.05 in 2TY medium, then a serial 10-fold dilution was spotted on chloramphenicol-infused agar plates (5 μL per spot). Growth of JW1114_EV was observed at higher dilutions of 10^4-fold. LB-agar plates containing 0, 0.5, or 1 μg/mL chloramphenicol were used. Plates were incubated overnight at 37°C in a benchtop incubator (21% O_2). The plates were photographed using the UVI-Pro Chemi visualisation chamber and software.

Alternatively, differences in basal growth rate may stem from over-expression of the green fluorescent protein (GFP) tag, the production of which may place significant energetic demands on the replicating bacteria. In order to exclude the expression of GFP as a confounding factor, ycfD WT was expressed from an alternative vector (pTRC99a), which enables protein expression without an epitope tag (Figure 3.12) (22).

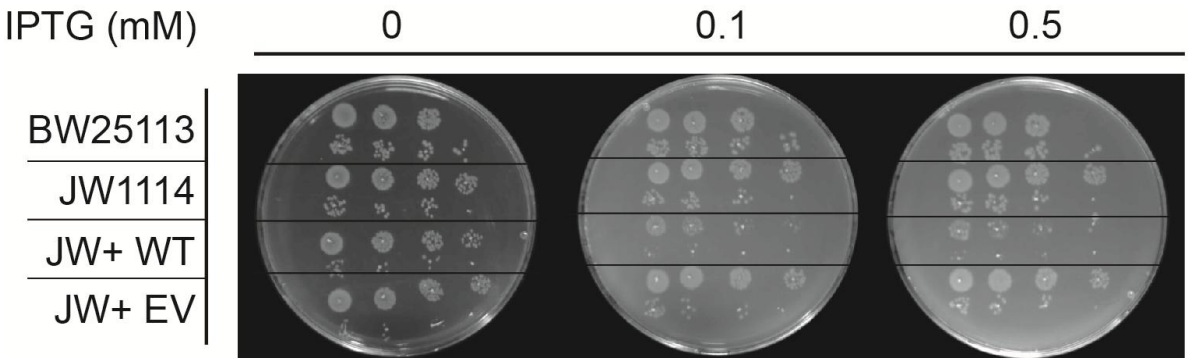


Figure 3.12: Growth of BW25113, JW1114, and JW1114 reconstituted with wild-type ycfD (WT) or empty vector (EV), grown in the presence of IPTG (0, 0.1 and 0.5 mM) reveals toxicity of ycfD expression. Cultures were diluted to OD_{600} 0.05 in 2TY medium, then a serial 10-fold dilution was spotted on IPTG-infused agar plates (5 μL per spot). Plates were incubated overnight at 37°C in a benchtop incubator (21% O_2). The plates were photographed using the UVI-Pro Chemi visualisation chamber and software.

In the absence of chloramphenicol, induction of protein expression by IPTG (0.1 mM and 0.5 mM) resulted in ~10^3-fold decrease in colony formation for the JW1114 strain over-expressing ycfD WT ('JW+WT'), but not the empty-vector ('JW+EV', Figure 3.12), measured by observing the dilution at which colonies were still observable for each strain. At 0 mM IPTG, colonies were observed at 10^7 dilution of each strain; at 0.1 and 0.5 mM IPTG, all strains apart from 'JW+WT' showed colony formation at 10^7-dilution, whereas colonies were observed for 'JW+WT' only to 10^4-fold dilution. It was concluded that ycfD over-expression leads to chloramphenicol-independent growth retardation which likely originates from binding of L16 and its sequestration from the ribosome. The demonstration of a ycfD-dependent effect was therefore not possible by genetic manipulation; a small-molecule inhibitor of ycfD was required to demonstrate a ycfD-dependent effect on chloramphenicol sensitivity.

3.5. YcfD inhibitor screening

In order to identify a selective, cell permeable ycfD inhibitor, an inhibitor screening method was developed. Initially, ycfD was produced recombinantly and the reaction conditions optimised for screening of small molecule inhibitors. The ycfD-pET28a construct with an *N*-terminal His-tag (kindly provided by Dr. Wei Ge) was transformed in BL21 (DE3) Gold according to the standard protocol and purified by HisTrap Ni(II)-affinity chromatography followed by size-exclusion chromathography using S75 (300 mL) size-exclusion column (Amersham) to produce >85% pure protein.

Activity of ycfD was measured by monitoring the hydroxylation of ycfD substrate fragment L16 by MALDI-MS. Upon incubation of ycfD with an appropriate concentration of Fe(II), 2OG, ascorbate and L16 fragment peptide (predicted mass 2152 Da), a +16 Da shift was observed by MALDI-MS, indicating ycfD-dependent hydroxylation (Figure 3.13).

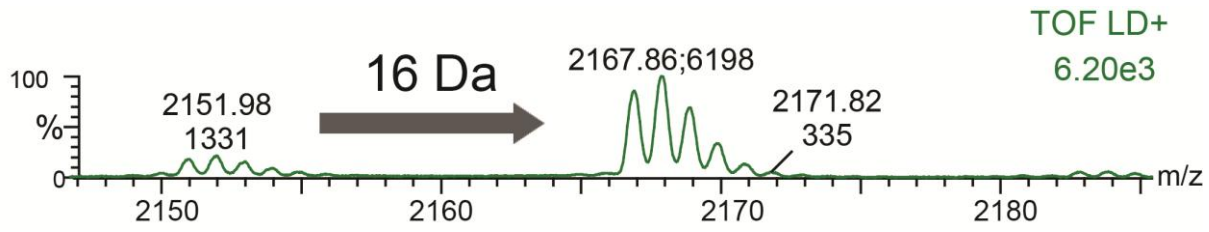


Figure 3.13: A representative MALDI-MS spectrum of ycfD-catalysed hydroxylation of L16 fragment peptide (20-mer; 100 μM; expected mass: 2152 Da) in the presence of Fe(II), 2OG and ascorbate (at 100 μM, 200 μM and 1 mM, respectively) in HEPES pH 8.0 at 37°C.

3.5.1. pH optimisation for ycfD activity

The dependence of ycfD activity on pH was assessed by incubating the enzyme with co-factors/co-substrates in buffers at their pKa value at a chosen pH (pH range 6.0-10.0 was tested; Figure 3.14). Tris-HCl buffer (50 mM, pH 7.5) decreased ycfD activity at low concentration of peptide substrate (10 μM) relative to HEPES (data not shown); HEPES/EPPS were therefore chosen as buffers for the pH range 7.0-8.0.YcfD activity was maximal at pH 8.0 (Figure 3.14). HEPES (50 mM, pH 8.0) was used for all subsequent kinetic studies.

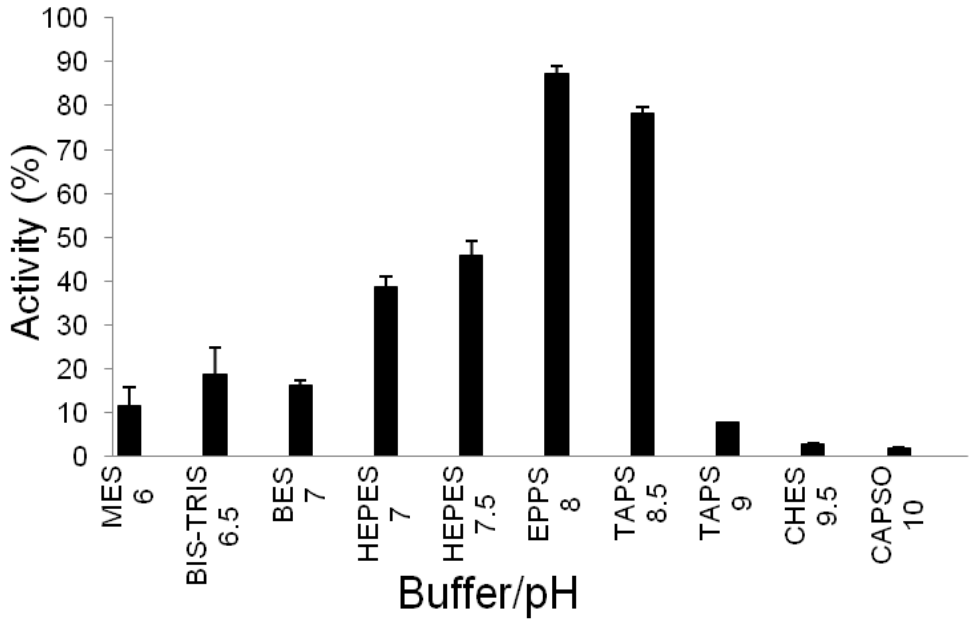


Figure 3.14: Determination of pH dependence of ycfD activity. YcfD (1 μM) was incubated with Fe(II), Asc, 2OG and L16 substrate peptide (2 μM, 1 mM, 200 μM, and 10 μM, respectively) in buffers at selected pH values. Stock solutions of individual co-factors/co-substrates was prepared in HEPES buffer, pH 7.5; subsequent dilutions were made in selected buffers. Reactions were allowed to proceed for 30 min at 37°C before quenching with equal volume of $CF_3COOH_{(aq)}$ and analysis by MALDI-MS.

3.5.2. Optimising kinetic parameters for ycfD activity screening

Enzyme inhibitor screening requires optimisation of the kinetic parameters for optimal catalytic activity. YcfD kinetics were studied by monitoring the L16 fragment peptide hydroxylation using MALDI-MS, for which a 20-mer peptide (KPITEKPLAVR^{81}MGKGKGNVE) was synthesised using solid-phase peptide synthesis, followed by HPLC purification. Enzyme inhibitor potency can be measured by the half maximal inhibitory concentration (IC_{50}), which is defined by the concentration of an inhibitor required to inhibit the rate of a given process by half; a low IC_{50} value suggests a potent inhibitor. In competitive inhibition assays, determination of IC_{50} values requires the endogenous ligand present at low concentration (at or below its K_m^{app} value), to ensure competitive binding to the enzyme active site, and that reaction rates are measured in the linear phase. The K_m^{app} is defined as the substrate concentration required for half-maximal reaction rate based on the steady-state approximation of enzyme catalysis. At low substrate concentrations, $[S] << K_m$, and the rate depends linearly on [S]. At high substrate concentrations, $[S] >> K_m$, and the rate is independent of [S]. V_{max} is the maximal velocity for a reaction at a given enzyme concentration and does not represent a fundamental property of an enzyme.

Initially, the time-course for the hydroxylation reaction was determined at 10 μM and 100 μM of the *E. coli* L16 peptide fragment (20-mer) with other co-factors/co-substrates provided in excess (Fe(II), ascorbate and 2OG at 100 μM, 1 mM and 200 μM, respectively; Figure 3.15). The time-course revealed linearity between 0-8 min for hydroxylation of 100 μM of substrate (Figure 3.15b).

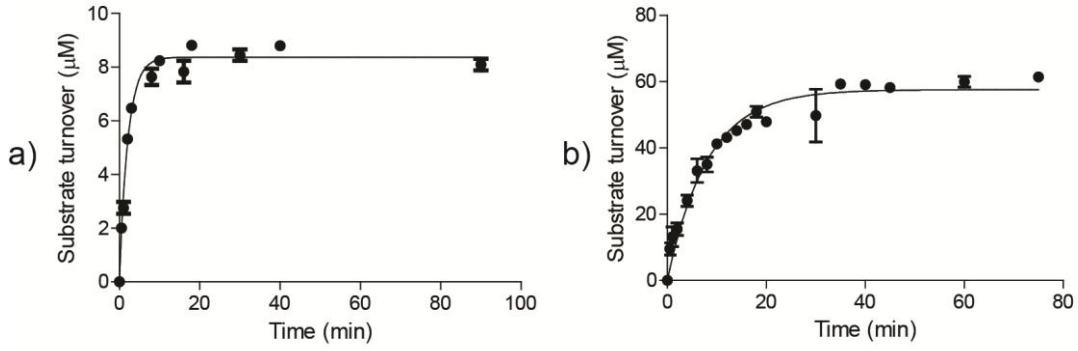


Figure 3.15: Time-dependent substrate turnover by ycfD. YcfD was incubated with a) 10 μM or b) 100 μM of *E. coli* L16 20-mer peptide and co-factors (Fe(II), 100 μM, ascorbate 1 mM, 2OG 200 μM) in 150 μL. 5 μL aliquots were removed at specified time points and quenched with equal volume of 1% TFA. Average value of background-substracted three independent reactions is shown, with error bars representing standard deviation from the mean.

It was envisaged that a potent ycfD inhibitor would be competitive with respect to the 2OG co-substrate. In order to ensure competitive conditions for inhibitor screening, the K_m^{app} value was determined for 2OG. All components were provided in excess, and the concentration of 2OG was systematically varied. Reactions were quenched in the linear phase (5 min) with equal volume of $CF_3COOH_{(aq)}$ (0.1%) and substrate turnover was analysed by MALDI-MS. K_m^{app} value for 2OG was 8.6 ± 3.3 μM (Figure 3.16), comparable to other characterised 2OG oxygenases (23).

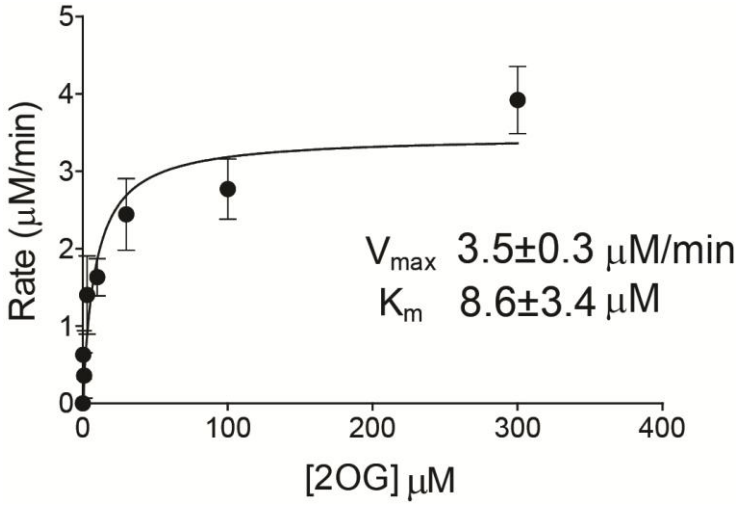


Figure 3.16: Rate dependence of ycfD on 2OG. YcfD (1 μM) was incubated with Fe(II), ascorbate, substrate peptide (at 100 μM, 1 mM and 100 μM, respectively) and a serial dilution (300-0 μM) of 2OG. Reaction was allowed to proceed for 5 min before quenching with equal volume of CF_3COOH(aq) (0.1 %). Results were analysed by MALDI-MS. Reactions were carried out in triplicate; error bars denote standard deviation from the mean**.**

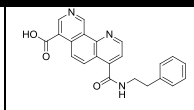
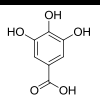
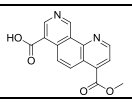
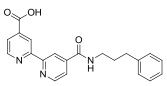
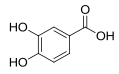
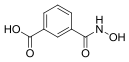
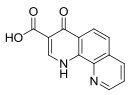
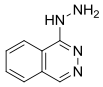
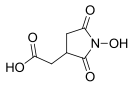
3.5.3. *In vitro* MALDI-MS Inhibitor Screen

In order to identify potential ycfD inhibitors for inhibition of L16 hydroxylation in cells, an *in vitro* screen of >190 known 2OG oxygenase inhibitors, assembled in the laboratory of Prof. Chris Schofield based on inhibition studies of human and bacterial 2OG oxygenases (24) was then tested. The library comprises a diverse set of 2OG mimetics (e.g. *N*-oxalyl glycine, 2,4-pyridine dicarboxylic acid), iron chelators (e.g. desferioxamine, bipyridyls) and natural products (e.g. daminozide). Compounds from the inhibitor library were screened in duplicate at 100 μM and as single measurements at 20 μM. The compounds that showed consistent inhibition of ycfD activity at both concentrations (<10% activity at 100 μM and <40% at 20 μM) were considered as hits (highlighted in Table 3.2).

Table 3.1: Average activity of ycfD incubated with inhibitors from the 2OG mimetic library. YcfD (1 μM) was incubated with 2OG mimetics (100/20 μM), Fe(II), 2OG, ascorbate and L16 fragment peptide (20-mer) at 20μM , 10 μM, 2 mM and 50 μM for 20 min at 37°C before quenching with equal volume of CF_3COOHaq (1% vol/vol). Ammonium citrate was added (20 mM) to improve MALDI-MS detection. YcfD activity (%) is reported relative to control (DMOG only) as a heat map, with red fields indicating low ycfD activity in the presence of an inhibitor, orange fields indicating intermediate activity, and green fields indicating high activity of ycfD.

	μM	1	2	3	4	5	6	7	8	9	10	11
1A	100	30.2±8.3	99.6±0.8	2.3±3.2	0±0	40.8±50.8	91.4±2.5	104.8±5.1	102.7±3.7	83.8±13.2	99.0±6.6	105.3±6.1
	20	82.6	85.5	*31.3*	29.2	87.7	89.7	91.2	69.7	98.9	68.9	93.1
1B	100	99.2±7.7	102.12±5.9	94.0±9.0	99.0±2.4	32.3±45.6	0	93.2±3.8	94.0±6.9	85.8±13.0	48.3±0.9	84.6±7.3
	20	93.2	88.6	*73.4*	73.7	96.4	65.5	72.3	78.8	90.9	95.9	91.1
1C	100	96.9	101.5±1.5	97.1	26.8	15.4	0±0	3.9±5.5	98	90.9	104.2	103.1±5.1
	20	83.0	89.9	*73.3*	74.4	65.4	93.9	81.8	89.5	89.3	91.2	91.7
1D	100	100.5±1.2	89.7±17.4	58.8±6.5	14.1±10.8	71.8±29.1	90.0±10.6	98.4±3.5	58.5±31.9	0±0	95.2±7.5	68.6±1.0
	20	90.3	97.1	*91.7*	68.4	92.8	89.0	98.7	100.0	59.4	95.4	80.8
1E	100	0±0	77.3±29.5	86.0±0.5	98.5±3.7	82.9±27.8	85.2±15.9	88.0±24.2	45.7	85.2±13.2	99.1±6.2	91.4±16.9
	20	101.3	98.8	*85.1*	81.4	107.9	105.5	107.3	92.8	107.2	92.8	102.9
1F	100	104.6±10.6	114.8±42.3	58.8±51.5	79.4±24.3	31.3±13.6	41.2±58.2	71.6±12.0	53.5±36.5	80.9±10.8	114.4±45.6	88.9±7.7
	20	100.1	92.4	*94.8*	96.4	88.4	93.2	109.1	107.0	107.5	26.3	93.5
1G	100	88.9±6.5	99.7±15.6	84.8±5.5	66.8±1.1	100.6±2.8	79.6±9.3	20.3±14.5	23.6±14.4	62.8±5.7	54.5±9.7	22.4±31.7
	20	102.9	96.8	*106.3*	89.5	66.2	98.9	85.2	99.9	98.5	99.0	79.8
1H	100	72.7±16.4	0±0	92.8±2.9	85.7±3.9	100.3±0.8	15.2±16.5	7.2±9.1	36.5±51.6	93.2±6.1	2.7±3.9	5.7±8.1
	20	79.0	87.7	*96.6*	99.2	96.3	86.5	58.2	101.8	90.5	101.6	37.6
2A	100	18.4±2.3	25.2±35.7	62.8±2.8	86.0±10.9	102.3±8.0	107.2±4.9	110.3±4.3	101.8±19.8	3.8±5.4	0±0	113.0±39.3
	20	76.4	86.0	79.3	78.2	88.3	77.1	100.3	95.4	47.5	35.8	88.3
2B	100	100.8±15.0	102.7±14.0	7.2±4.1	0±0	91.8±2.3	48.4±68.4	12.0±3.4	0.3±0.5	48.9±4.4	96.1±0.4	30.1±42.6
	20	79.5	82.5	33.0	55.1	63.7	80.1	58.6	36.7	78.8	70.3	74.9
2C	100	0±0	91.8±3.3	4.0±5.6	83.8±4.0	78.8±23.0	56.8±55.5	90.1±2.0	92.1±14.0	97.9±10.3	93.3±1.1	45.1±26.7
	20	33.1	82.5	75.1	73.3	92.2	83.2	77.8	77.5	74.0	83.8	72.7
2D	100	97.0±2.7	92.4±2.4	95.0±6.2	3.9±5.6	94.4±12.0	59.9±52.9	93.6±18.0	98.0±12.3	8.4±6.7	75.5±16.3	98.2±15.5
	20	74.2	81.0	80.1	56.0	81.3	85.5	86.4	78.4	65.1	84.7	81.4
2E	100	87.8±2.2	103.6±7.0	16.4±13.6	93.8±1.0	4.4±6.2	94.1±8.0	66.7±7.9	18.0±7.9	95.5±1.5	3.3±4.6	99.3±21.6
	20	68.3	83.4	85.5	86.2	69.6	86.7	83.6	64.7	85.6	40.0	89.7
2F	100	103.8±4.0	94.1±1.9	6.3±8.9	89.7±1.8	95.5±1.5	96.0±5.6	100.8±13.6	95.1±12.2	99.3±6.4	18.3±14.9	79.5±41.3
	20	85.4	68.0	64.7	79.3	86.8	86.4	88.7	85.9	92.8	93.7	93.0
2G	100	102.7±7.0	98.5±4.5	96.4±1.8	22.75±6.7	89.8±17.9	72.1±35.4	0±0	85.8±7.8	99.0±3.3	104.1±3.4	98.7±15.8
	20	85.9	85.9	87.7	80.6	74.6	91.7	34.2	84.0	91.4	79.1	72.3
2H	100	84.0±16.5	83.7±8.6	84.8±2.2	96.2±2.4	91.3±21.5	87.4±41.0	56.1±46.0	74.0±31.9	51.9±73.4	60.9±30.8	64.8
	20	85.3	63.8	94.1	66.7	80.9	82.2	78.8	77.8	91.2	85.6	77.2

Table 3.2: Chemical structures of the ten most potent inhibitors of ycfD activity determined at 100 μM and 20 μM in a single-point activity assay as detailed above.

Well	Structure	% Activity at 100 μM	% Activity at 20 μM	Well	Structure	% Activity at 100 μM	% Activity at 20 μM
1A4		0±0	29.2	2A10		0±0	35.8
1A3		2.3±3.2	31.3	2B8		0.3±0.5	36.7
2B3		7.2±4.1	33.0	1H11		5.7±8.1	37.6
2C1		0±4.1	33.1	2E10		3.3±4.6	39.9
2G7		0±0	34.2	2A9		3.8±5.4	47.5

3.5.4. Differential scanning fluorimetry (DSF) screen of ycfD stabilisers

An orthogonal screening approach was then applied to screen for potential ycfD inhibitors by the measurement of protein stabilisation using differential scanning fluorimetry (DSF). Small molecule binders/inhibitors stabilise the tertiary fold of an enzyme due to increased stability of protein-inhibitor complex, thus increasing the temperature of protein unfolding. This stabilisation effect is reflected in a higher temperature of structural denaturation compared to untreated control, which can be quantified as the 'melting temperature shift' (ΔT_m, or T_m shift). Various methods exist to measure the stability of protein structure, e.g. circular dichroism (CD), differential scanning calorimetry (DSC) and differential scanning fluorimetry (DSF). The latter provides the most rapid and inexpensive screening method for identification of protein ligands based on ΔT_m that can be executed in medium throughput (48-well plates) and in small volumes (50 μL). Measurements do not vary significantly with buffer composition and measurement of blanks, except to exclude the possibility of compound fluorescence, is not necessary. The downside of the measurement is its indirectness; DSF relies on the binding of a fluorescent dye to hydrophobic residues of a protein, which are assumed to be buried in a hydrophobic core under normal conditions of protein folding (Figure 3.17) (25). DSF has been used extensively by academic and industrial researchers as an initial approach to ligand identification, because changes in T_m shift indicate strength of compound binding and provide a proxy for inhibitor potency. However,

DSF assays should be interpreted in combination with results of activity-based assays, such as MALDI-MS activity assay for 2OG oxygenases, in order to exclude compounds that bind allosterically and do not affect enzyme activity.

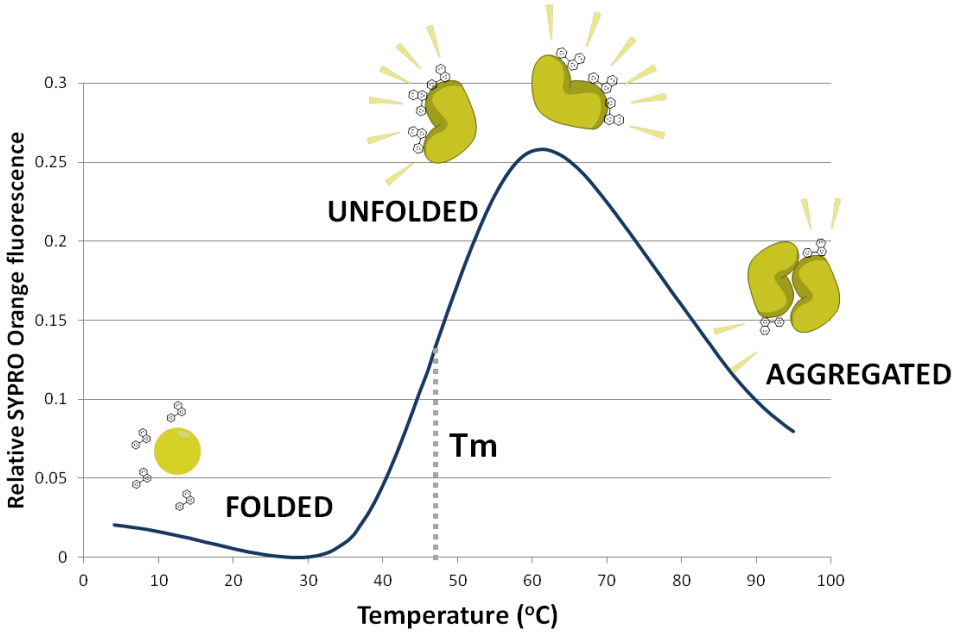


Figure 3.17: Differential scanning fluorimetry (DSF) measures the fluorescence signal of a hydrophobic dye. The dye binds to hydrophobic patches of a protein, which are buried in the inside of a protein under normal conditions. During protein denaturation (due to increased temperature or chaotropic agents) the hydropohibc patches are exposed, leading to binding of the dye concomitant with an increase in fluorescence signal. After maximum temperature at which proteins unfold, aggregation through hydrophobic interactions leads to a decrease in fluorescence signal.

A selection of compounds from the 2OG analogue library (400 μM) was therefore tested against ycfD (2 μM) in the presence of Mn(II) (50 μM), which is required for binding of metal-chelating compounds to the enzyme. The ten most stabilising compounds were tabulated and compared to the results of the activity screen (Figure 3.18, Table 3.3).

The comparison of MALDI-MS activity screen and the DSF-based T_m shift identified a limited number of common hits, which included compounds 2G7 (5-carboxy-8-hydroxyquinoline, IOX1), 1A4 and 1A3 (bipyridyl derivatives).

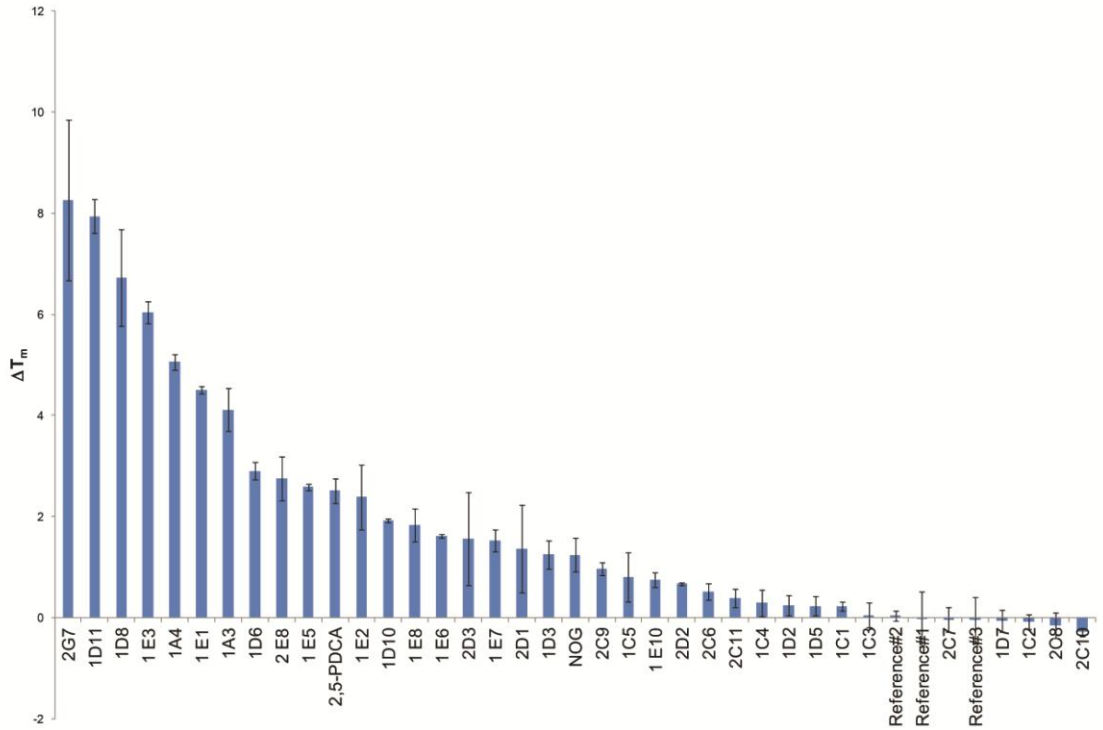
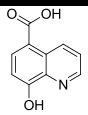
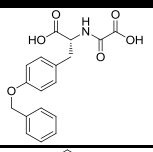
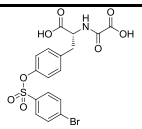
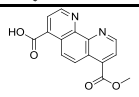
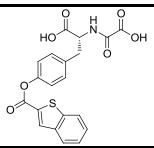
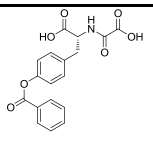
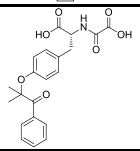
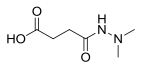
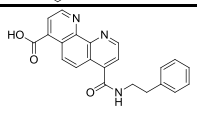
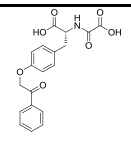


Figure 3.18: DSF screening of ycfD stabilisers. DSF analysis of ycfD protein stabilisation by small molecule inhibitors was conducted by incubating a selection of compounds from the 2OG analogue library was tested at 400 μM final concentration in the presence of ycfD (2 μM) and Mn(II) (50 μM) in order to enable binding of metal-chelating compounds to the enzyme. Sypro Orange was diluted in 50 mM HEPES, pH 7.5. Fluorescence of Sypro Orange was measured using Opticon monitor software intergrated with a thermal cycler (BioRad) between 25-95°C in 1°C steps. Potency of stabilisation is quantified by ΔT_m relative to positive control (DMSO). T_m shifts are reported relative to the control (DMSO). Compounds were screened in triplicate, average values are reported.

Table 3.3: Structures of the ten most potent stabilisers of ycfD as determined by DSF, ordered by decreasing ΔT_m.

	Well	Structure	ΔT_m (°C)		Well	Structure	ΔT_m (°C)
1	2G7		8.3±1.6	**6**	1E1		4.5±0.1
2	1D11		7.9±1.6	**7**	1A3		4.1±0.4
3	1D8		6.7±1.0	**8**	1D6		2.9±0.2
4	1E3		6.0±0.2	**9**	1D6		2.8
5	1A4		5.0±0.2	**10**	1E5		2.6±0.1

IOX1 is an inhibitor of the JMJD2 family of oxygenases (26), which contain a JmjC-domain and are structurally related to ycfD. Broad-spectrum activity against other 2OG oxygenases in humans has also been reported (23). A potent scaffold identified by both screens was a 4-carboxy-2,2'-bipyridyl scaffold, which also inhibits histone demethylases *in vitro* (27). It was envisaged that non-specific effects may be observed due to its strong Fe^{2+}-chelating properties (24). Interestingly, *N*-oxalyl-D-tyrosine (NOYD) derivatives, which were the most potent stabilisers of ycfD by the DSF assay did not inhibit the activity of ycfD, as measured by MALDI-MS, potentially suggesting allosteric binding that does not interfere with substrate or co-substrate binding. This result highlighted the discrepancy between activity- and stability-based assays and the need to perform them in conjunction. T_m shift assays measure the relative stabilisation of a protein by a compound, and not enzyme activity, and the two are not necessarily correlated (25).

Because of its reported biological activity against the JmjC-domain containing oxygenases in humans, as well as the observed potency of inhibition and stabilisation of ycfD, compound 2G7 (IOX1) was pursued as a potential ycfD inhibitor in cells.

3.5.5. Determination of IC_{50} values for IOX1 analogues against ycfD

The potency of IOX1 as a ycfD inhibitor *in vitro* was then quantified. At a limiting concentration of 2OG (20 μM, $2\times K_m^{app}$ value) to ensure competitive inhibition, its IC_{50} value was 38 μM (Figure 3.19a). The IC_{50} value for the IOX1 analogue 5-sulfo-8-hydroxyquinoline was 169 μM (Figure 3.19b), suggesting that the 5-carboxy group is important for interaction with the ycfD backbone, which was further corroborated by the observation that 5-chloro-7-iodo-8-hydroxyquinoline (clioquinol, a known inhibitor of FIH (28)), did not show marked inhibition of ycfD up to 1 mM concentrations (Figure 3.19c). The IC_{50} value for *N*-oxalyl glycine, an unreactive 2OG analogue and a pan-2OG oxygenase inhibitor, was 466 μM. Consistent with its relatively high IC_{50}, the ΔT_m value was relatively low (1.2±0.3°C, compared to 8.3±1.5°C for IOX1; Figure 3.18).

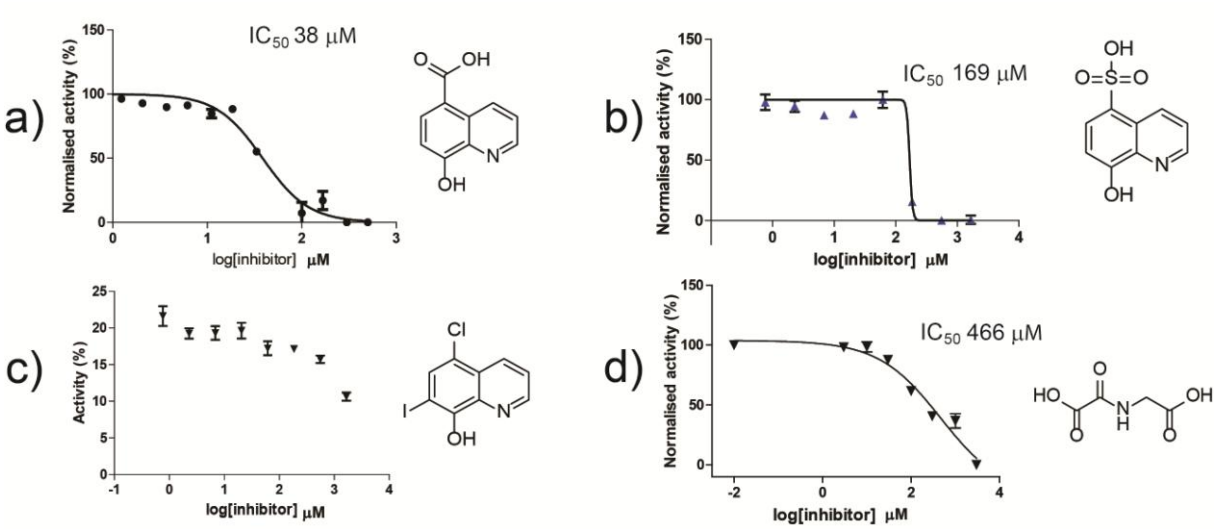


Figure 3.19: IC_{50} value determination of IOX1 analogues and NOG against ycfD. Inhibitors a) 5-carboxy-8-hydroxyquinoline, b) 5-sulfo-8-hydroxyquinoline, c) clioquinol and d) NOG were incubated with ycfD (1 μM) enzyme, 500 μM Fe, 1 mM ascorbate, 20 μM 2OG, 100 μM L16 fragment peptide) for 15 min at 37°C before quenching with CF_3COOHaq (1% vol/vol). Reactions were performed in triplicate and analysed by MALDI-MS. The mean of three experiments is shown with error bars denoting the standard deviation of the mean. Inhibition curves were fitted to the data using the *log Inhibitor vs. normalised response – Variable slope* function in GraphPad Prism (Version 5.0).

A model of IOX1 binding to the ycfD active site was proposed based on the homology of ycfD and JMJD2A active sites (Figure 3.20). The model identified a strong interaction between the 5-carboxy moiety of IOX1 with R140 (3.9-4.1 Å), which is involved in 2OG-binding (see Chapter 5). Furthermore, the side chains of S199 (OH-COOH distance 3.7 Å) and N197 (side chain amide-carboxylate distance 3.5 Å) may be involved in hydrogen-bonding interaction with the 5-carboxylate of IOX1. The 8-hydroxy substituent and the quinoline nitrogen are proposed to chelate the active site metal (1.7 and 1.9 Å distance, respectively). Importantly, the model identified C2 as a potential site for further functionalisation of the IOX1 scaffold without disrupting the binding to the ycfD active site. Due to the presence of hydrophobic residues at the back of the binding pocket, substitutions at C6 and C7 are unlikely to afford improved binding potency (Figure 3.20).

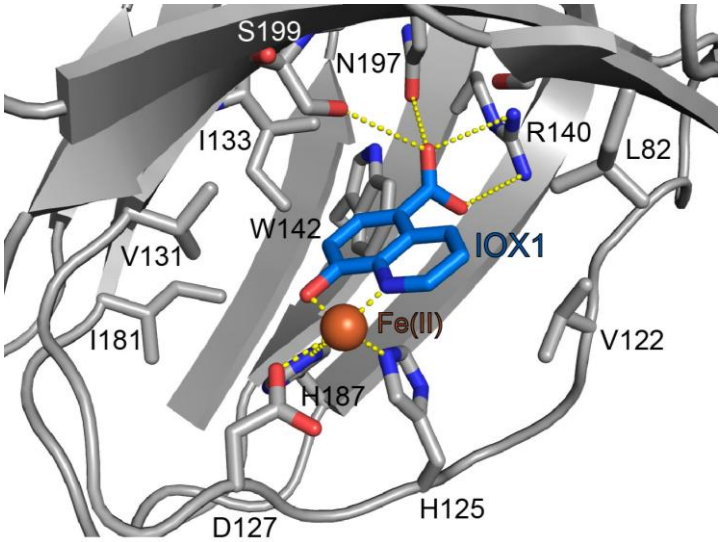


Figure 3.20: A Protein-ligand docking simulation of ycfD-IOX1 complex. A crystal structure of JMJD2A in complex IOX1 (PDB ID: 3NJY (26)) was used for docking studies. After eliminating the substrate, atomic charges were assigned to the model of KDM4A through the restrained electrostatic potential (RESP) methods using AMBER. Because of the need for electronic charge redistribution resulting from central ferrous ion and its ligand group, density functional theory (DFT) calculations were carried out with geometry optimization of a simplified model for ferrous ion in complex with the amino acid residues and the ligand 5-carboxy-8-hydroxyquinoline (IOX1). Modelling was performed by Hwan Choi.

3.5.6. IOX1 treatment increases the antibiotic potency of chloramphenicol

To date, ycfD is the only identified JmjC-domain containing a 2OG oxygenase in *E. coli* and IOX1 was assumed not to have other inhibition targets in *E. coli*. It was therefore tested as a ycfD inhibitor in cells. Dimethyl *N*-oxalyl glycine (DMOG), a cell-permeable analogue of NOG (29), was tested as a negative control (the IC_{50} of NOG was >12-fold greater than for IOX1, Figure 3.19). BW25113 was grown in the presence of chloramphenicol and IOX1/DMOG and the growth rate was compared to JW1114 treated with chloramphenicol only (Figure 3.21). DMOG did not inhibit the growth of BW25113 and JW1114 differentially, consistent with NOG being a relatively poor inhibitor of ycfD. Consistent with its strong inhibition of ycfD *in vitro*, IOX1 achieved growth inhibition of BW25113; at 2.5 μg/mL chloramphenicol, 1 mM IOX1 resulted in a 6.3 h delay in reaching the mid-log phase, equivalent to JW1114 treated with chloramphenicol alone. At lower chloramphenicol concentrations, the growth delay of IOX1-treated BW25113 was greater than of JW1114: at 1 μg/mL the growth delay for IOX1-treated BW25113 was 5 h, and for JW1114 it was 2.3 h. Both JW1114 and IOX1-treated BW25113 reached optical density of 0.8, compared to 1.2 for untreated BW25113. Consistent with its inhibition of ycfD *in vitro*, the results in cells suggested that IOX1 potentiates the effects of

chloramphenicol against BW25113 which were similar to the effects of chloramphenicol on JW1114 alone. However, the slower growth of IOX1-treated BW25113 compared to JW1114 at 1 μg/mL chloramphenicol suggested that IOX1 may also have off-target effects at 1 mM (Figure 3.21A).

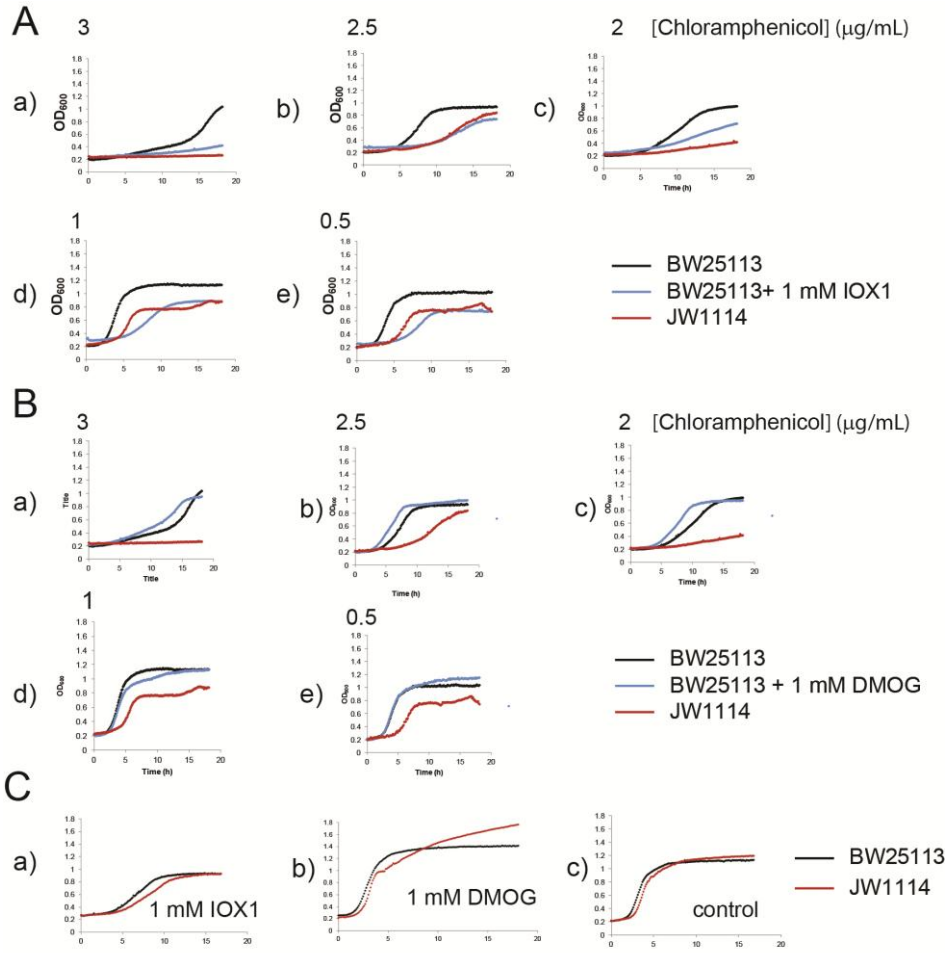


Figure 3.21: Treatment of BW25113 with IOX1, but not DMOG, increases its sensitivity to chloramphenicol. A) BW25113 (black), BW25113 treated with 1 mM IOX1 (blue) or JW1114 (red) were grown in the presence of chloramphenicol at a) 3 μg/mL, b) 2.5 μg/mL, c) 2.0 μg/mL, d) 1.0 μg/mL or e) 0.5 μg/mL. B) BW25113 (black), BW25113 treated with 1 mM DMOG (blue) or JW1114 (red) were grown in the presence of the same concentrations of chloramphenicol as in A. C) control growths of BW25113 and JW1114 in the presence of a) 1 mM IOX1, b) 1 mM DMOG or c) no inhibitor. OD_{600} was measured spectrophotometrically for 18 h after inoculation with BW25113 culture diluted to OD_{600} 0.05 in 2TY medium.

Mass spectrometric analysis of L16 hydroxylation confirmed that IOX1 leads to nonspecific effects above 0.5 mM (Figure 3.22). A dose-dependent decrease in L16 hydroxylation was observed between 0-0.25 mM IOX1; at 0.5 mM, a further -14 Da shift was observed, consistent with demethylation of L16 (Figure 3.22). L16 is *N*-terminally methylated by an unknown methyltransferase (30). A high concentration of IOX1 may inhibit the activity of the methyltransferase by chelation of iron or other transition metals, which may be required for the activity of the methyltransferase.

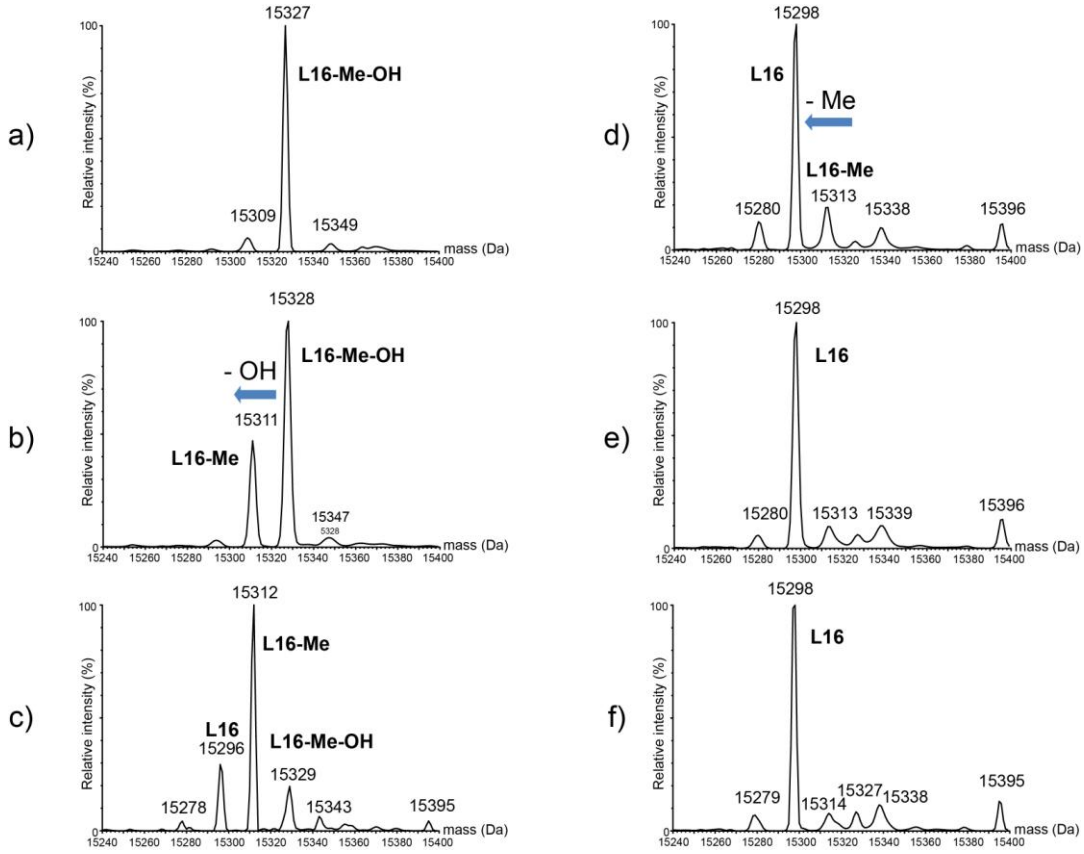


Figure 3.22: LC-MS analysis of ribosomal protein L16 modification. Deconvoluted ESI-MS spectra of L16 treated with IOX1 demonstrate a dose-dependent decrease in hydroxylation of L16 and a further loss of methylation. a) 0 mM, b) 0.1 mM, c) 0.25 mM, d) 0.5 mM, e) 1.0 mM and f) 1.5 mM IOX1 was used to treat BW25113 grown in 2TY medium at 37°C. Cultures were harvested at OD_{600} 0.8 and ribosomes extracted as detailed in Materials and Methods.

Dose-dependent effects of IOX1 on BW25113 growth were then measured between 0-0.5 mM IOX1. At 1 μg/mL chloramphenicol, treatment of BW25113 with 0.25 mM IOX1 resulted in growth delay of < 1h, and a reduction of maximum optical density from 1.6 (untreated BW25113) to 0.9 (IOX1-treated BW25113). Treatment with 0.5 mM IOX1 resulted in growth comparable to untreated JW1114; JW1114 treated with 0.25 mM IOX1 demonstrated a growth delay of 2.5 h relative to untreated JW1114, but did not lead to a reduction in maximum optical density (Figure 3.23). At higher doses of chloramphenicol, a dose-dependent decrease in maximum optical density was observed for IOX1-treeated BW25113; growth of JW1114 was completely inhibited. The interpretation of the data was confounded by the fact that at 0.25 mM IOX1 there was ~20% residual L16 hydroxylation, whereas at 0.5 mM IOX1, L16 was also >80% demethylated, which may contribute to some of the observed growth defects of BW25113; dissection of the individual effects of hydroxylation and methylation is impossible without the knowledge of the L16 methyltransferase

(with which the methyltransferase knock-out could be subjected to the chloramphenicol sensitivity assay). Given that the majority of the known 2OG oxygenase inhibitors are iron chelators, it will be challenging to develop a ycfD inhibitor that would not, at least partly, lead to demethylation of L16.

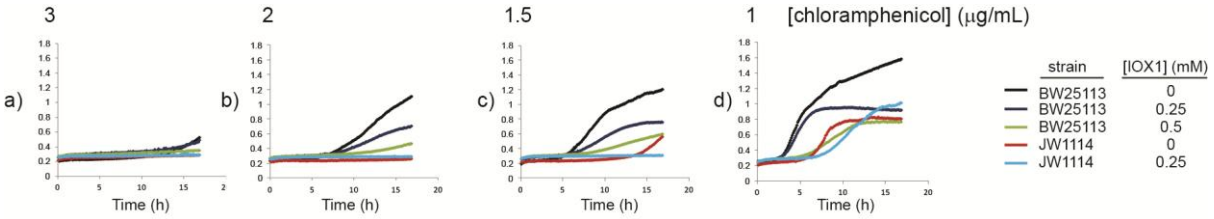


Figure 3.23: Dose-dependent effect of chloramphenicol on BW25113 and JW1114 treated with 0, 0.25 and 0.5 mM IOX1. BW25113 (black), BW25113 treated with 0.25 mM IOX1 (blue), BW25113 treated with 0.5 mM IOX1, JW1114 (red) and JW1114 treated with 0.25 mM IOX1 (cyan) were grown in the presence of chloramphenicol at a) 3 μg/mL, b) 2.0 μg/mL, c) 1.5 μg/mL, and d) 1.0 μg/mL. OD_{600} was measured spectrophotometrically for 16 h after inoculation with BW25113 and JW1114 cultures diluted to OD_{600} 0.05 in 2TY medium.

3.6. Conclusions

Post-translational modifications of the rRNA are known to cause antibiotic resistance. For example, methylation (mono- and dimethylations, catalysed by methyltransferases from Erm family) of A2058 confers resistance to MLS_B antibiotics. Mutations of ribosomal proteins, including L16, result in antibiotic resistance; L16 mutations A50, R51, I52, T55 and R56 in *S. pneumoniae*, *S. aureus* and *E. faecium* confer evernimicin resistance (31). However, to the best of my knowledge, no ribosomal protein post-translational modification has been implicated in antibiotic resistance. In this chapter it was demonstrated that ycfD-catalysed hydroxylation of the ribosomal protein L16 acts to increase resistance to the PTC inhibitors chloramphenicol and clindamycin/lincomycin. Although evidence was complicated by the inability to operate in an isogenic model of ycfD (likely due to relative toxicity of over-expressing a strong interactor of a key ribosomal proteins), chloramphenicol was shown to selectively inhibit the growth of *ycfD*-deleted strain. Furthermore, the chloramphenicol sensitivity of the *E. coli* wild-type strain could be increased by inhibiting the activity of ycfD by removing co-factors required for catalytic activity (Fe(II) and O_2, achieved by treatement with iron-chelator desferioxamine and growth in hypoxic chamber, respectively). Chloramphenicol sensitivity could also be increased by using a ycfD inhibitor, IOX1, which inhibited ycfD with IC_{50} of 38 μM *in vitro*. Mass spectrometry analysis revealed that IOX1 treatment above 0.5 mM resulted in non-

specific effects on L16 methylation, which limited the interpretation of chloramphenicol sensitivity in co-treatment with IOX1. A more selective inhibitor of ycfD will be required to avoid secondary effects on L16 methylation in order to demonstrate the link between L16 hydroxylation and chloramphenicol sensitivity using a chemical probe.

The finding that *ycfD* deletion increases chloramphenicol sensitivity is significant, because oxygen dependence of chloramphenicol sensitivity *via* a 2OG oxygenase-mediated modification has not previously been reported. This finding is in agreement with the data available from a genome-wide gene deletion screen of KO fitness; analysis of phenotypic correlations between gene deletion strains revealed that *ycfD* KO strain JW1114 has an inverse phenotypic correlation with the *MarR* deletion strain, the transcriptional repressor of multiple antibiotic resistance, including to chloramphenicol (32). The inverse correlation with the phenotype of a repressor of antibiotic resistance is consistent with *ycfD* increasing antibiotic resistance.

The therapeutic potential of a post-translational modification mediating antibiotic resistance is very appealing, because it provides an opportunity for medicinal targeting of ribosome-modifying enzymes, for example ycfD, which may be more druggable than the ribosome itself. In co-treatment with an existing antibiotic, such as chloramphenicol, a small molecule inhibitor would achieve a potentiated antibiotic effect.

The data presented in this chapter provides a proof-of-concept study that it is in principle possible to selectively inhibit ribosomal activity by exploiting differential levels of ribosomal protein PTMs. The clinical implication of inhibition of modified ribosomes extends beyond microbiology; the existence of human homologues of ycfD, NO66 and MINA53, suggests that human ribosomes may be unhydroxylated, and thereby susceptible to antibiotic inhibition, in hypo-oxygenated growth environment (e.g. solid tumours). Alternatively, ribosomes may be unhydroxylated in cancers harbouring genetic variations that deactivate ribosomal oxygenases. In Chapter 6, a demonstration that clinical cancer tissues carry deactivating mutations of ROX genes, resulting in unhydroxylated ribosomal proteins, will be presented. The existence of selectively modified ribosomal proteins in

clinical cancer tissues could in principle provide a biochemical mark that could be exploited for selective inhibition using a ribosomal antibiotic with selectivity for a subset of ribosomal PTMs.

While potentiated antibiotic effects can be achieved through a combination treatment with a ribosomal oxygenase (ROX) inhibitor, leading to a much improved potency of the antibiotic, it is in principle possible to develop a chemical agent that would incorporate both activities in one molecule. It was recently demonstrated that the dichloromoiety of chloramphenicol can be replaced by large, aromatic substituents without the loss of binding affinity for the ribosome (9,33). The work presented in this chapter identified a potentially useful scaffold for achieving ycfD inhibition *in vivo*, which resulted in increased sensitivity to chloramphenicol. The C2 position on the hydroxyquinoline scaffold of IOX1 was identified as a potential anchor point for further chemical modifications without disrupting the binding to ycfD. Future work may therefore focus on development of a dual-function chemical agent (Figure 3.24) that would incorporate both ROX- and ribosome-inhibition properties, thereby potentiating the effect of the antibiotic alone.

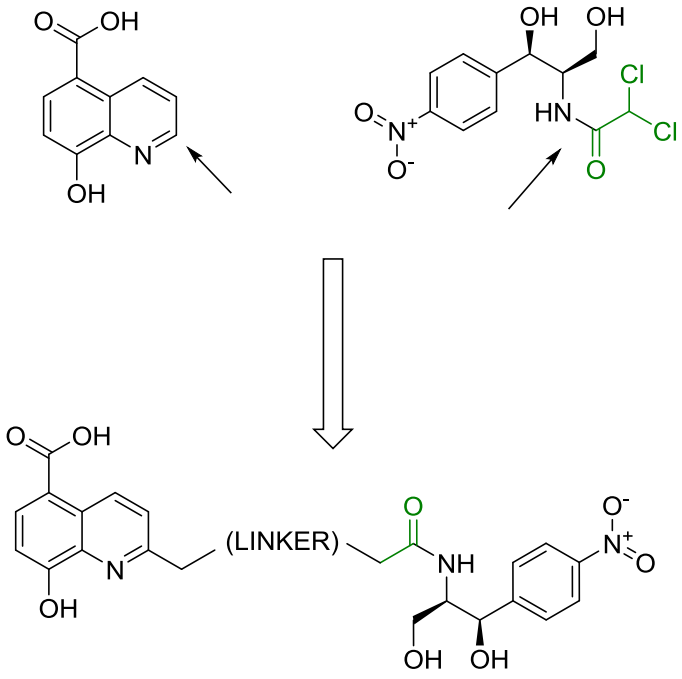


Figure 3.24: A dual-action inhibitor could in principle be synthesised to incorporate both ROX- and ribosome-inhibitory properties, thereby enhancing the potency of the antibiotic alone, by sensitising the ribosome via reduction of ribosomal protein hydroxylation.

3.7. References

1. Spahn, C. M., and Prescott, C. D. (1996) Throwing a spanner in the works: antibiotics and the translation apparatus. *J Mol Med (Berl)* **74**, 423-439
2. Poehlsgaard, J., and Douthwaite, S. (2005) The bacterial ribosome as a target for antibiotics. *Nat Rev Microbiol* **3**, 870-881
3. Wilson, D. N. (2011) On the specificity of antibiotics targeting the large ribosomal subunit. *Ann Ny Acad Sci* **1241**, 1-16
4. Decatur, W. A., and Fournier, M. J. (2002) rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-351
5. Skinner, R., Cundliffe, E., and Schmidt, F. J. (1983) Site of action of a ribosomal RNA methylase responsible for resistance to erythromycin and other antibiotics. *J Biol Chem* **258**, 12702-12706
6. Weisblum, B. (1995) Erythromycin resistance by ribosome modification. *Antimicrob Agents Chemother* **39**, 577-585
7. Navarre, W. W., Zou, S. B., Roy, H., Xie, J. L., Savchenko, A., Singer, A., Edvokimova, E., Prost, L. R., Kumar, R., Ibba, M., and Fang, F. C. (2010) PoxA, yjeK, and elongation factor P coordinately modulate virulence and drug resistance in Salmonella enterica. *Mol Cell* **39**, 209-221
8. Dunkle, J. A., Xiong, L., Mankin, A. S., and Cate, J. H. (2010) Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc Natl Acad Sci U S A* **107**, 17152-17157
9. Bulkley, D., Innis, C. A., Blaha, G., and Steitz, T. A. (2010) Revisiting the structures of several antibiotics bound to the bacterial ribosome. *Proc Natl Acad Sci U S A* **107**, 17158-17163
10. Hansen, J. L., Ippolito, J. A., Ban, N., Nissen, P., Moore, P. B., and Steitz, T. A. (2002) The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Mol Cell* **10**, 117-128
11. Pestka, S. (1967) The action of streptomycin on protein synthesis in vitro. *Bull N Y Acad Med* **43**, 126-148
12. Nierhaus, D., and Nierhaus, K. H. (1973) Identification of Chloramphenicol-Binding Protein in Escherichia-Coli Ribosomes by Partial Reconstitution. *Proc Natl Acad Sci U S A* **70**, 2224-2228
13. Maimets, T., Remme, J., and Villems, R. (1984) Ribosomal protein L16 binds to the 3'-end of transfer RNA. *FEBS Lett* **166**, 53-56
14. Bock, L. V., Blau, C., Schroder, G. F., Davydov, II, Fischer, N., Stark, H., Rodnina, M. V., Vaiana, A. C., and Grubmuller, H. (2013) Energy barriers and driving forces in tRNA translocation through the ribosome. *Nat Struct Mol Biol* **20**, 1390-1396
15. Contreras, A., Barbacid, M., and Vazquez, D. (1974) Binding to ribosomes and mode of action of chloramphenicol analogues. *Biochim Biophys Acta* **349**, 376-388
16. Pongs, O., Bald, R., and Erdmann, V. A. (1973) Identification of Chloramphenicol-Binding Protein in Escherichia-Coli Ribosomes by Affinity Labeling. *Proc Natl Acad Sci U S A* **70**, 2229-2233
17. Pongs, O., and Messer, W. (1976) The chloramphenicol receptor site in Escherichia coli in vivo affinity labeling by monoidoamphenicol. *J Mol Biol* **101**, 171-184
18. Siibak, T., Peil, L., Donhofer, A., Tats, A., Remm, M., Wilson, D. N., Tenson, T., and Remme, J. (2011) Antibiotic-induced ribosomal assembly defects result from changes in the synthesis of ribosomal proteins. *Mol Microbiol* **80**, 54-67
19. Corbett, M. D., and Chipko, B. R. (1978) Synthesis and Antibiotic Properties of Chloramphenicol Reduction Products. *Antimicrob Agents Ch* **13**, 193-198
20. Rothbauer, U., Zolghadr, K., Muyldermans, S., Schepers, A., Cardoso, M. C., and Leonhardt, H. (2008) A versatile nanotrap for biochemical and functional studies with fluorescent fusion proteins. *Mol Cell Proteomics* **7**, 282-289

21. Moore, V. G., Atchison, R. E., Thomas, G., Moran, M., and Noller, H. F. (1975) Identification of a ribosomal protein essential for peptidyl transferase activity. *Proc Natl Acad Sci U S A* **72**, 844-848
22. Amann, E., Ochs, B., and Abel, K. J. (1988) Tightly Regulated Tac Promoter Vectors Useful for the Expression of Unfused and Fused Proteins in Escherichia-Coli. *Gene* **69**, 301-315
23. Hopkinson, R. J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K. H., Li, X. S., Kristensen, J. B. L., King, O. N. F., Chan, M. C., Yeoh, K. K., Choi, H., Walport, L. J., Thinnes, C. C., Bush, J. T., Lejeune, C., Rydzik, A. M., Rose, N. R., Bagg, E. A., McDonough, M. A., Krojer, T. J., Yue, W. W., Ng, S. S., Olsen, L., Brennan, P. E., Oppermann, U., Muller, S., Klose, R. J., Ratcliffe, P. J., Schofield, C. J., and Kawamura, A. (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci* **4**, 3110-3117
24. Rose, N. R., McDonough, M. A., King, O. N., Kawamura, A., and Schofield, C. J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev* **40**, 4364-4397
25. Niesen, F., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **9**, 2212-2221
26. King, O. N., Li, X. S., Sakurai, M., Kawamura, A., Rose, N. R., Ng, S. S., Quinn, A. M., Rai, G., Mott, B. T., Beswick, P., Klose, R. J., Oppermann, U., Jadhav, A., Heightman, T. D., Maloney, D. J., Schofield, C. J., and Simeonov, A. (2010) Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5**, e15535
27. Chang, K. H., King, O. N., Tumber, A., Woon, E. C., Heightman, T. D., McDonough, M. A., Schofield, C. J., and Rose, N. R. (2011) Inhibition of histone demethylases by 4-carboxy-2,2'-bipyridyl compounds. *ChemMedChem* **6**, 759-764
28. Moon, H., Han, S., Park, H., and Choe, J. (2010) Crystal structures of human FIH-1 in complex with quinol family inhibitors. *Mol Cells* **29**, 471-474
29. Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Gielbert, J., Gaskell, S. J., von Kriegsheim, A., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001) Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **292**, 468-472
30. Brosius, J., and Chen, R. (1976) The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett* **68**, 105-109
31. Belova, L., Tenson, T., Xiong, L., McNicholas, P. M., and Mankin, A. S. (2001) A novel site of antibiotic action in the ribosome: interaction of evernimicin with the large ribosomal subunit. *Proc Natl Acad Sci U S A* **98**, 3726-3731
32. Nichols, R. J., Sen, S., Choo, Y. J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K. M., Lee, K. J., Wong, A., Shales, M., Lovett, S., Winkler, M. E., Krogan, N. J., Typas, A., and Gross, C. A. (2011) Phenotypic landscape of a bacterial cell. *Cell* **144**, 143-156
33. Johansson, D., Jessen, C. H., Pohlsgaard, J., Jensen, K. B., Vester, B., Pedersen, E. B., and Nielsen, P. (2005) Design, synthesis and ribosome binding of chloramphenicol nucleotide and intercalator conjugates. *Bioorg Med Chem Lett* **15**, 2079-2083

4. Biochemical characterisation of a thermophilic 2OG oxygenase ycfD$_{RM}$

Contents

4. Biochemical characterisation of a thermophilic 2OG oxygenase ycfD$_{RM}$ 128

4.1. Biochemical characterisation of a thermophilic 2OG oxygenase ycfD$_{RM}$ 130

4.1.1. Aims 130

4.2. Bioinformatic search for ycfD homologues 131

4.2.1. Thermophilic bacterium *Rhodothermus marinus* contains a *ycfD* homologue 132

4.3. Ribosomal protein analysis of *Rhodothermus marinus* 133

4.3.1. Large scale growth 133

4.3.2. Lysis of *Rhodothermus marinus* cells 134

4.3.3. LC-MS analysis of *Rhodothermus marinus* ribosomal proteins 135

4.3.4. L16$_{RM}$ is hydroxylated on R82 139

4.4. Cloning of *ycfD$_{RM}$* from *Rhodothermus marinus* gDNA 141

4.4.1. Regional order neural network analysis of ycfD$_{RM}$ 141

4.4.2. Codon usage and GC-content of *ycfD$_{RM}$* 142

4.4.3. Optimisation of PCR conditions for *ycfD$_{RM}$* amplification from gDNA 143

4.5. Expression and purification of ycfD$_{RM}$ 144

4.6. Activity studies of recombinant ycfD$_{RM}$ 147

4.6.1. YcfD$_{RM}$ hydroxylates L16$_{RM}$ fragment peptide *in vitro* 147

4.6.2. MS/MS analysis confirms hydroxylation occurs on R82 148

4.6.3. L16$_{RM}$ hydroxylation by ycfD$_{RM}$ is Fe(II) and 2OG-dependent..... 149

4.7. Kinetic studies on recombinant ycfD$_{RM}$..... 150

4.8. Investigations of the L16$_{RM}$ binding mode to ycfD$_{RM}$ 153

4.8.1. YcfD$_{RM}$ catalysis of cyclic L16 peptides 153

4.9. YcfD substrate selectivity..... 156

4.9.1. Identification of L16 residues conferring ycfD$_{RM}$ substrate selectivity 158

4.9.2. Alanine scanning study confirms E^{-2} as the key residue for ycfD$_{RM}$ catalysis..... 159

4.9.3. L16$_{EC}$ A79E variant improves hydroxylation kinetics with ycfD$_{EC}$..... 161

4.9.4. Studying ycfD$_{RM}$ substrate selectivity *in vivo*..... 162

4.10. Biochemical comparison of ycfD$_{EC}$ and ycfD$_{RM}$: biocatalytic potential..... 164

4.10.1. Primary sequence comparison 165

4.10.2. Comparison of enzymatic activity and stability..... 166

4.10.3. Circular dichroism studies 169

4.10.4. Temperature dependence of activity..... 172

4.10.5. Co-factor thermostability studies..... 173

4.11. Oxygen dependence of ycfD$_{RM}$ catalytic rate 174

4.12. Conclusions..... 176

4.13. References..... 177

4.1. Biochemical characterisation of a thermophilic 2OG oxygenase ycfD$_{RM}$

Chapter 2 described studies on activity of ycfD as a ribosomal hydroxylase catalyzing arginyl hydroxylation of *E. coli* L16. The identification of a novel class of 2OG oxygenase substrates raised questions about the structural aspects of ribosome hydroxylation. Co-crystallisation attempts with the *E. coli* L16 fragment proved unsucessful and it was hypothesised that the affinity for the substrate was too low to allow trapping of the substrate in the enzyme active site under crystallisation conditions.

Structural understanding of ycfD substrate recognition and arginyl catalysis was therefore pursued using alternative approaches. Two main approaches were considered: i) chemical cross-linking of the substrate with the enzyme and ii) identification of a homologous enzyme-substrate pair with a higher binding affinity. The first approach has been used successfully on 2OG oxygenases, for example to determine the structure of ABH3 in complex with double-stranded DNA (1). However, the approach is laborious, requires investigation of many variants of both the enzyme and the substrate and may yield artefactual results stemming from non-specific interactions between chemically reactive substrates and enzyme. The second approach relies on the existence of homologous enzyme/substrate pairs, with similar substrate specificity but different biophysical properties, which may lead to a different crystallographic behaviour. Thermophilic organisms are most likely to contain enzymes with high affinity for the substrate, because their catalysis has to proceed under conditions where either elevated temperature or high salt concentrations interfere with the binding process; thermophilic enzymes are therefore characterised by a low K_m^{app} for the substrate and high k_{cat}/K_m^{app} ratios (2). A bioinformatic analysis was therefore required to search for thermophilic/halophilic organisms whose genomes contain putative homologoues of ycfD.

4.1.1. Aims

In order to gain structural understanding of ribosomal protein hydroxylation by JmjC-domain containing 2OG oxygenases, the work described in Chapter 4 sets out to address the following objectives:

i) conduct a bioinformatic search of ycfD conservation to identify homologues with significantly different (extremophilic) biophysical properties amenable to crystallisation;

ii) clone, purify and express a ycfD homologue(s) with extremophilic character;

iii) characterise a ycfD extremophilic homologue(s) biochemically;

iv) identify suitable enzyme-substrate pair(s) for co-crystallisation of ycfD in complex with substrate.

4.2. Bioinformatic search for ycfD homologues

A BLAST search was undertaken with the *E. coli* ycfD protein sequence as a search query against all prokaryotic phyla; sequences with >35% identity were deemed as homologous whereas those with 10-35% identity were considered as potentially homologous (human NO66 and MINA53 are 8% identical to ycfD). Sequences with >30% identity were identified in phyla *Proteobacteria* and *Bacteroidetes*. *Actinobacteria* and α-*Proteobacteria* were found to contain partially homologous sequences (12/25% identity/similarity with *E. coli* ycfD was found in α-*Proteobacteria* and 12/27% in *Actinobacteria*). Despite their relatively high overall homology with ycfD, protein sequence alignments of the putative DSBH domains revealed greater overall homology with NO66 (Figure 4.1) and were thus labelled as potential NO66 homologues. The search was then extended to simple eukaryotes; chlorophytes *Micromonas pusilla* and *Ostreococcus tauri/marinus* were found to contain two distinct genes coding for proteins sharing homology with ycfD and with human NO66, respectively (*O. tauri* contains a gene with 24/40% identity/similarity with *E. coli* ycfD and 7/15% with human NO66 (gi|308807571) and a gene with 15/24% identity/similarity with human NO66 and 9/21% with *E. coli* ycfD (gi|308805368O) (Figure 4.1). No ycfD homologues were found in other eukaryotes, suggesting that ycfD function is redundant in higher eukaryotes and was lost throughout evolution. Over 90% of all ycfD homologues were identified in *Proteobacteria*, in particular in γ-*Proteobacteria* which shared a high degree of homology (>90%) with *E. coli* ycfD but were not considered of significant interest.

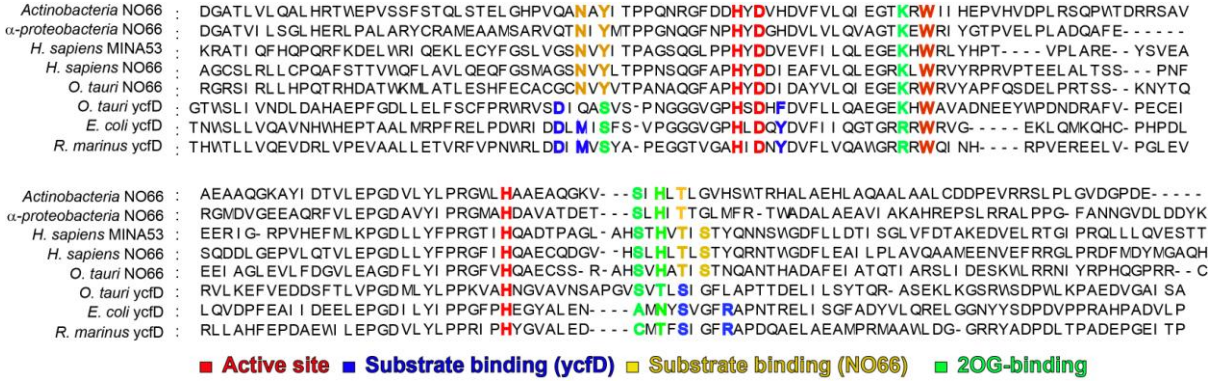


Figure 4.1: Protein sequence alignment of ycfD and NO66 homologues from *Actinobacteria* (*Arthrobacter aurescens*, gi: 4637841**); α-*Proteobacteria* (*Janaschia sp.*, gi:** 3932826**); *H. sapiens* MINA53 (gi:** 84864**); *H. sapiens* NO66 (gi:** 79697**); *O. tauri* NO66 (gi: 308805368); *O. tauri* ycfD (gi: 308807571); *E. coli* ycfD (gi: 268315578); *R. marinus* ycfD (gi: 268315578).**

4.2.1. Thermophilic bacterium *Rhodothermus marinus* contains a *ycfD* homologue

Of particular interest were homologues identified in two species from phylum *Bacteroidetes* which possess extremophilic traits: a halophile *Salinibacter ruber* and a thermo- and halophile *Rhodothermus marinus.* The putative ycfD homologue from *R. marinus* (henceforth labelled ycfD$_{RM}$) showed 38% identity with *E.coli* ycfD (henceforth labelled ycfD$_{EC}$) as well as homology with JmjC-domain containing 2OG oxygenases FIH, NO66 and MINA53 (Figure 4.2). As the motivation behind searching for extremophilic homologues of ycfD was large-scale production of the protein for crystallographic studies, as well as demonstrating that the enzyme catalyses ribosomal hydroxylation, two factors were considered when selecting the organism for further studies: i) likelihood of thermophilic properties and ii) ease of laboratory culturing of the organism. The ycfD homologues from both *S. ruber* and *R. marinus* were likely to possess biochemically interesting properties due to the extereme living environments of the organisms and both therefore fulfilled condition i). However, the analysis of available literature suggested that *S. ruber* has only been cultured in small quantities, which suggested that large-scale culture may be technically challenging (3); *R. marinus* culturing, however, has been performed on a large scale (4), suggesting that it should be feasible to culture the organism in our laboratory at a scale required for analysis of ribosomal proteins. *R. marinus* (strain DSM 4525) was therefore aquired from the American Type Culture Collection (ATCC) and used for extraction of genomic DNA (gDNA) and further biological studies.

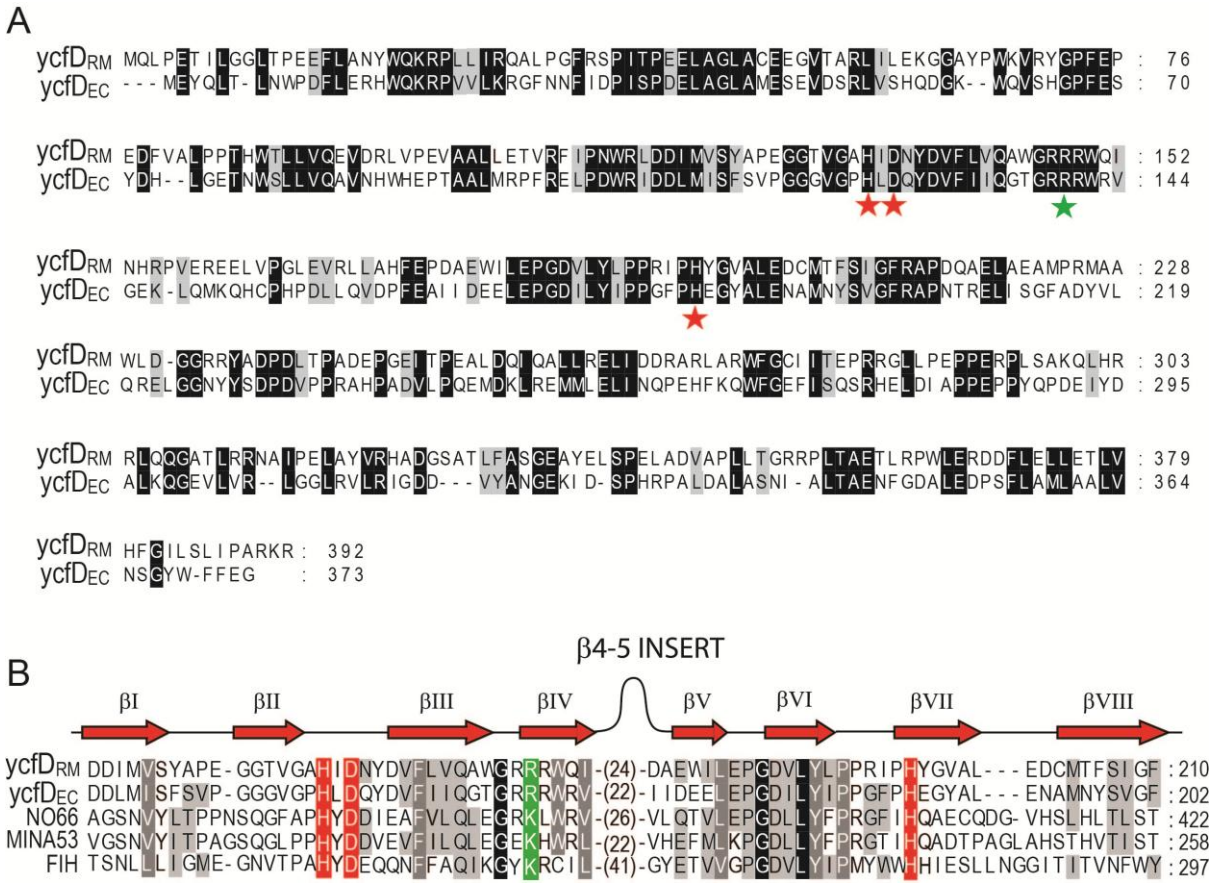


Figure 4.2: Protein sequence alignments of ycfD$_{RM}$/ycfD$_{EC}$, of ROX and FIH from *R. marinus*, *E. coli* and *H. sapiens* suggest that all characteristic 2OG oxygenase features are conserved in the putative 2OG oxygenase ycfD$_{RM}$ from *R. marinus*. ClustalW was used to generate protein sequence alignments, which were edited in GeneDoc 2.7. A) Protein sequence alignment of ycfD$_{RM}$ and ycfD$_{EC}$. Red stars indicate Fe(II)-binding residues and the green star indicates the basic residue that binds the 2OG C-5 carboxylate. B) Protein sequence alignment of the double-stranded beta-helix (DSBH)-containing JmjC domains of prokaryotic and human ROX and FIH. Secondary structure assignment is according to the crystal structure of FIH (PDB: 1H2K, (5)). Red arrows denote β-sheets with Roman numerals indicating the strand numbering within the DSBH.

4.3. Ribosomal protein analysis of *Rhodothermus marinus*

R. marinus is a thermo- and halophilic, obligate aerobe, first isolated from submarine alkaline hot-springs in Iceland (6). Its optimal growth at 65°C is significantly higher than for all other organisms with characterized 2OG oxygenases. A number of thermophilic enzymes have been studied from *R. marinus*, including DNA polymerases and extremely thermostable cellulases which are of significant biotechnological interest (7).

4.3.1. Large scale growth

Growth of *R. marinus* was induced in ATCC #1599 broth from glycerol stock. The organism was grown in a 50 mL Falcon tube floating sideways (in order to maximise aeration of the liquid media) in

a water bath (70°C) for two days, until OD_{600} reached ~0.7 and no longer increased. The growth of 100 mL strong starter culture, induced by dilution of culture at maximum OD_{600} into ATCC #1599 broth, was monitored for 4 days, and a gradual increase in the intesity of pigmentation was observed with time (Figure 4.3).

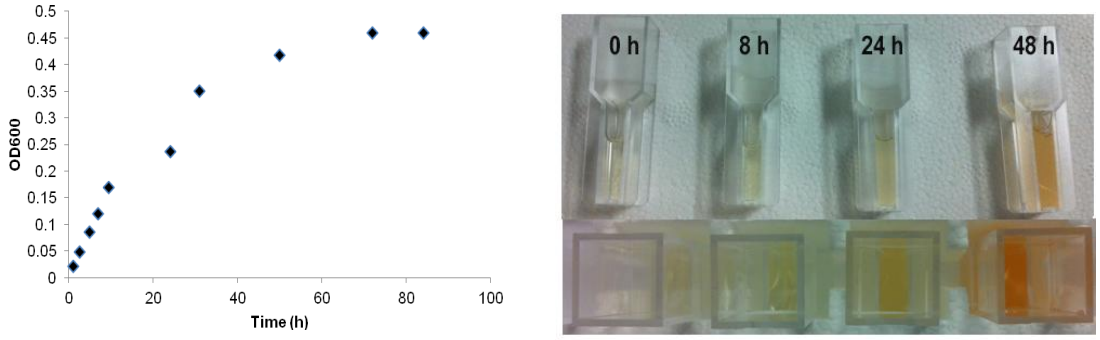


Figure 4.3: Growth curve and change of pigmentation of *R. marinus* grown in ATCC #1599 broth (100 mL) at 70°C in a water bath. *R. marinus* growth rate was measured by removing 1 mL aliquots at defined time-points and measuring OD_{600}. Pigmentation increased over time, correlating with the growth phase. Pigmentation did not change after 48 h.

When the strong starter culture reached maximum OD_{600}, a large scale growth was induced by 1:100 dilution into ATCC #1599 broth in two types of flasks: 600 mL PYREX Erlenmeyer flasks, or 1 L plastic baffled flasks, which were predicted to provide better aeration. They were grown simultaneously in an Innova44 shaking incubator (New Brunswick Scientific) overnight. In separate batches, cultures were grown at 60, 65, 70, 75, 77 and 80°C. It was observed that glass flasks supported growth up to 75°C, whereas plastic baffled flasks supported growth up to 80°C.

4.3.2. Lysis of *Rhodothermus marinus* cells

In order to lyse the cell paste, two methods were attempted. Initially, sonication was tested using a Sonics Vibracell sonicator. After centrifugation of the lysate, the supernatant was clear and colourless, suggesting that no lysis had occurred. It was therefore assumed that the standard sonication technique is not adequate for lysis of *R. marinus*, which contains a triple cell wall (6).

The second attempted lysis procedure used a bead-beater tissue homogeniser Precellys-24 (Bertin Technologies), which is typically used for disruption of yeast cells. Two types of beads were tested: glass beads (Precellys Glass Kit 0.1/0.5 mm in 7 mL container) and ceramic beads (Precellys Ceramic

Kit, 2.8 mm in 7 mL container). As shown in Figure 4.4, glass beads (samples 1 and 2) achieved a higher degree of lysis compared to ceramic beads (sample 3), as judged by the intensity of supernatant coloration after centrifugation. 6 cycles of lysis (sample 2) achieved more thorough lysis than 3 cycles (sample 1) and were considered optimal for further lysis.

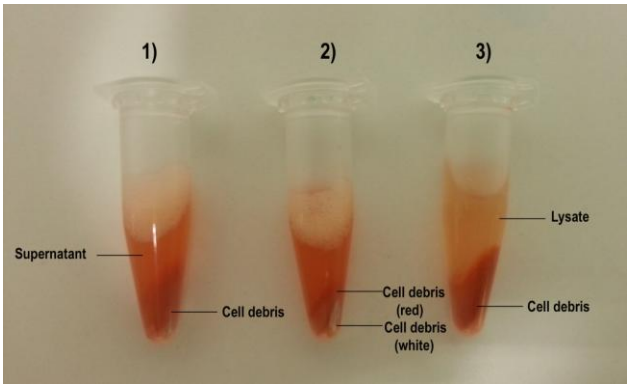


Figure 4.4: *R. marinus* lysis using bead-beating. Two types of beads were tested: glass beads (samples 1 and 2) and ceramic beads (sample 3). A bead-beater tissue homogeniser (Precellys 24; Bertin Technologies) was used with two types of beads: glass beads (Precellys Glass Kit 0.1/0.5 mm in 7 mL container) and ceramic beads (Precellys Ceramic Kit, 2.8 mm in 7 mL container). The shaking parameters were 6,500 rpm in 20 s bursts, followed by a cooling period of 1 min on ice. 1 mL aliquots were removed from the sample and centrifuged at 14,000 rpm for 3 min to assess the degree of lysis. Glass beads (1 and 2) achieved a higher degree of lysis compared to ceramic beads (3). 6 cycles of lysis (sample 2) achieved more thorough lysis than 3 cycles (sample 1).

4.3.3. LC-MS analysis of *Rhodothermus marinus* ribosomal proteins

Intact-protein MS analysis of *R. marinus* ribosomal proteins used a modification of the UPLC-ESI-MS methodology described in Chapter 2 (Figure 4.5). Notably, leucine enkephalin was used for on-line correction of mass, achieving superior mass accuracy compared to external myoglobin calibration used in previous chapters. Out of 54 predicted ribosomal proteins in *R. marinus*, 48 were identified based on intact masses and their identities confirmed by MS/MS (Tables 1-2).

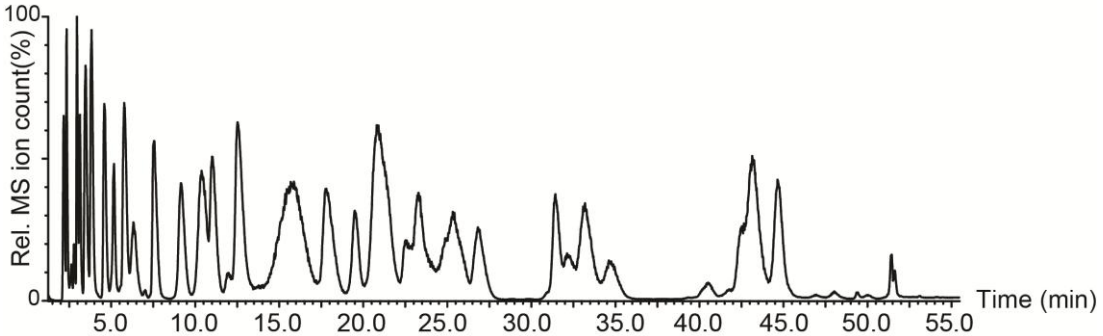


Figure 4.5: UPLC chromatogram of *R. marinus* ribosomal proteins. Isolation was achieved using bead-beating lysis procedure followed by sucrose density centrifugation and acetone precipitation. ESI-TOF-MS was used to determine intact masses of *R. marinus* ribosomal proteins, which were compared with predicted masses based on assigned Uniprot entries.

Table 4.1: ESI-TOF-MS deconvoluted masses of *R. marinus* 30S ribosomal proteins. Column 1 lists the ribosomal proteins analysed; column 2 lists predicted unmodified masses based on DNA sequence information; column 3 lists experimentally observed masses for individual proteins; column 4 lists calculated differences between predicted and experimental masses; column 5 indicates a prediction of *N*-terminal methionine loss (NML) ('Yes' if Δ(Da) is between 129-132 Da or a composite thereof, 'No' otherwise); column 6 indicates differences between predicted and observed masses after accounting for NML; column 7 indicates predicted post-translational modifications based on the calculated mass differences; column 8 lists % sequence coverage by MS/MS-based peptide identification after partial trypsin digestion.

1	2	3	4	5	6	7	8
Protein	Sequence mass (Da)	Experimental mass (Da)	Δ (Da)	Methionine lost?	ΔNML (Da)	PTM	% MS/MS coverage
S1	80651.5	n.d.	n.a.	n.a.	n.a.	n.a.	49.9
S2	31152.8	n.d.	n.a.	n.a.	n.a.	n.a.	89.5
S3	28438.8	n.d.	n.a.	n.a.	n.a.	n.a.	84.3
S4	23726.6	23597.0	-129.6	Yes	1.6		91.0
S5	19516.5	19384.9	-131.6	Yes	-0.4		91.2
S6	17188.2	17056.8	-131.4	Yes	-0.2		83.3
S7	17847.7	17847.3	-0.4	No	-0.4		95.5
S8	14960.3	14828.8	-131.5	Yes	-0.3		95.5
S9	15219.6	15160.1	-59.5	Yes	72	n.d.	74.4
S10	11500.5	11369.0	-131.5	Yes	-0.3		93.1
S11	13979.9	13863.4	-116.5	Yes	14.7	Me	42.6
S12	13757.1	13624.2	-132.9	Yes	-1.7		74.2
S13	14462.9	14331.1	-131.8	Yes	-0.6		78.6
S14	10454.4	10322.8	-131.6	Yes	-0.4		80.9
S15	10671.4	10671.0	-0.4	No	-0.4		83.0
S16	23179.8	23048.2	-131.6	Yes	-0.4		66.8
S17	11986.9	11855.2	-131.7	Yes	-0.5		75.2
S18	7595.9	7464.0	-131.9	Yes	-0.7		55.4
S19	10891.7	10760.1	-131.6	Yes	-0.4		58.5
S20	10008.8	9877.4	-131.4	Yes	-0.2		29.8
S21	8251.6	8119.9	-131.7	Yes	-0.5		40.3

Table 4.2: ESI-TOF-MS deconvoluted masses of *R. marinus* 50S ribosomal proteins. Column 1 lists the ribosomal proteins analysed; column 2 lists predicted unmodified masses based on DNA sequence information; column 3 lists experimentally observed masses for individual proteins; column 4 lists calculated differences between predicted and experimental masses; column 5 indicates a prediction of *N*-terminal methionine loss (NML) ('Yes' if Δ(Da) is between 129-132 Da or a composite thereof, 'No' otherwise); column 6 indicates differences between predicted and observed masses after accounting for NML; column 7 indicates predicted post-translational modifications based on the calculated mass differences. column 8 lists % sequence coverage by MS/MS-based peptide identification after partial trypsin digestion.

1	2	3	4	5	6	7	8
Protein	Sequence mass (Da)	Experimental mass (Da)	Δ (Da)	Methionine lost?	Δ NML (Da)	PTM	% MS/MS coverage
L1	25877.8	25746.1	-131.7	Yes	-0.5		80.9
L2	30710.3	30579.0	-131.3	Yes	-0.1		96.0
L3	22999.7	22868.0	-131.7	Yes	-0.5		94.8
L4	24128.5	24127.5	-1.0	No	-1.0		98.6
L5	21822.5	21690.5	-132.0	Yes	-0.8		99.5
L6	20273.6	20142	-131.6	Yes	-0.4		95.1
L7	12988.9	12857.1	-131.8	Yes	-0.6		96.0
L9	18216	n.d.	n.a.	n.a.	n.a.	n.a.	83.0
L10	18722.7	18591.5	-131.2	Yes	0.0		89.2
L11	15635.3	15631.0	-4.3	Yes	126.9	9 Me	71.2
L13	16703.4	16703.1	-0.3	No	-0.3		98.0
L14	13475.6	13475.2	-0.4	No	-0.4		100.0
L15	16420.9	16420.5	-0.4	No	-0.4		78.4
L16	16275.3	16306.0	30.7	No	30.7	Me,OH	84.7
L17	24984.9	24969.0	-15.9	No	-15.9		37.6
L18	13512.6	13512.1	-0.5	No	-0.5		81.5
L19	14247.4	14115.9	-131.5	Yes	-0.3		79.0
L20	13561.8	13430.5	-131.3	Yes	-0.1		59.1
L21	13416.6	13416.7	0.1	No	0.1		92.4
L22	13588.9	13588.6	-0.3	No	-0.3		91.6
L23	11506.5	11374.9	-131.6	Yes	-0.4		87.0
L24	13577.7	13446	-131.7	Yes	-0.5		66.1
L25	22065.1	n.d.	n.a.	n.a.	n.a.	n.a.	64.4
L27	9295.6	9164.0	-131.6	Yes	-0.4		57.0
L28	9054.6	8923.0	-131.6	Yes	-0.4		69.2
L29	8571.0	8570.6	-0.4	No	-0.4		63.4
L30	7733.1	7601.6	-131.5	Yes	-0.3		66.7
L31	9497.7	9497.4	-0.3	No	-0.3		100.0
L32	7748.1	7612.7	-135.4	Yes	-4.2		66.2
L33	6687.8	6556.2	-131.6	Yes	-0.4		80.7
L34	6228.4	6090.7	-137.7	Yes	-6.5		34.6
L35	7907.5	7775.9	-131.6	Yes	-0.4		51.5
L36	4523.3	n.d.	n.a.	n.a.	n.a.	n.a.	26.3

The observed mass of S11$_{RM}$ was 13,863.4±0.4 Da, consistent with a single methylation and NML, as observed in *E. coli* where S11$_{EC}$ has been shown to contain an *N*-terminally mono-methylated alanine

residue (Table 4.3) (8). The mass of S12$_{RM}$ was 13624.2 Da, consistent with NML. β-Methylthiolation of S12$_{RM}$, present in S12$_{EC}$ (9), was not observed. The observed mass of L11$_{RM}$ was 15,631±0.4 Da, consistent with addition of 9 methyl groups. In *E. coli*, these modifications were found to be catalysed by PrmA (10), a putative homologue of which is present in *R. marinus*. In *E. coli*, A2, K4 and K40 are trimethylated (10), MS/MS analysis identified putative methylation sites on K3 and K4 (data not shown); the third trimethylation site was not identified. Whilst in *E. coli* the proteins L7/L12 are K82-monomethylated in response to cold-shock (11), no methylation was observed in *R. marinus*, possibly as a result of its growth at high temperature, nor was L7 $_{RM}$ acetylated to L12 $_{RM}$. Two forms of L32$_{RM}$ were found, one unmodified, another 32 Da larger than predicted, consistent with two oxidative modifications. MS/MS analysis suggested methionine oxidations on M32$_{RM}$ and M40$_{RM}$ (Figure 4.6). The functional significance of this modification is unclear.

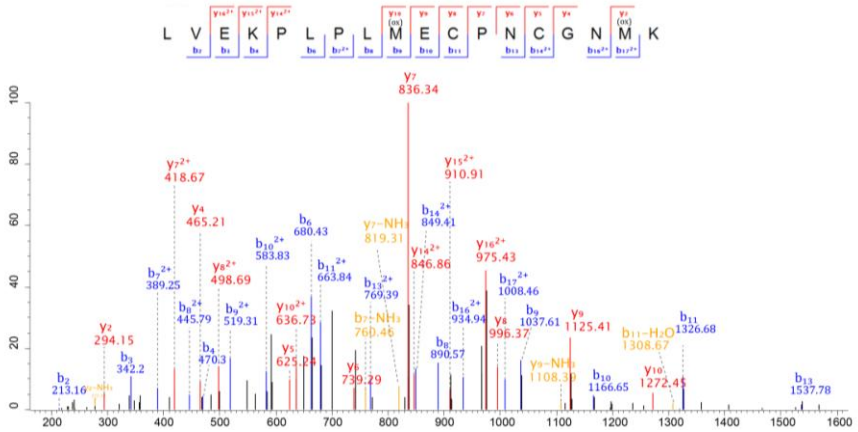


Figure 4.6: MS/MS spectrum of doubly oxidized trypsin-digested fragment from L32 reveals hydroxylation on M32 and M40. Fragmentation method: CID, Maxquant score 139. Figure kindly provided by Jacob Bush.

Table 4.3: Summary of ribosomal protein post-translational modifications observed in *E. coli* and *R. marinus* ribosomal proteins.

Ribosomal protein (30S)	*E. coli*	*R. marinus*	Ribosomal Protein (50S)	*E. coli*	*R. marinus*
S5	Ac	-	L3	Me	-
S9	-	unknown	L7	Ac/Me	-
S11	Me	Me	L11	9 × Me	9 × Me
S12	β-Methylthiolation	-	L12	Me	Not observed
S18	Ac	-	L16	2 × OH, Me	OH, Me
			L32	-	2 × OH
			L33	Me	-

4.3.4. L16$_{RM}$ is hydroxylated on R82

Consistent with the presence of a *ycfD* homologue in the *R. marinus* genome, two co-eluting proteins were observed at retention time 25.3 min with masses 16,289 and 16,306 Da, assigned as unhydroxylated and hydroxylated L16$_{RM}$ (Figure 4.7A). MS/MS analysis of trypsin-digested ribosomal proteins identified the hydroxylation site as R82 of L16$_{RM}$ (Figure 4.7C), analogous to L16$_{EC}$ R81, which is the hydroxylation substrate of ycfD$_{EC}$; oxidative modification of M83 was ruled out by MS/MS. The mass of unhydroxylated L16$_{RM}$ is consistent with monomethylation, presumably on the *N*-terminal methionine-residue (the location of the methyl group could not be determined; in *E. coli*, L16 is methylated on *N*-terminal methionine (12) and the *N*-terminal fragment of L16$_{RM}$ was not observed by MS/MS).

In contrast to *E. coli*, where L16$_{EC}$ hydroxylation was complete at optimal growth temperature (37°C), incomplete L16$_{RM}$ hydroxylation was observed at optimal growth temperature of *R. marinus* (70°C). It was noticed that the extent of L16$_{RM}$ hydroxylation differed depending on what type of flask was used for *R. marinus* growth. Because baffled and unbaffled flasks lead to different oxygenation of growth media (13) and oxygen solubility decreases with temperature (14), it was investigated whether L16$_{RM}$ hydroxylation may be affected by aeration or temperature.

As expected from better oxygenation of the media, hydroxylation was higher in baffled than in unbaffled flasks (65% and 35%, respectively). When *R. marinus* was cultured between 60-80°C in both types of flasks, an approximately linearly decreasing level of L16$_{RM}$ hydroxylation was observed (Figure 4.7B): in baffled flasks L16$_{RM}$ hydroxylation decreased from 80-60%; in unbaffled flasks the hydroxylation decreased from 60-25% as the growth temperature increased from 60-75°C. These results suggested that the hydroxylation is not limited by temperature itself but by a temperature-dependent factor. Furthermore, it presented an indication that the enzyme catalysing the hydroxylation, assumed to be the homologue of ycfD$_{EC}$, may act as an oxygen sensor.

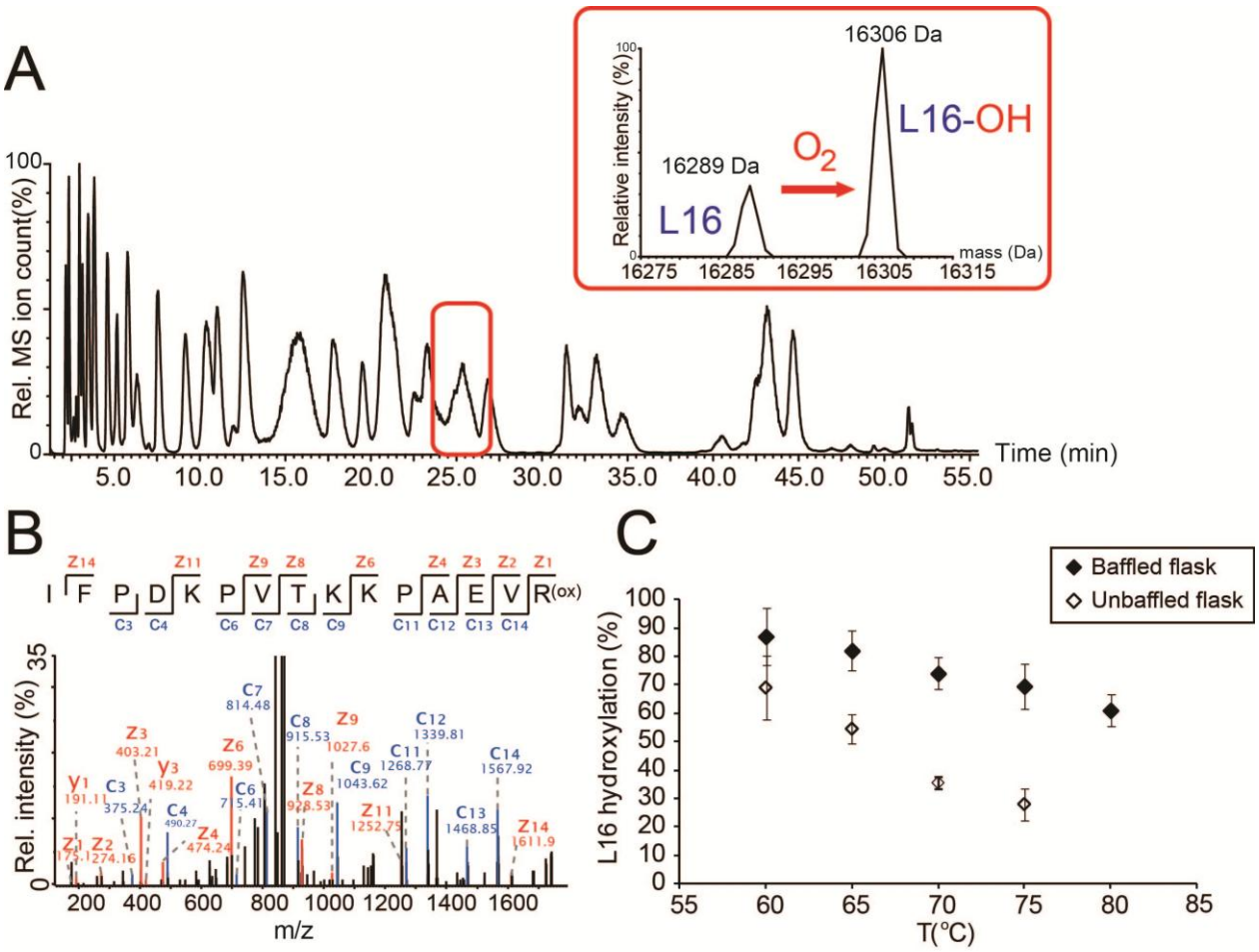


Figure 4.7: The hydroxylation of L16$_{RM}$ occurs on residue R82 and varies with temperature. A) The deconvoluted ESI-MS spectrum (inset) from the highlighted peak eluting at ~25 min, revealed partial hydroxylation of L16$_{RM}$ (predicted mass 16275 Da). L16$_{RM}$ also undergoes mono-methylation. B) MS/MS spectrum of trypsin-digested modified peptide fragment ^{68}IFPDKPVTKKPAEVR^{82} from L16$_{RM}$ reveals hydroxylation on R82. Fragmentation by ETD; Maxquant score 201. C) Hydroxylation of L16$_{RM}$ decreases with increasing growth temperature and varies for growth in baffled and unbaffled flasks. L16$_{RM}$ whole-protein mass was determined by ESI-MS spectrometry. Intensity of MS signals corresponding to unhydroxylated and hydroxylated L16$_{RM}$ were used to quantify the extent of L16$_{RM}$ hydroxylation. The average of two independent MS experiments is shown with error bars denoting s.d. from the mean.

4.4. Cloning of *ycfD$_{RM}$* from *Rhodothermus marinus* gDNA

To investigate whether the observed hydroxylation of L16$_{RM}$ was catalysed by ycfD$_{RM}$, the enzyme was then biochemically characterized *in vitro*. The *ycfD$_{RM}$* gene was cloned from *R. marinus* gDNA and ligated into a plasmid vector suitable for expression in *E. coli*. For ease of purification, a vector containing either a hexa-histidyl tag or a glutathione S-transferase (GST) tag, to enable purification using Ni- or GST-affinity chromatography, respectively, would be suitable.

4.4.1. Regional order neural network analysis of ycfD$_{RM}$

A number of 2OG oxygenases have been found to contain flexible or disordered regions which interfere with correct folding and lead to protein aggregation during purification procedures (Dr. M. McDonough, personal communication). In order to avoid protein misfolding and aggregation, soluble proteins are produced by truncation of disordered regions, which are not crucial for catalytic activity.

A theoretical prediction of relative disorder of ycfD$_{RM}$ secondary structure was undertaken using regional order neural network (RONN) software (https://app.strubi.ox.ac.uk/RONN/) to identify regions with a high degree of disorder. The results were compared to an analogous prediction for ycfD$_{EC}$, as well as a putative *ycfD* homologue from a psychrophilic ('cold-loving') bacterium *Psychrobacter arcticus*: ycfD$_{PA}$ (Figure 4.8). Overall, ycfD$_{RM}$ demonstrates a lower degree of structural disorder than ycfD$_{EC}$, with the exception of residues 300-320, where the disorder is greater by 0.1-0.25 units of disorder. Previous work on ycfD suggested that the *C*-terminal domain of ycfD (residues 297-373) is crucial for catalytic activity (Maggie Ho, DPhil thesis, University of Oxford); it was therefore decided that despite the relatively higher degree of disorder of that region of *ycfD$_{RM}$*, the gene would be cloned in its entirety, including the *C*-terminal domain (residues 1-392).

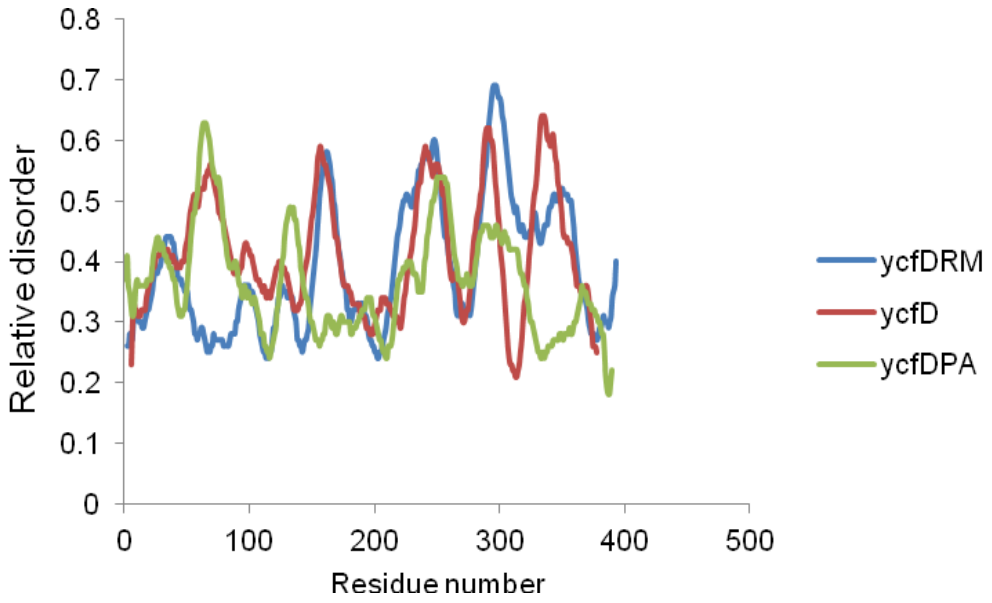


Figure 4.8: Theoretical prediction of relative secondary structure disorder, based on primary sequence of three ycfD homologues from a thermophilic, mesophilic and psychrophilic bacterium. Blue: ycfD$_{RM}$ (*R. marinus)*, red: ycfD$_{EC}$ (*E. coli*), green: ycfD$_{PA}$ (*P. arcticus*).

4.4.2. Codon usage and GC-content of *ycfD$_{RM}$*

A common issue with successful protein expression in bacteria is codon usage. Organisms vary widely in their use of codons for expression of particular amino acid residues, therefore production of a heterologous protein can benefit from 'codon-optimisation', which involves replacing codons in the gene of interest which are rarely used in a target organisms, with ones that the target organisms uses at a higher frequency. In *ycfD$_{EC}$*, 65% of codons are from the most commonly (>90%) used set of codons and only 5% of codons have a lower than 30% occurrence in *E. coli*. In *ycfD$_{RM}$*, 49% of codons are from the commonly used codon set (>90% usage), whereas 12% are from the rarely used set (>30% usage, Figure 4.9). This proportion of rare codons may be limiting for successful expression in *E. coli* and an empirical determination of expression levels was required to assess whether codon-optimisation would be required for successful expression.

Another important consideration for successful cloning procedure is the GC-content of a gene of interest. Genes with high GC-content are difficult to clone due to high secondary structure of both the gene and PCR primers, as well as non-specific binding of primers. The *ycfD$_{RM}$* gene has a GC-content of 68.0%, compared to 54.4% for *ycfD$_{EC}$* (Figure 4.9). This classifies it as GC-rich, requiring the use

of specialized DNA polymerases. Primers to amplify the gene of interest (forward: CACAGCCACGACGTTTACC; reverse: GGAATCACGTTGACCCAGTC) were specifically designed to achieve the lowest possible GC-content, by containing several nucleotides upstream and downstream of the gene of interest.

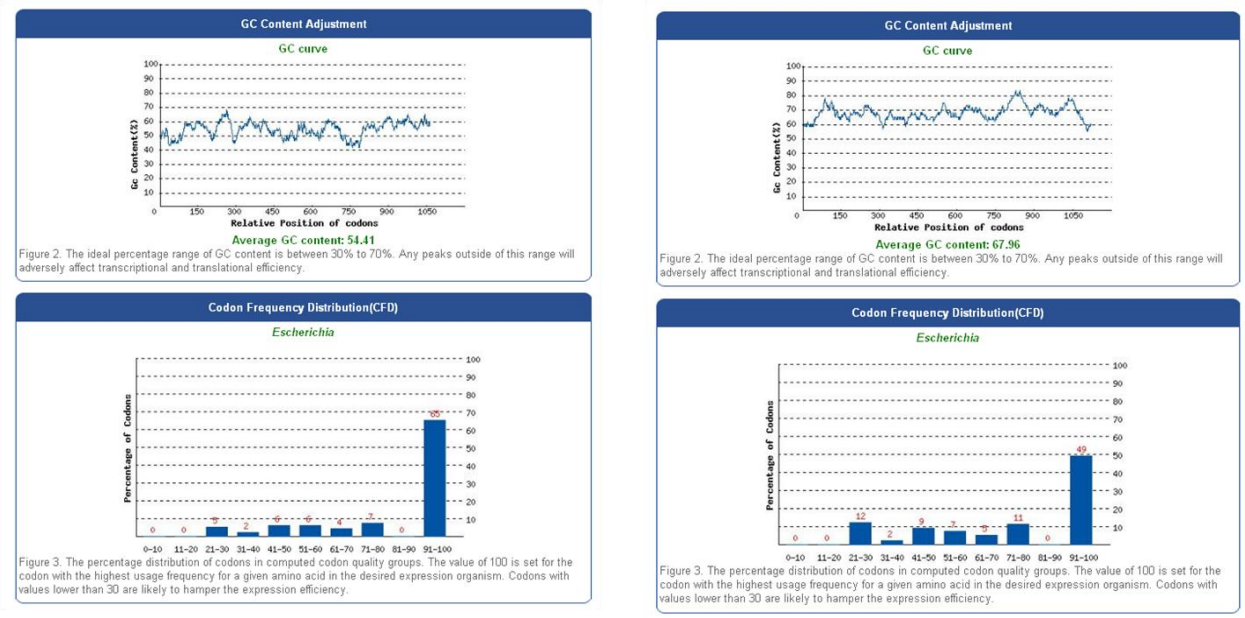


Figure 4.9: GC content (top) and codon frequency distribution (bottom) of *ycfD*$_{EC}$ and *ycfD*$_{RM}$ genes from *E. coli* (left) and *R. marinus* (right). The *ycfD*$_{RM}$ gene has a GC-content of 68.0%, compared to 54.4% for *ycfD*$_{EC}$. This classifies it as GC-rich. In *ycfD*$_{EC}$, 65% of codons are from the most commonly (>90%) used set of codons and only 5% of codons have a lower than 30% occurrence in *E. coli*. In the *ycfD*$_{RM}$ gene, 49% of codons are from the commonly used codon set (>90% usage), whereas 12% are from the rarely used set (>30% usage). Rare-codon analysis was performed using publically available software (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis).

4.4.3. Optimisation of PCR conditions for *ycfD*$_{RM}$ amplification from gDNA

Initially, standard laboratory DNA polymerases (Pfu Turbo and Taq) were used for cloning of *ycfD*$_{RM}$. Upon several failed attempts which tested the effects of template concentration (10-100 ng/mL), annealing temperature (60-67°C) and additives (0-10% DMSO) it was envisaged that specialised DNA polymerases which contain additives optimised for high GC content are required. BIO-X-ACT LongMix (BioLine) DNA polymerase was found to yield PCR products of desired molecular weight (Figure 4.10).

Several annealing temperatures were then tested: 60°C yielded products of the right size, but a high degree of unspecific amplification was observed; 67°C yielded a single band of the right molecular

weight, but the yield was low; 64°C afforded a strong, clean band at 1,500 bp, the expected molecular weight of the gene of interest (Figure 4.10b).

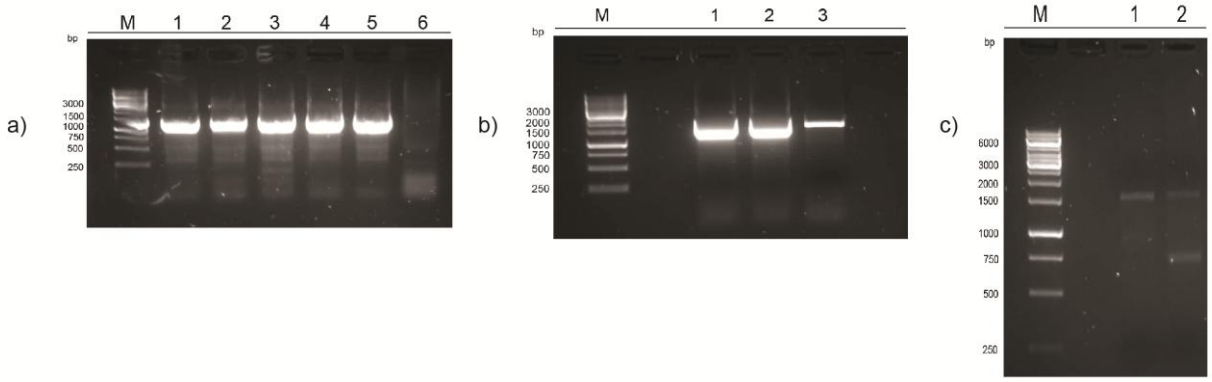


Figure 4.10: Optimisation of annealing temperature for PCR amplification of *ycfD$_{RM}$* from *R. marinus* gDNA. PCR reaction products from reactions with annealing temperatures of a) 60°C, b) 64°C, and c) 67°C; all other cycling parameters were constant.

The Bio-X-Act Long DNA amplification kit contains a Taq polymerase, which appends a single-nucleotide (thymine) overhang to the amplified stretch of DNA. This feature enables TA cloning, allowing the PCR product to be ligated into a vector with single nucleotide (adenine) overhang (e.g. pGEMT-Easy) immediately upon amplification from genomic DNA. This approach allows the products of PCR to be stored safely in a high-copy vector from which further manipulation is possible without the need for repeating costly PCR amplification. In order to produce high quantities of soluble protein, the PCR product of *ycfD$_{RM}$* amplification was then sub-cloned into pET28a and pGEX-6P-1 expression plasmids.

4.5. Expression and purification of ycfD$_{RM}$

The pET28a and pGEX-6P-1 expression plasmids containing *ycfD$_{RM}$* were transformed into *E. coli* strain BL21(DE3). Expression trials were conducted at 3 temperatures (18°C, 30°C and 37°C) incubated after IPTG-induction for either 18 h (18°C and 30°C) or 4 h (37°C); IPTG concentrations tested were 0.1 and 0.5 mM. The protein content of soluble and insoluble fractions was visualised by SDS-PAGE and relative quantities of protein bands corresponding to the correct molecular mass (44 kDa and 65 kDa for the pET28a and pGEX-6P-1 constructs, respectively) were assessed.

The pGEX-6P-1-ycfD$_{RM}$ construct afforded less soluble protein product compared to pET28A-ycfD$_{RM}$ construct. A band at ~65 kDa, corresponding to the ycfD$_{RM}$-GST fusion protein, was strongest for

cells incubated at 18°C in the presence of 0.1 mM IPTG overnight (Figure 4.11b). A relatively strong band at ~40 kDa likely corresponds to a truncated protein product lacking a GST-tag. Overall, ycfD$_{RM}$ production was highest when cells containing the pET28a-ycfD$_{RM}$ vector were incubated at 30°C overnight (Figure 4.11a) in the presence of 0.1 mM IPTG. This construct and these expression conditions were therefore used in all subsequent studies (Figure 4.11a).

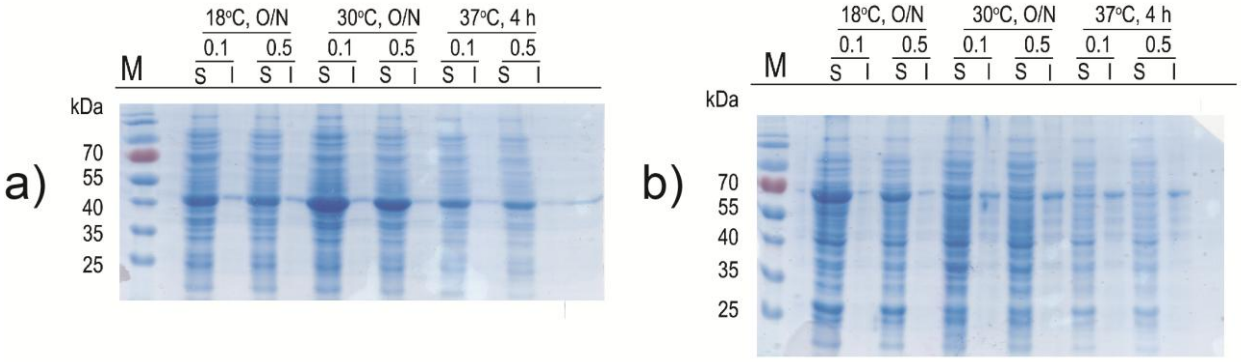


Figure 4.11: Expression of ycfD$_{RM}$ in a) pET28a and b) pGEX-6P-1. M: molecular weight marker; S: soluble fraction; I: insoluble fraction; 0.1/0.5: concentrations of IPTG used (in mM); expression temperature and duration after IPTG induction indicated. Predicted molecular mass of ycfD$_{RM}$ from pET28a and pGEX-6P-1 is 44.2 kDa and 65 kDa, respectively.

The *N*-terminal hexahistidyl tag enables the use of immobilised metal affinity chromatography (IMAC) procedures which rely on the strength of interaction between Ni^{2+} attached to the resin and a poly-histidyl tag attached to the protein of interest. The interaction is relatively strong and enables the trapping of a protein of interest from a cell lysate with high specificity. Protein is eluted from the column using a gradient of imidazole, which competes with the hexahistidyl tag for binding Ni^{2+} ions on the chromatographic resin.

YcfD$_{RM}$ was initially purified using a HisTrap HP (GE Healthcare) Ni^{2+}-affinity column attached to an AKTA purifier system (Amersham Pharmacia Biotech), eluted using a linear gradient of imidazole. Fractions containing high amounts of protein at ~40 kDa, consistent with the mass of ycfD$_{RM}$ (predicted at 43 kDa) were further purified using a size exclusion chromatographic column. SDS-PAGE gel of eluted fractions revealed a strong protein band at ~40 kDa (>85% purity), with small amounts of contaminant proteins at 25 kDa and 15 kDa; the sum of fragment masses approximately correspond to the parent mass (40 kDa), suggesting that degradation or site-specific cleavage may have occurred (Figure 4.12b).

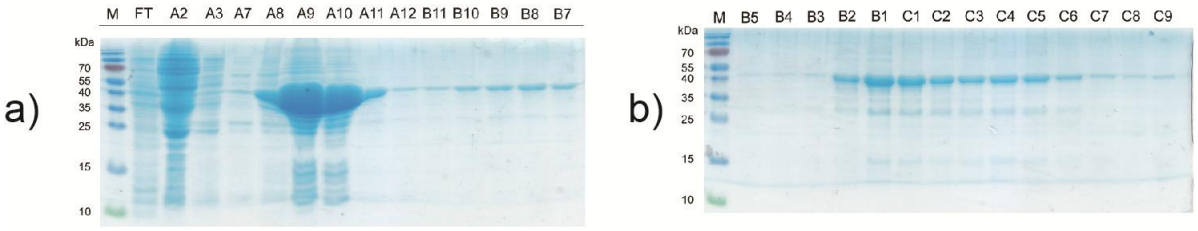


Figure 4.12: SDS-PAGE gel of ycfD$_{RM}$ purification. A) Cell lysate was applied onto a 5 mL HisTrap HP column (left) followed by b) size exclusion chromatographic separation of fractions A8-A11 using a 300 mL S75 size exclusion column (right). FT=flow-through. A2-C9: chromatographic fractions as labelled by AKTA-purification software.

As crystallisation studies were planned (see Chapter 5), it was envisaged that a higher degree of purity needed to be achieved in order to obtain diffracting crystals. Therefore, an alternative type of IMAC resin was tested, which is known to afford higher purity, albeit in lower yields. The tris(carboxymethyl)ethylene diamine (TED) resin chelates Ni^{2+} in a penta-coordinate fashion, thus leaving only one site available for binding of the poly-histidyl tag. The purification of ycfD$_{RM}$ using a Ni-TED column afforded >95% pure protein after the first round of purification (Figure 4.13). Although yields were lower, it was noted that cell lysate used for the TED purification contained a lower amount of ycfD$_{RM}$ (Figure 4.13, lane 2). After buffer exchange using a PD10 column a degree of degradation into a 25 kDa and 15 kDa fragments was again observed, but the contaminating fragments were removed during size-exclusion chromatography (Figure 4.13).

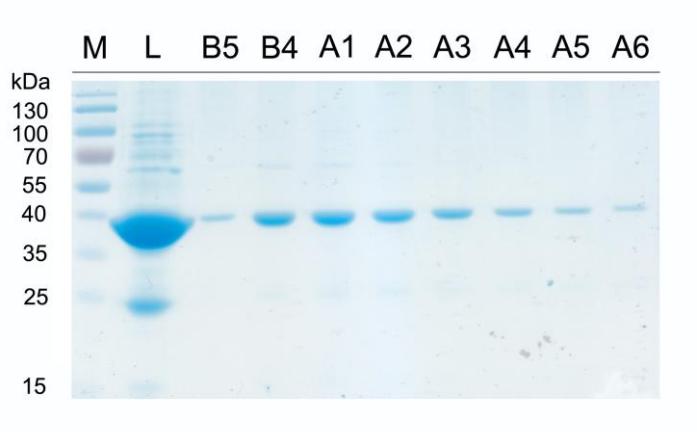


Figure 4.13: Purification of ycfD$_{RM}$ by Ni-TED and size exclusion chromatography. YcfD$_{RM}$ was initially purified using the Ni-TED column. After the washing step, the protein was eluted with 250 mM imidazole. Fractions containing large quantities of relatively pure protein were applied onto S75 size exclusion column (300 mL); the procedure successfully eliminated the contaminants at ~25 kDa. M: molecular weight marker; L: load; B5-A6: collected fractions.

4.6. Activity studies of recombinant ycfD$_{RM}$

YcfD$_{EC}$ was already characterised as a 2OG oxygenase using the radioactive 2OG decarboxylation assay and its target identified as R81 on ribosomal protein L16$_{EC}$ (15). It was therefore hypothesised that ycfD$_{RM}$ might be responsible for hydroxylation of the homologous ribosomal protein L16$_{RM}$. Should incubation of ycfD$_{RM}$ with L16$_{RM}$ substrate fragment in the presence of 2OG-oxygenase cofactors/co-substrates (Fe(II), 2OG, and ascorbate) result in an enzyme- and co-factor dependent shift of +16 Da, the enzyme could be classified as an active 2OG oxygenase. This approach avoids the use of the dangerous and laborious $^{14}CO_2$ decarboxylation assay, traditionally used upon identification of a novel oxygenase.

4.6.1. YcfD$_{RM}$ hydroxylates L16$_{RM}$ fragment peptide *in vitro*

YcfD$_{RM}$ was incubated with a L16$_{RM}$ substrate fragment (20-mer) in the presence of 2OG, Fe^{2+} and ascorbate. Incubation was carried out at 65°C based on the fact that the optimal temperature for *R. marinus* growth is 65-70°C. Incubation of the substrate fragment with ycfD$_{RM}$ and co-factors resulted in a +16 Da shift relative to incubation of the substrate fragment with co-factors alone, suggesting that ycfD$_{RM}$ indeed catalyses an oxidative modification of L16$_{RM}$ substrate fragment (Figure 4.14).

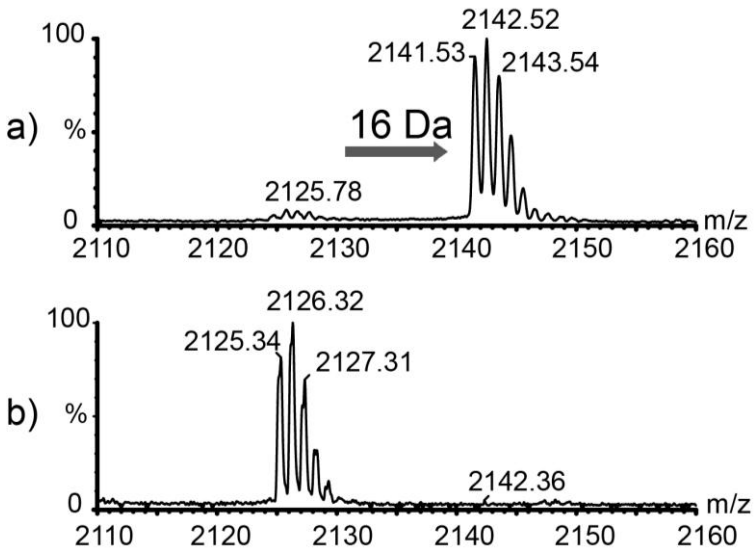


Figure 4.14: MALDI-MS spectrum of ycfD$_{RM}$-dependent hydroxylation of *R. marinus* L16 substrate fragment. L16$_{RM}$ substrate fragment (KPVTKKPAEVRMGKGKGSVE, 100 μM) was incubated with a) ycfD$_{RM}$ (1 μM), 2OG, (200 μM), Fe^{2+} (100 μM) and ascorbate (1 mM) or b) 2OG, Fe^{2+} and ascorbate alone for 30 min at 65°C. The reaction was quenched by addition of equal amount of CF_3COOH_{aq} (1% vol/vol) and analysed by MALDI-MS. A +16 Da mass shift is consistent with a ycfD$_{RM}$-dependent oxidative modification.

4.6.2. MS/MS analysis confirms hydroxylation occurs on R82

The L16 substrate fragment contains a 'VRM' sequence motif. Although ycfD-dependent hydroxylation has already been demonstrated to occur on arginine in *E.coli*, it could not be ruled out that the oxidative modification of the *R. marinus* sequence occurs on the neighbouring methionine. Methionines are prone to oxidation in the presence of reactive oxygen species (ROS) (16), and since the levels of these are presumably elevated at high temperatures (65°C), there is an increased likelihood of methionine oxidation. A tandem mass-spectrometry (MS/MS) study was undertaken to elucidate the site of modification. First, the L16$_{RM}$ substrate fragment was incubated for 1 h with 2OG, Fe^{2+} and ascorbate in the presence and absence of ycfD$_{RM}$. Samples were desalted before analysis on an Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, DE) by Jacob Bush, who also carried out all data-processing. The presence of z9, c11 and y10 ions suggested that ycfD$_{RM}$-dependent hydroxylation occurs on R82, and not M83 (Figure 4.15). YcfD$_{RM}$ is thus a *bona fide* 2OG oxygenase catalysing post-translational hydroxylation of R82 on ribosomal protein L16$_{RM}$.

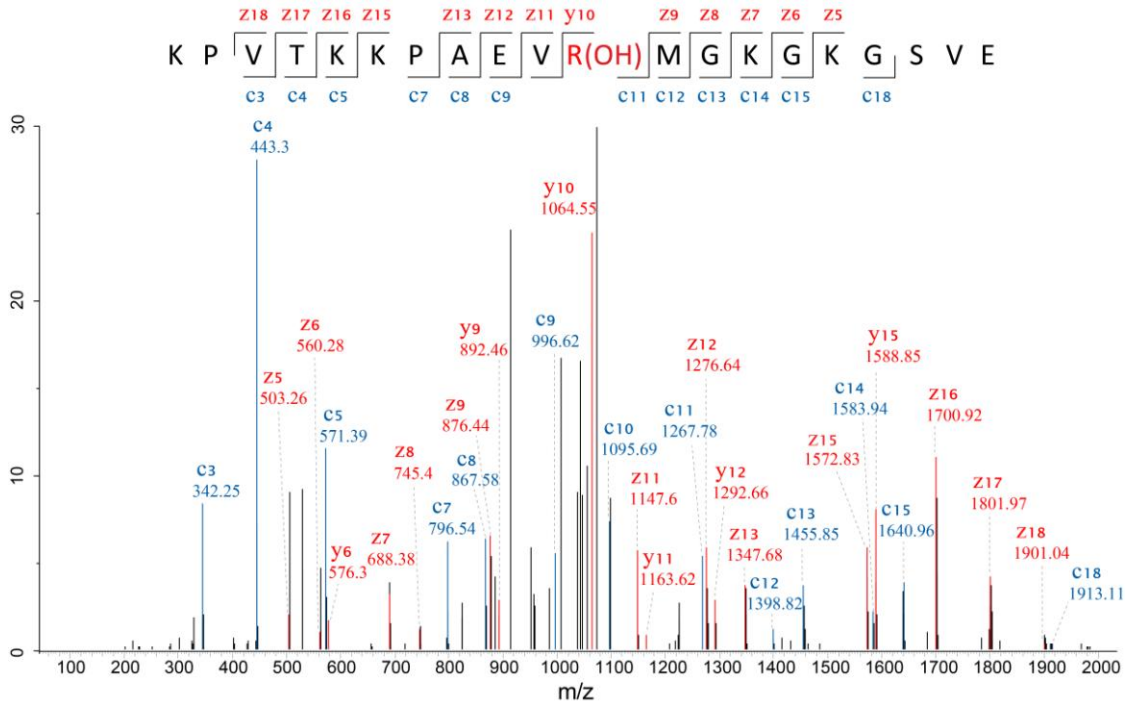


Figure 4.15: MS/MS (ETD) spectrum of hydroxylated L16$_{RM}$ fragment demonstrating hydroxylation of R82. Note that ETD ionization leads to an alternative mode of cleavage, producing c/z ions instead of the more conventional b/y ions. Figure kindly provided by Jacob Bush.

4.6.3. L16$_{RM}$ hydroxylation by ycfD$_{RM}$ is Fe(II) and 2OG-dependent

In order to confirm *ycfD$_{RM}$* as a *bona fide* 2OG oxygenase, the dependence of ycfD$_{RM}$ activity on co-factors Fe^{2+} and ascorbate, and co-substrate 2OG was investigated. Two forms of ycfD$_{RM}$ were tested: metallated, referring to the protein that has been purified through HisTrap and gel filtration columns; and *apo*, whereby ycfD$_{RM}$ purified as above was treated with the iron-chelator ethylenediamine-tetraacetic acid (EDTA), which was subsequently removed from solution by buffer-exchange into HEPES buffer. The results suggested that the oxidative modification of L16$_{RM}$ substrate fragment depended on ycfD$_{RM}$, 2OG and a mixture of Fe^{2+} and ascorbate (Figure 4.16). The *apo* form of ycfD$_{RM}$ was not active in the absence of exogenously added Fe^{2+}, suggesting that the transition metal is crucial for its activity. Furthermore, the metallated form retained 53% activity in the absence of exogenously added Fe^{2+}; the activity further decreased to background levels if ascorbate was also removed, which suggested that ycfD$_{RM}$ co-purifies with Fe^{3+} which requires ascorbate-mediated reduction to Fe^{2+} for activity. Activity of both the metallated and *apo* forms of ycfD$_{RM}$ demonstrated an approximately 50% loss in catalytic activity in the absence of ascorbate, suggesting that both exogenously added and co-purifying Fe^{2+} are prone to irreversible oxidation in the absence of ascorbate. This dependence on a reducing agent is interesting given that the presence of ascorbate has not been reported in prokaryotes. Although a different reducing agent may play the role of ascorbate in *E. coli* and *R. marinus* the search for such an agent is beyond the scope of this thesis.

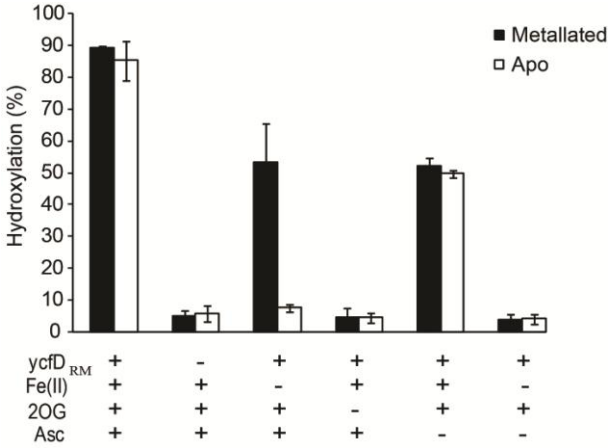
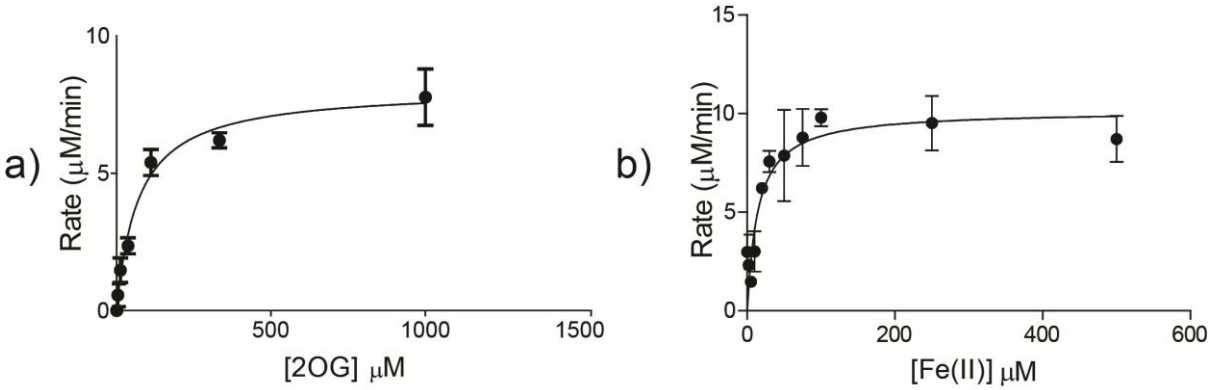


Figure 4.16: Co-factor dependence of ycfD$_{RM}$. L16 substrate fragment (KPVTKKPAEVRMGKGKGSVE, 100 μM) was incubated in a reaction mixture from which the enzyme, co-factors and co-substrates (ycfD$_{RM}$, Fe^{2+}, 2OG and ascorbate: 1 μM, 100 μM, 200 μM and 1 mM, respectively), were systematically removed. *Apo*- and metallated forms of ycfD$_{RM}$ were tested. Catalytic reaction was carried out in 50 mM HEPES (pH 7.5) at 65°C in triplicate. Mean value is shown, with error bars representing ± 1 s.d.

4.7. Kinetic studies on recombinant ycfD$_{RM}$

In order to quantify the kinetics of ycfD$_{RM}$ dependence on 2OG and Fe^{2+}, a K_m^{app} for 2OG and Fe^{2+} was determined. A time-point within the linear section of the hydroxylation curve was chosen, such that the signal-to-noise ratio was sufficient for meaningful interpretation of results across the entire range of cofactor/co-substrate concentrations. The results of the experiment suggested that K_m^{app} for 2OG is 70±12 μM (Figure 4.17a), significantly higher than for other 2OG oxygenases (17). K_m^{app} for Fe^{2+} was determined as 15±5 μM (Figure 4.17b).



		Fe(II)	2OG
YcfD$_{RM}$	V_{max} (μM/min)	10.2±0.9	8.1±0.4
	K_m (μM)	15±5	70±12

Figure 4.17: Rate-dependence of ycfD$_{RM}$ on 2OG and Fe^{2+}. Rate-dependence of ycfD$_{RM}$ on a) 2OG and b) Fe^{2+} was determined by incubating the enzyme with a serial dilution of co-factor, in the presence of saturating concentrations of all other co-factors (100 μM Fe^{2+}, 1 mM ascorbate, 400 μM 2OG and 400 μM substrate peptide). *Apo*-ycfD$_{RM}$ was used to determine the dependence on Fe^{2+}. Reactions were quenched with an equal volume of CF_3COOH_{aq} (1%) after 16 min and results analysed by MALDI-MS. Experiments were conducted in triplicate, with average values shown and error bars denoting ±1 s.d.

High temperatures forced the evolution of thermophilic enzymes into adaptation of their structures to cope with increased thermal energy, which is disruptive for tertiary/quaternary structure. The rigidity of a thermophilic enzyme at its optimum temperature should match the rigidity of its mesophilic homologue at ambient temperature and therefore has to increase relative to a mesophile, while balancing the flexibility of the active site (18). Thermophilic enzymes thus frequently contain an increased number of salt-bridges, hydrogen bonds and buried hydrophobic interactions compared to their mesophilic homologues, which rigidify their tertiary and quaternary structures (19).

Moreover, whilst increasing the rate of chemical reactions, thermal noise is disruptive for enzyme-substrate interactions, which have to be relatively strong for substrate binding compared to mesophilic homologues. K_m^{app} values therefore typically decrease from psychrophilic, via mesophilic to thermophilic enzymes, whereas the k_{cat}/K_m ratio increases (2).

In order to determine the K_m^{app} value for ycfD$_{RM}$ hydroxylation of L16$_{RM}$, reaction kinetic curves were first determined at 35°C (to enable comparison with ycfD) and 55°C. Fitting the curves with a linear regression function (GraphPad Prism version 5.02) suggested that in the linear range of the reaction, the slope at 35°C is 8.2±0.4 μM/min^{-1}, whereas at 55°C it is 25.4±2.6 μM/min^{-1}, i.e. 3.1-times faster (Figure 4.18).

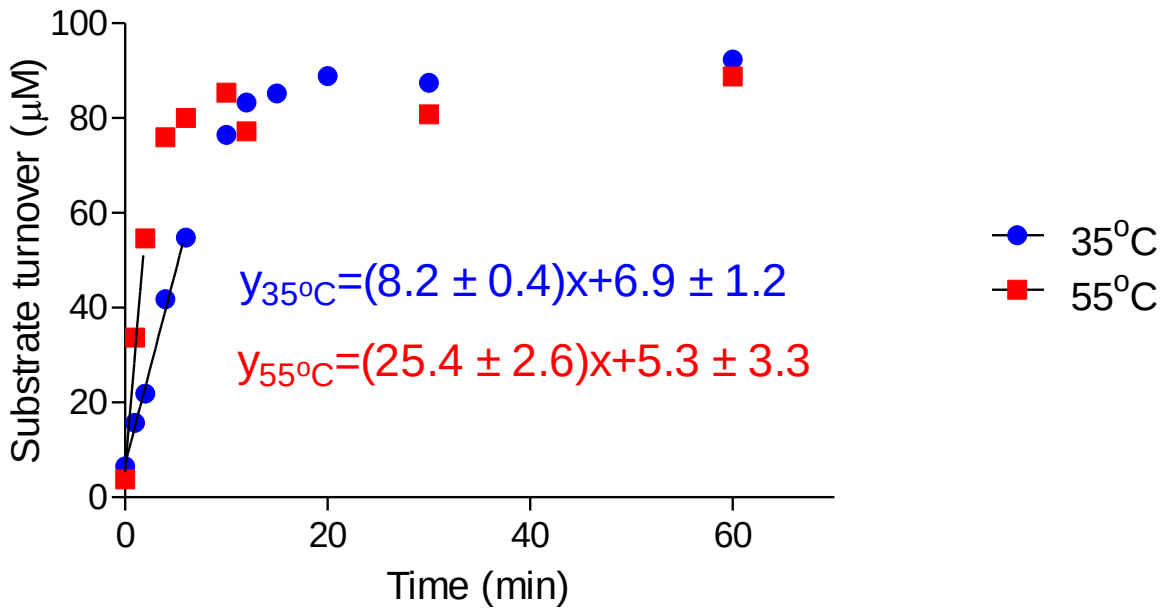


Figure 4.18: Time-dependent hydroxylation of ycfD$_{RM}$ at 35°C and 55°C. YcfD$_{RM}$ (1 μM) was incubated with L16$_{RM}$ (100 μM), 2OG (200 μM), Fe^{2+} (100 μM) and ascorbate (1 mM) at 35°C (blue circles) or 55°C (red squares). An aliquot (5 μL) was removed from the reaction at indicated time-points and quenched with an equal volume of CF_3COOH_{aq} (1%). Single data-points were determined for each time-point, as shown on the graph. GraphPad Prism version 5.02 was used to fit a linear regression function between t=0 and t=6 min (35°C) and t=0 to t=2 min (55°C).

For the determination of $K_m^{app}{}_{(substrate)}$, t=1 min was chosen for substrate concentrations < 1 mM and 3 min for concentrations >1 mM. The substrate peptide was then titrated and the rates measured by MALDI-MS analysis of peptide hydroxylation. $K_m^{app}{}_{(substrate)}$ for ycfD$_{RM}$ at 35°C was 208±66 μM (Figure 4.19).

K_m^{app} increased with temperature, from 208±66 μM at 35°C to 390±228 μM at 55°C (Figure 4.19). Despite repeating the experiment several times, it was not possible to reduce the error at the higher concentration points, leading to uncertainty in the measured K_m^{app} for the reaction at 55°C. V_{max} increased with temperature (from 66±7 μM/min at 35°C to 93±25 μM/min at 55°C; Figure 4.19).

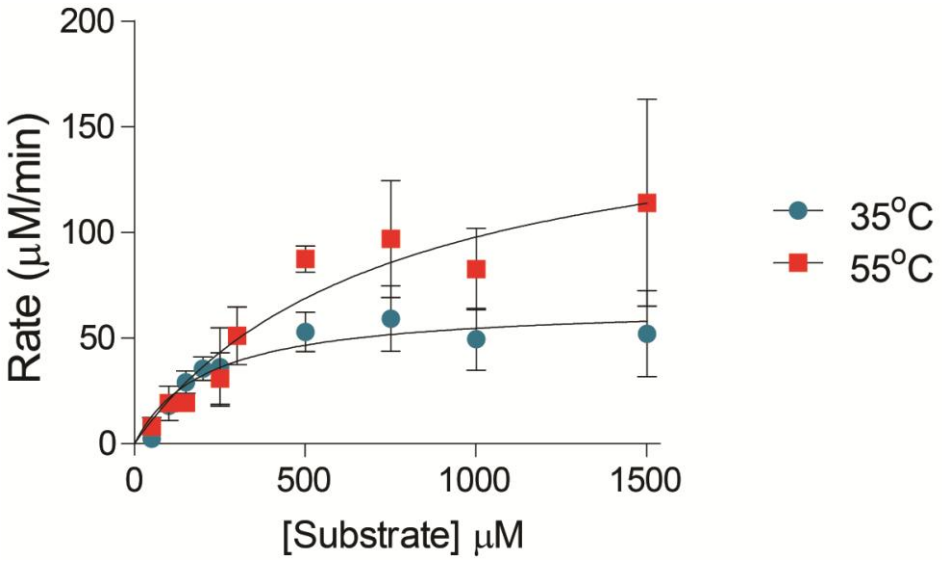


Figure 4.19: Rate-dependence of ycfD$_{RM}$ catalysis on the concentration of L16$_{RM}$ substrate peptide. YcfD$_{RM}$ (2 μM) was incubated with indicated concentrations of L16$_{RM}$(20-mer) at 35˚C (blue circles) and 55˚C (red squares) in the presence of 2OG, Fe^{2+} and ascorbate (2 mM, 100 μM and 1 mM, respectively). Reactions were quenched after 1 min by addition of equal volume of CF_3COOH_{aq} (1%) and results analysed by MALDI-MS. An average of three reactions is demonstrated with error bars denoting ±1 s.d.

Within experimental error, the k_{cat}/K_m ratio remained unchanged at 0.002±0.0006 $s^{-1}\mu M^{-1}$, consistent with a requirement for the enzyme to balance the increased affinity (lower K_m^{app}) and lower reaction rates (lower k_{cat}) at lower temperature and the decreased affinity (higher K_m^{app}) and higher reaction rates (higher k_{cat}) at higher temperature.

Table 4.4: Kinetic parameters for hydroxylation of L16$_{RM}$ by ycfD$_{RM}$ at different temperatures.

35°C		55°C	
V_{max}	66±7 μM/min	V_{max}	93±25 μM/min
K_m^{app}	208±66 μM	K_m^{app}	390±228 μM
k_{cat}	0.55 s^{-1}	k_{cat}	0.78 s^{-1}
k_{cat}/K_m	0.0026 $s^{-1}\mu M^{-1}$	k_{cat}/K_m	0.0020 $s^{-1}\mu M^{-1}$

4.8. Investigations of the L16$_{RM}$ binding mode to ycfD$_{RM}$

Next, the binding of mode of L16$_{RM}$ to ycfD$_{RM}$ was investigated. Two general modes of substrate binding have been observed for 2OG-dependent protein oxygenases: i) the substrate binds as a loop, with the substrate residue protruding into the core of the active site or ii) the substrate binds in a more linear manner, with secondary structure unfolded during the process of hydroxylation. Example of mode i) is binding of methylated histones to JMJD2A histone demethylase (20), whereas an example of ii) is binding of HIF-1α to FIH (21). Both JMJD2A and FIH are 2OG-dependent oxygenases from the JmjC-domain containing protein subfamily, and are therefore distant structural homologues of ycfD$_{RM}$. This variability in substrate binding modes of JmjC-domain containing oxygenases makes a prediction of the substrate binding mode for ycfD$_{RM}$ challenging.

4.8.1. YcfD$_{RM}$ catalysis of cyclic L16 peptides

Cyclic peptide inhibitors have recently become a productive route to design potent and selective inhibitors of 2OG oxygenases and other enzyme families (22). In the thermophilic bacterium *Thermus thermophilus*, for which the structure of the 50S subunit has been solved (23), the conserved residue R82 is located at the apex of the L16 beta-turn; this observation prompted the design of a cyclic peptide mimic of the L16 secondary structure, which 'locks' the conformation of the substrate as a loop. It was hypothesised that ycfD$_{RM}$ would only catalyse the hydroxylation of a cyclic L16$_{RM}$ peptide if the latter binds in the active site as a loop, rather than a straight chain.

The aim was to design, synthesise and test a cyclic peptide fragment of the ribosomal protein L16 beta turn. Models of cyclic peptides were built using the structure of the 50S ribosomal subunit of *T. thermophilus* (PDB: 2wdl (23)) under the supervision of Dr Martin Muenzel. The "builder" tool function in Pymol (DeLano Scientific, Version 1.3) was used to generate models. The peptide models were extracted from the .pdb model and imported into ChemBio3D (Cambridgesoft, Version 12.0). Their geometry was minimized using the "MM2" tool. Energy-minimised molecules were re-imported into Pymol and suitable cyclic peptide candidates were chosen after visual inspection (Figure 4.20).

A 14-mer cyclic peptide was designed and synthesised (KKPAEVRMGKGKGdC, peptide CP1) and cyclised using the cysteine alkylation procedure of Suga and co-workers (24). Initially, CP1 was tested as a substrate for ycfD$_{RM}$; the result suggested that it was a substrate for ycfD$_{RM}$ with a high degree of hydroxylation (>95%; Figure 4.20).

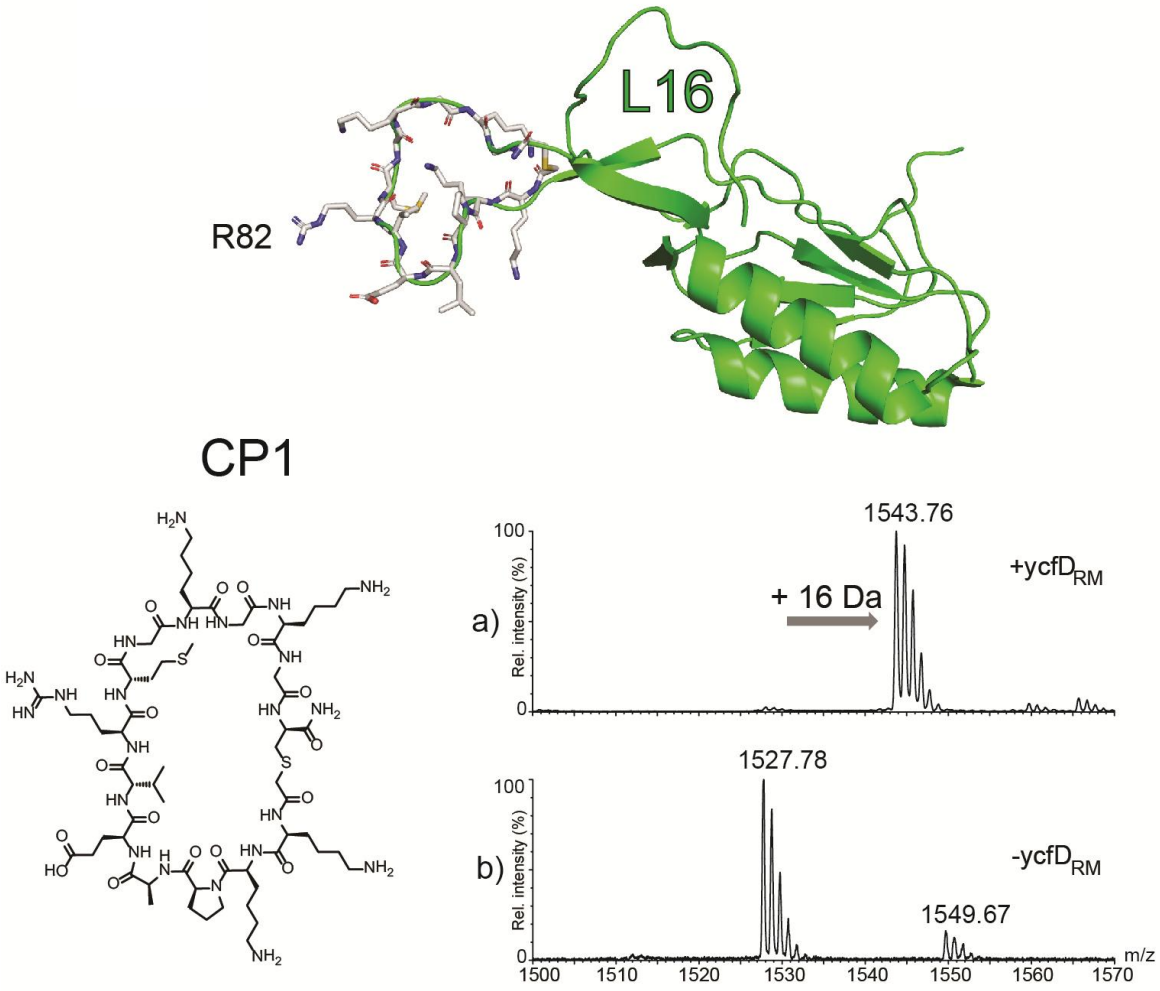


Figure 4.20: Model of L16$_{RM}$ cyclic peptide CP1 designed based on the *T. thermophilus* L16 protein from the structure of the 50S subunit (pdb: 2WDL) with MALDI-MS spectra showing that cyclic peptide CP1 undergoes ycfD$_{RM}$-dependent hydroxylation. CP1 (KKPAEVRMGKGKGdC, 100 μM) was incubated with ycfD$_{RM}$ (1 μM), 2OG (200 μM), Fe^{2+}(100 μM) and ascorbate (1 mM) for 15 min at 55°C, before quenching with an equal volume of CF_3COOH_{aq} (1%) and analysed by MALDI-MS (calculated mass of CP1 is 1528.8 Da).

Next, the cyclic peptide was tested in a competition study with the linear variant of the L16$_{RM}$ substrate fragment (20-mer). Kinetic experiments were performed by Aayan Hussein, under supervision of the author. Equal amounts of cyclic and linear peptides (100 μM) were incubated with ycfD$_{RM}$ (1 μM), 2OG (200 μM), Fe^{2+}(100 μM) and ascorbate (1 mM) at 55°C. At specified time-

points, aliquots were removed from the reaction mixture and quenched. The result indicates that CP1 is hydroxylated 9.5±2.6-times faster than the linear variant (Figure 4.21).

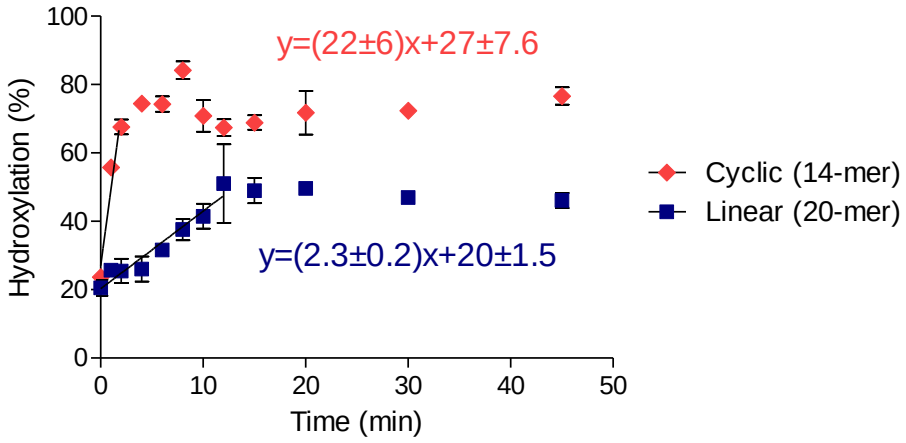


Figure 4.21: Competition study between *CP1* (blue diamonds) and linear L16 (20-mer, red squares) peptides. Equal amounts of cyclic and linear peptides (100 µM each) were incubated with ycfD$_{RM}$ (1 µM), 2OG (200 µM), Fe^{2+}(100 µM) and ascorbate (1 mM) at 55°C. At specified time-points, aliquots were removed from the reaction mixture and quenched by addition of equal volume of CF_3COOH_{aq} (1%). Results were analysed by MALDI-MS. Experiments were conducted in triplicate, with error bars denoting standard deviation from the mean.

K_m^{app} for CP1 was then determined at 35°C, in order to allow comparison with the linear variant determined previously. All conditions were replicated, in order to allow unbiased comparison of kinetic parameters. The K_m^{app} for the 14-mer cyclic peptide was found as 75±23 µM; and v_{max} was 62±4 µMmin^{-1} (Figure 4.22). Comparison of kinetic parameters for the linear and cyclic variants (Table 4.5) thus revealed that the K_m^{app} is 2.7-times lower for the cyclic variant, suggesting that it is a better substrate for ycfD$_{RM}$ than the linear variant. Because v_{max} values remained unchanged (0.47 and 0.50 s^{-1} for cyclic and linear variant, respectively), the mechanism of catalysis or their mode of binding likely does not differ.

Table 4.5: K_m determination for L16 peptide variants with ycfD$_{RM}$. Reactions were performed in triplicate. K_m and V_{max} values were determined by fitting a Michaelis Menten function in GraphPad Prism (version 5.02).

Substrate fragment	T [°C]	K_m (µM)	V_{max} (µM/min)	k_{cat} (s^{-1})
R. marinus L16 (20mer)	35	208±66	66±7	0.55
R. marinus L16 (20mer)	55	390±228	93±25	0.78
R. marinus cyclic L16 (CP1, 14 mer)	35	75±23	62±4	0.47

Overall, these results suggested that the mode of binding of L16$_{RM}$ substrate fragment in the ycfD$_{RM}$ active site more closely resembles a closed loop than open linear chain. The main difference between the two variants likely arises from the entropic term of pre-organisation. Cyclisation of L16 reduces its conformational freedom, thereby probably reducing the entropic penalty of binding and decreasing the negative free energy of binding. Crystallographic studies would be required to gain deeper understanding of the mode of binding of L16$_{RM}$ protein to ycfD$_{RM}$.

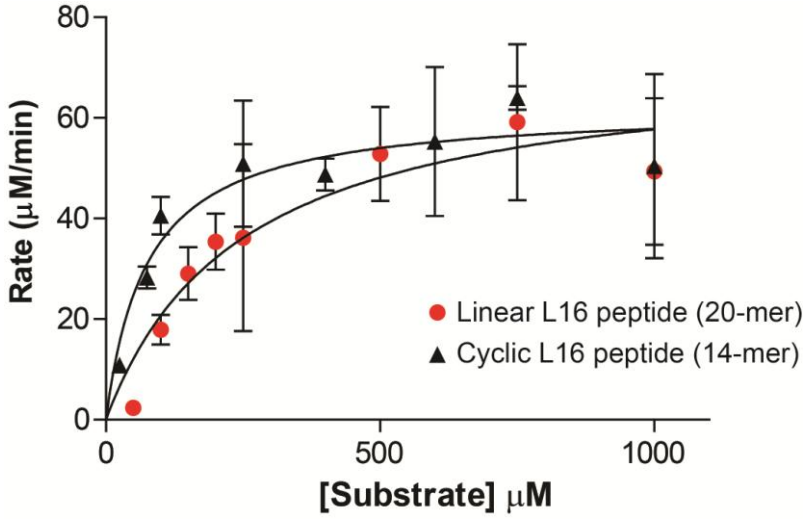


Figure 4.22: Kinetics of ycfD$_{RM}$-catalysed hydroxylation of the linear (KKPAEVRMGKGKGSVE, red circles) and cyclic (KKPAEVRMGKGKGdC, black triangles) L16 fragment peptide variants. L16$_{RM}$ linear (20-mer) and cyclic peptide (14-mer, CP1) fragments were incubated with ycfD$_{RM}$ (2 μM), 2OG (200 μM), Fe^{2+} (100 μM), and ascorbate (1 mM) for 30 sec (25-100 μM peptide) or 1 min (250 μM – 1 mM) at 37°C. Reactions were quenched by addition of an equal volume of CF_3COOH_{aq} (1%) and results analysed by MALDI-MS. Reactions were conducted in triplicate, with representative spectra shown.

4.9. YcfD substrate selectivity

The cross-reactivity of ycfD$_{RM}$ and ycfD$_{EC}$ was then examined. Initially, ycfD$_{RM}$ was incubated with substrate fragments derived from L16 homologues from *R. marinus* (L16$_{RM}$ positive control) and *E. coli* (L16$_{EC}$). Standard reaction conditions for ycfD$_{RM}$ were used.

Surprisingly, the extent of hydroxylation varied significantly between the two substrates. As expected, ycfD$_{RM}$-catalysed hydroxylation of L16$_{RM}$ was near-complete (94%), whereas the hydroxylation of L16$_{EC}$ was 3%, suggesting that L16$_{EC}$ is not a substrate for ycfD$_{RM}$ (Figure 4.23). The L16$_{EC}$ 20-mer sequence tested differs from the L16$_{RM}$ sequence at 5 positions, namely -8, -6, -3, -2 and +7. An

individual residue or a combination of residues within the 20-mer substrate fragment must therefore dictate substrate specificity of ycfD$_{RM}$. It was hypothesised that this result signifies an important mechanistic aspect of ycfD$_{RM}$ catalysis, and was thus pursued further.

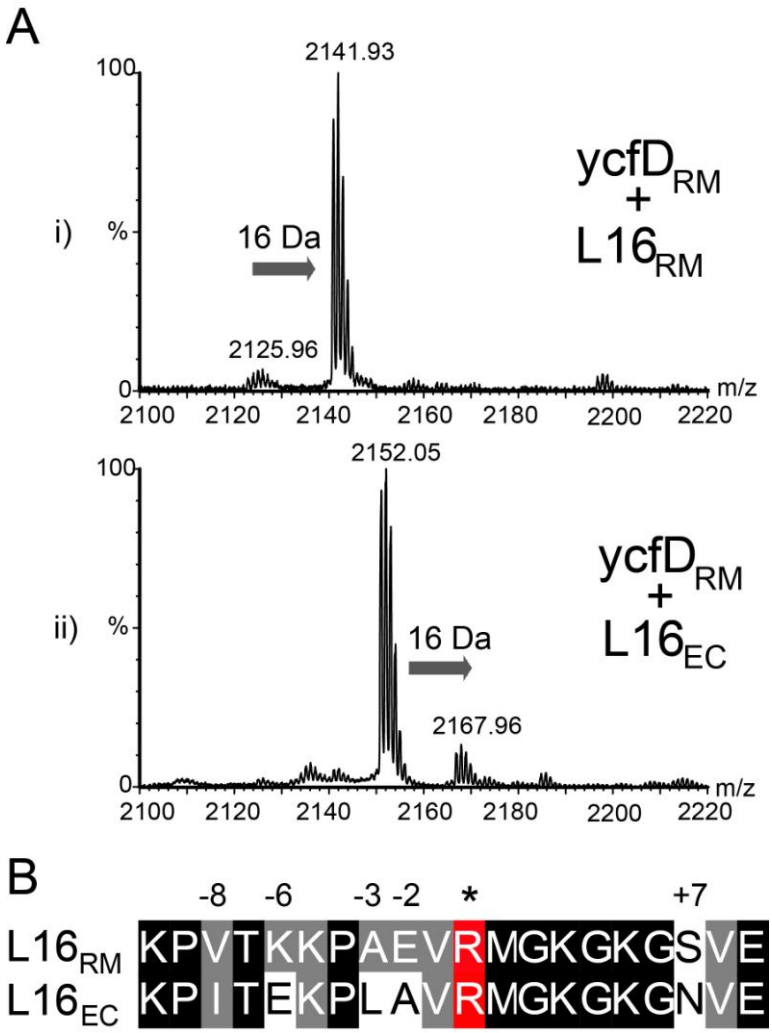


Figure 4.23: YcfD$_{RM}$-dependent catalysis of L16 substrate fragments from i) *R. marinus*, ii) *E. coli*. Substrate fragments (20-mers, 100 μM) were incubated with ycfD$_{RM}$ (1 μM), 2OG (200 μM), Fe^{2+} (100 μM), and ascorbate (1 mM) for 30 min at 65°C. Reactions were quenched by addition of an equal volume of CF_3COOH_{aq} (1%) and results analysed by MALDI-MS. Reactions were conducted in triplicate, with representative spectra shown. B) Sequence alignment of 20-mer L16 peptide fragments from *R. marinus* (L16$_{RM}$) and *E. coli* (L16$_{EC}$). Residue numbering relative to ycfD$_{RM/EC}$ target residue (R82/R81) is indicated above the alignment.

Initially, the experiment was repeated with ycfD$_{EC}$. Unlike for ycfD$_{RM}$, the results showed that, at least under the conditions tested, ycfD$_{EC}$ did not demonstrate similar sequence specificity. Moreover, ycfD$_{EC}$ catalysed the hydroxylation of L16$_{RM}$ to a greater extent than L16$_{EC}$ (88±7% and 59±2%, respectively), suggesting that it may itself possess a higher affinity for the *R. marinus*, than for the *E. coli* substrate sequence (Figure 4.24).

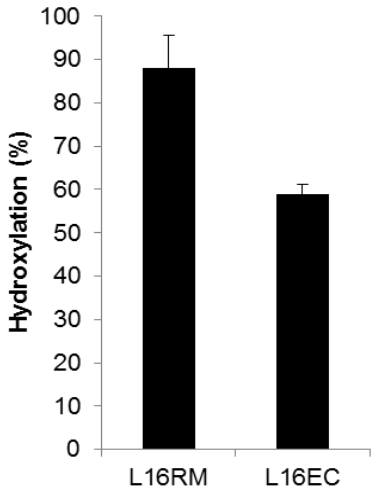


Figure 4.24: YcfD$_{EC}$ catalyses hydroxylation of L16$_{RM}$ and L16$_{EC}$ substrate fragments (20-mers). YcfD$_{EC}$ was incubated with L16$_{RM}$ and L16$_{EC}$ substrate fragments (KKPAEVRMGKGKGSVE and KPITEKPLAVRMGKGKGNVE, both 100 μM) 2OG (200 μM), Fe^{2+} (100 μM), and ascorbate (1 mM) for 30 min at 37°C. Reactions were quenched by addition of an equal volume of CF_3COOH_{aq} (1%) and results analysed by MALDI-MS. Reactions were conducted in triplicate, with error bars denoting ±1 s.d. from the mean.

4.9.1. Identification of L16 residues conferring ycfD$_{RM}$ substrate selectivity

It was therefore interesting to define what variations in the L16 sequence renders ycfD$_{RM}$ unable to accept L16$_{EC}$ as a substrate. In order to exclude cumulative effects of several variations working in concert, a set of permutated variants was designed and synthesised, aimed at gradually 'evolving' the L16$_{RM}$ sequence into L16$_{EC}$. Peptides were synthesised on a peptide multisynthesiser (operated by Dr. Wei Ge). Their purity was assessed by MALDI-MS and was >80% for every peptide. Enzymatic peptide hydroxylation (50 μM) was analysed for relative to no-enzyme control reactions. A number of peptides revealed partial (< 10%) non-enzymatic oxidation, which likely occurred under the conditions tested (60°C in the presence of Fe(II), ascorbate and 2OG, potentially leading to oxidative Fenton chemistry). The results revealed that for ycfD$_{RM}$-catalysed hydroxylation, all peptides containing Ala at position -2 relative to R82 were <10% hydroxylated. The 15 peptides that were hydroxylated >28% (Figure 4.25A) could be further stratified by considering position -6; all peptides containing K^{-6} were hydroxylated >60%, whereas only one peptide containing E^{-6} was hydroxylated >60%, all others were hydroxylated between 28 and 58%. The results thus suggested that position -2 was crucial for positive recognition of the substrate by ycfD$_{RM}$. Peptides containing A^{-2} were not hydroxylation substrates of ycfD$_{RM}$, thereby explaining the lack of reactivity of ycfD$_{RM}$ against L16$_{EC}$, which contains A at position -2.

A

Sequence	%OH over control	-6	-2
KPITKKPLEVRMGKGKGSVE	78.8	K	E
KPVTKKPAEVRMGKGKGSVE	76.0	K	E
KPVTKKPLEVRMGKGKGSVE	75.7	K	E
KPITKKPLEVRMGKGKGNVE	72.9	K	E
KPVTKKPAEVRMGKGKGNVE	69.6	K	E
KPVTKKPLEVRMGKGKGNVE	67.9	K	E
KPITEKPAEVRMGKGKGNVE	66.4	E	E
KPITKKPAEVRMGKGKGNVE	63.5	K	E
KPVTEKPLEVRMGKGKGSVE	58.1	E	E
KPITEKPAEVRMGKGKGSVE	54.3	E	E
KPVTEKPAEVRMGKGKGSVE	49.9	E	E
KPVTEKPAEVRMGKGKGNVE	48.6	E	E
KPITEKPLEVRMGKGKGSVE	47.6	E	E
KPITEKPLEVRMGKGKGNVE	37.1	E	E
KPVTEKPLEVRMGKGKGNVE	28.9	E	E
KPVTKKPAAVRMGKGKGSVE	6.7	K	A
KPVTKKPLAVRMGKGKGNVE	6.2	K	A
KPITKKPAAVRMGKGKGSVE	3.5	K	A
KPITEKPAAVRMGKGKGNVE	2.5	E	A
KPVTEKPLAVRMGKGKGNVE	2.4	E	A
KPITKKPAAVRMGKGKGNVE	1.9	K	A
KPITKKPLAVRMGKGKGNVE	1.5	K	A
KPITEKPAAVRMGKGKGSVE	1.2	E	A
KPITEKPLAVRMGKGKGSVE	0.6	E	A
KPITKKPLAVRMGKGKGSVE	0.0	K	A
KPVTKKPLAVRMGKGKGSVE	-1.1	K	A
KPVTEKPLAVRMGKGKGSVE	-1.3	E	A
KPITEKPLAVRMGKGKGNVE	-2.7	E	A
KPVTEKPAAVRMGKGKGSVE	-3.4	E	A
KPVTEKPAAVRMGKGKGNVE	-4.3	E	A
KPVTKKPAAVRMGKGKGNVE	-8.8	K	A

B

Sequence	%OH over control	-6	-2
KPITKKPLEVRMGKGKGNVE	83.0	K	E
KPVTKKPLAVRMGKGKGNVE	81.4	K	A
KPITKKPLEVRMGKGKGSVE	78.1	K	E
KPVTKKPAAVRMGKGKGNVE	77.3	K	A
KPITKKPLAVRMGKGKGSVE	76.4	K	A
KPVTEKPAAVRMGKGKGNVE	74.2	E	A
KPVTKKPAEVRMGKGKGSVE	73.8	K	E
KPITKKPAAVRMGKGKGSVE	73.3	K	A
KPVTKKPAAVRMGKGKGSVE	72.8	K	A
KPITEKPAEVRMGKGKGNVE	72.2	E	E
KPITEKPLAVRMGKGKGSVE	71.2	E	A
KPITKKPAAVRMGKGKGNVE	70.4	K	A
KPVTKKPAEVRMGKGKGNVE	69.7	K	E
KPITEKPAEVRMGKGKGSVE	69.2	E	E
KPITKKPAEVRMGKGKGNVE	69.2	K	E
KPVTKKPLEVRMGKGKGNVE	69.2	K	E
KPVTKKPLEVRMGKGKGSVE	69.0	K	E
KPVTEKPLEVRMGKGKGSVE	68.7	E	E
KPITEKPAAVRMGKGKGSVE	67.3	E	A
KPVTEKPLAVRMGKGKGNVE	65.8	E	A
KPVTEKPAAVRMGKGKGSVE	65.8	E	A
KPITEKPAAVRMGKGKGNVE	64.9	E	A
KPVTKKPLAVRMGKGKGSVE	63.8	K	A
KPITEKPLEVRMGKGKGNVE	63.1	E	E
KPVTEKPAEVRMGKGKGNVE	56.8	E	E
KPVTEKPLEVRMGKGKGNVE	56.3	E	E
KPITKKPLAVRMGKGKGNVE	53.7	K	A
KPVTEKPLAVRMGKGKGSVE	45.3	E	A
KPITEKPLEVRMGKGKGSVE	43.6	E	E
KPITEKPLAVRMGKGKGNVE	28.5	E	A
KPVTEKPAEVRMGKGKGSVE	25.0	E	E

Figure 4.25: MALDI-MS analysis of % activity of A) $ycfD_{RM}$ and B) $ycfD_{EC}$ against the L16 variant peptide library containing all permutations of variations between $L16_{EC}$ and $L16_{RM}$. Peptide variants (50 μM) were reacted with $ycfD_{RM}$ or $ycfD_{EC}$ (1 μM) in the presence of 2OG (200 μM), Fe^{2+} (100 μM) and ascorbate (1 mM). Reactions were allowed to proceed for 10 min at 60°C/37°C ($ycfD_{RM}$/$ycfD_{EC}$, respectively) in a 96-well plate. Reactions were then quenched by addition of 20 μL CF_3COOH_{aq} (1%) and analysed by MALDI-MS. A control reaction containing no enzyme was performed for each peptide; the difference between +/- enzyme reactions is shown as % hydroxylation over control.

When $ycfD_{EC}$ was reacted with the peptide library, the results suggested that position -2 plays a less important role for peptide hydroxylation (Figure 4.25B). All peptide variants were hydroxylated to >25%. Within the top 50% of hydroxylated peptides (all of which were hydroxylated to > 69%) the ratio of A^{-2} to E^{-2}-containing peptides was 1.1:1, suggesting that $ycfD_{EC}$ does not show a strong preference for residue -2. Interestingly, however, 73% of the top 50% peptides contained K at position -6 (K^{-6}). Since the endogenous residue in $L16_{EC}$ at that position is E^{-6}, this suggested that $ycfD_{EC}$ shows sequence preference for the non-endogenous $L16_{RM}$ over the endogenous $L16_{EC}$ sequence.

4.9.2. Alanine scanning study confirms E^{-2} as the key residue for $ycfD_{RM}$ catalysis

In order to investigate the importance of other residues within 10 residues N-/C-terminal from the target residue R82, a set of peptides was prepared with each residue substituted with alanine.

A set of 20 L16$_{RM}$ (20-mer) analogues was prepared using a peptide multisynthesiser (operated by Dr. Wei Ge), where each had a consecutive residue replaced by alanine. Individual peptides (50 μM) were reacted with ycfD$_{RM}$ (1 μM) in the presence of 2OG (200 μM), Fe^{2+} (100 μM) and ascorbate (1 mM). Reactions were allowed to proceed for 10 min at 60°C before quenching and analysis by MALDI-MS.

The results revealed the importance of three residues in a 20-mer L16 sequence (Figure 4.26): the E80A variant was hydroxylated 4±0.5%, which is within the experimental error of the MALDI-MS experiment, thus corroborating the earlier finding that the E^{-2} residue is crucial for hydroxylation of L16 by ycfD$_{RM}$. The R82A variant also showed background levels of hydroxylation (6±1.1%) consistent with R82 being the substrate for hydroxylation. V81A variant was hydroxylated to 20±4.0%, suggesting that a hydrophobic residue is required at position -1 relative to R82. Interestingly, K^{-6} replacement with alanine led to a 10% decrease in activity compared to WT peptide (63±5.0% from 72±4.1%). Together with a result from the previous section, which suggested that the K76E variant led to reduction in hydroxylation by both ycfD$_{RM}$ and ycfD$_{EC}$, this result suggested the presence of an important interaction of position -6 with the enzyme backbone. It is likely that backbones of both ycfD$_{EC}$ and ycfD$_{RM}$ contain an acidic residue which preferentially interacts with a basic residue at position -6 of L16.

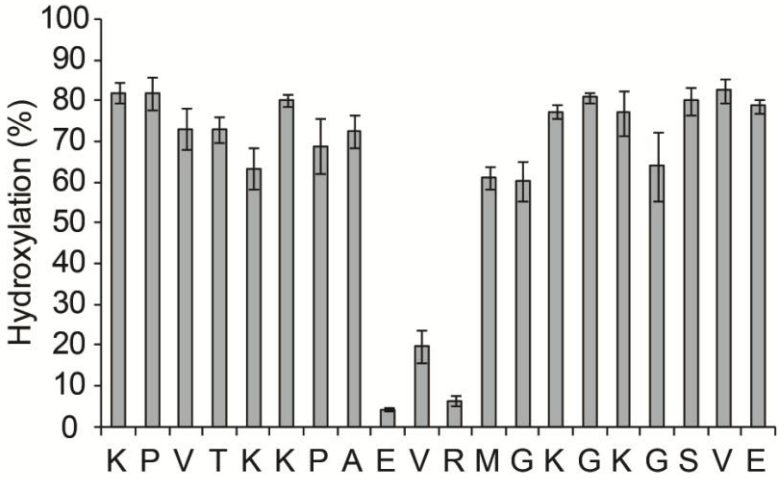


Figure 4.26: Alanine-walking experiment with ycfD$_{RM}$ and L16$_{RM}$ variants. Peptide variants (50 μM) were reacted with ycfD$_{RM}$ (1 μM) in the presence of 2OG (200 μM), Fe^{2+} (100 μM) and ascorbate (1 mM). Reactions were allowed to proceed for 10 min at 60°C in a 96-well plate. Reactions were then quenched by addition of 20 μL CF_3COOH_{aq} (0.1%) and analysed by MALDI-MS. Reactions were carried out in triplicate and a control reaction containing no enzyme was performed for each peptide. Triplicate values were averaged and the background subtracted for each peptide, with error bars signifying the standard deviation.

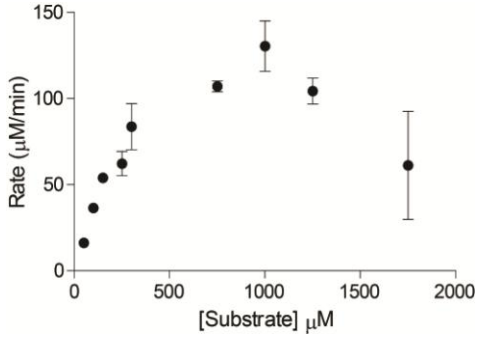
4.9.3. L16$_{EC}$ A79E variant improves hydroxylation kinetics with ycfD$_{EC}$

In order to quantify the different substrate affinities of ycfD$_{EC}$ and ycfD$_{RM}$ for their respective substrates, K_m^{app} values were determined for ycfD$_{EC}$ catalysis of L16$_{EC}$ and L16$_{RM}$; because the previous experiments revealed the importance of position -2, the K_m^{app} value for L16$_{EC}$ -2 variant A79E was also determined. The values were compared to previously determined K_m^{app} value for ycfD$_{RM}$ catalysis of L16$_{RM}$. As ycfD$_{RM}$ does not catalyse the hydroxylaton of L16$_{EC}$, the K_m^{app} value for this reaction could not be determined. The experiments were performed by Aayan Hussein under supervision of the author.

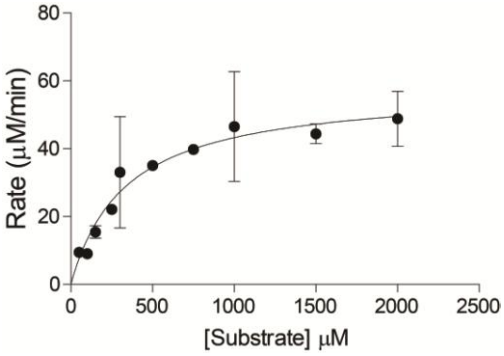
The rate dependence of ycfD$_{EC}$ catalysis of L16$_{EC}$ showed substrate inhibition (Figure 4.27). The experiment was independently repeated three times, with identical outcome. This suggested that the L16$_{EC}$ fragment sequence is not an optimal substrate for ycfD$_{EC}$. An accurate K_m^{app} value for ycfD$_{EC}$ hydroxylation of L16$_{EC}$ could not be determined; fitting of both Michaelis-Menten function and substrate inhibition function (GraphPad Prism, version 5.02) resulted in large error in K_m^{app} (Figure 4.27d). In contrast, ycfD$_{EC}$ hydroxylation of L16$_{EC}$ A79E did not show substrate inhibition (Figure 4.27b). K_m^{app} was 337±64 μM (comparable to ycfD$_{RM}$ catalysis of L16$_{RM}$ at 55°C, Table 4.4), which suggested that ycfD$_{EC}$ preferentially hydroxylates substrates with E^{-2} over A^{-2}. Interestingly, ycfD$_{EC}$ hydroxylation of L16$_{RM}$ substrate demonstrated the lowest K_m^{app} in the series, 164 ±37 μM. This suggested that residues other than -2 may play a role in determining the rate of ycfD$_{EC}$ catalysis. Residue at position -6 would be the next most likely candidate to exert control on the hydroxylation rate; due to time constraints the importance of that residue was not investigated further.

Overall, the results suggested that L16$_{RM}$ is a better substrate for ycfD$_{EC}$ than L16$_{EC}$, and identified the residue E^{-2} as being important for ycfD$_{RM}$ hydroxylation; E-2 also improves the affinity of ycfD$_{EC}$. It can be concluded, therefore, that ycfD$_{RM}$ and ycfD$_{EC}$ contain a key substrate-recognition site close to the active site, which mediates substrate selectivity in ycfD$_{RM}$ and promotes faster catalysis in ycfD$_{EC}$. A possibility that should be investigated in the future is that substrate selectivity of ycfD$_{RM}$ is temperature dependent, and is 'relaxed' at lower temperature, e.g. 37°C at which ycfD$_{EC}$ operates.

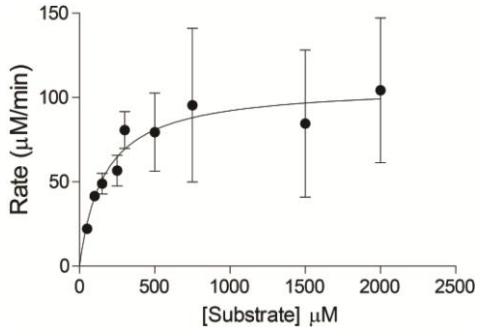
a)



b)



c)



d)

Peptide variant	K_m *(μM)*	V_{max} *(μM/min)*
L16$_{EC}$ WT	495 ±449	210 ±130
L16$_{EC}$ A79E	337 ±64	58 ±4
L16$_{RM}$ WT	164 ±37	53 ±7

Figure 4.27: K_m^{app} determinations for ycfD$_{EC}$ hydroxylation of L16 variants (all 20-mer peptides). A) ycfD$_{EC}$ hydroxylation of L16$_{EC}$ WT; b) ycfD$_{EC}$ hydroxylation of L16$_{EC}$ A79E; c) ycfD$_{EC}$ hydroxylation of L16$_{RM}$ WT; d) summary of kinetic parameters for ycfD hydroxylation of L16 variants.

4.9.4. Studying ycfD$_{RM}$ substrate selectivity *in vivo*

In order to verify the observed substrate selectivity in cells, ycfD$_{RM}$ was expressed in a *ycfD*-deleted background (*E. coli* strain JW1114) and hydroxylation of ribosomal protein L16 studied by whole-protein MS. YcfD$_{EC}$ expression was used as a positive control (see Chapter 2). Vector pTRC99a was chosen because it carries ampicillin resistance gene and a *lac* promoter and is therefore suitable for expression in a wide range of bacteria (25).

In order to sub-clone ycfD/ycfD$_{RM}$ into pTRC99A, both genes were amplified from the respective pET28A constructs. JW1114 was transformed with ycfD$_{EC}$_pTRC99a or ycfD$_{RM}$_pTRC99a. Standard ribosomal protein LCMS method was used to resolve the *E. coli* ribosomal proteins. A signal corresponding to protein mass of L16 (15311±1 Da, Figure 4.28c) eluted at 33.1 min.

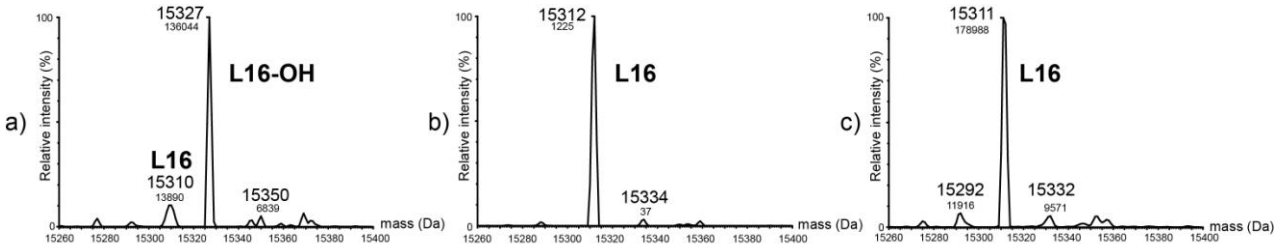


Figure 4.28: Deconvoluted ESI-MS spectra of *E. coli* ribosomal protein L16$_{EC}$ from ycfD knock-out strain (JW1114) transformed with a) ycfD_pTRC99 induced and grown at 18°C, b) ycfD$_{RM}$_pTRC99 induced and grown at 30°C; c) L16$_{EC}$ from the untransformed JW1114 strain.

L16$_{EC}$ from the control sample (ycfD_pTRC99 over-expression, Figure 4.28a) was >90% hydroxylated, suggesting that over-expression from pTRC99 was successful. Upon over-expression of ycfD$_{RM}$_pTRC99a, L16$_{EC}$ hydroxylation was <5% (Figure 4.28b), in agreement with the *in vitro* data presented earlier (Figure 4.23) and suggests that interspecies differences in L16 sequence between *E. coli* and *R. marinus* are sufficient to disable cross-reactivity of their respective ycfD homologues. It also suggested that peptide fragments of L16 are adequate model systems for studying ycfD$_{EC}$/ycfD$_{RM}$ activity *in vitro*. Furthermore, the result was important because it suggested that *E. coli* strain JW1114 cannot be used as a model strain for studying ycfD$_{RM}$ activity in cells.

Both observations were interesting from an evolutionary perspective. The Pfam database (Sanger Institute, University of Cambridge) contains data for all DNA-sequenced L16 homologues. A statistical analysis performed by the database indicates that the most highly conserved residue at position -2 (relative to R82) is glutamate (E^{-2}), whereas the most highly conserved residue at position -6 is lysine (K^{-6}) (Figure 4.29), both residues are found in the L16$_{RM}$, but not the L16$_{EC}$ sequence.

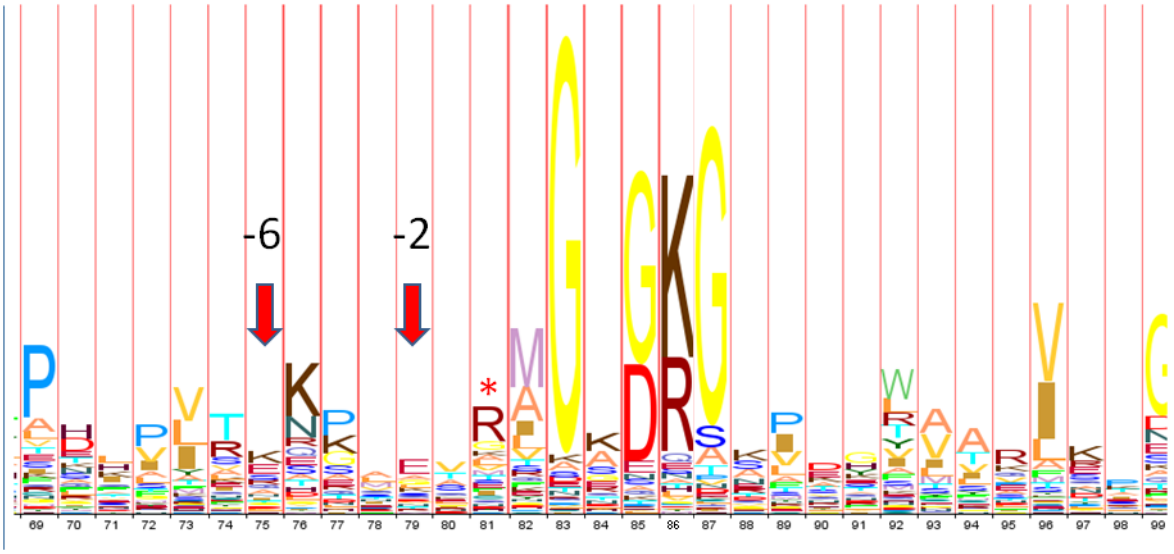


Figure 4.29: Pfam profile of L16 homologues from all sequenced genomes. Evolutionary conservation of residues is indicated by the height of letters (indicating individual amino acids). The profile suggests that G83 is the most widely conserved residue in L16, followed by G87, G85 and K863. Position of R81/2 is indicated by the red star. Positions -2 and -6 relative to R81/2 are indicated by red arrows. Source: Sanger Institute, Cambridge.

The search for alanine-containing L16 sequences was then conducted by searching through the Uniprot database. The results suggested that only *Enterobacteriales* strains *E. coli*, *Salmonella* and *Shigella* have acquired the E79A substitution (no other bacterial species was found with A79). Since ycfD$_{EC}$ demonstrated a preference for hydroxylation of peptides containing E^{-2} and K^{-6} this suggested that ycfD$_{EC}$ has not adapted to the substitutions present in *Enterobacteriales*. A more thorough evolutionary analysis would be required to substantiate this hypothesis.

4.10. Biochemical comparison of ycfD$_{EC}$ and ycfD$_{RM}$: biocatalytic potential

2OG oxygenases play roles in the biosynthesis of important natural products such as β-lactam antibiotics, including the penicillins and cephalosporins (26) and have been targets of bioengineering efforts to optimise the scale and diversity of antibiotic production by fermentation (27,28). 4-Hydroxyproline is produced commercially by fermentation using a recombinant 2OG-dependent proline 4-hydroxylase (29). Given the diverse range of stereospecific oxidative modifications that 2OG oxygenases catalyse, their biocatalytic potential is substantial. Thus, understanding the factors that would enhance their thermostability is of academic, as well as industrial significance. It is significant that ycfD$_{RM}$ is the first described thermophilic 2OG oxygenase; an investigation of its

biophysical properties, in comparison with ycfD$_{EC}$, will therefore define the extremophilic properties that a 2OG oxygenase could evolve and which could be used biotechnologically for the purposes outlined above.

4.10.1. Primary sequence comparison

Initially, the differences in the primary sequences of ycfD$_{EC}$ and ycfD$_{RM}$ were considered. Thermophilic proteins are frequently adapted for stability, but not activity, at high temperature, often by introduction of rigidifying salt-bridge interactions (30). Indeed, one broad study reported that the only common feature of thermophilic adaptation is an increase in number of ion pairs with increasing growth temperature (31).

The primary sequences of ycfD homologues from *E. coli* and *R. marinus*, as well as a putative ycfD homologue ycfD$_{PA}$ from the psychrophilic organism *Psychrobacter arcticus* were compared. Inclusion of the latter resulted in more pronounced trends in changes of amino acid composition. Consistent with reports from other organisms (19), an increase in abundance of highly charged (Arg, Glu), rigidifying (Pro) and small hydrophobic amino acids (Ala, Leu), and a decrease in oxidasable polar amino acids (Met), polar uncharged amino acids (Ser, Asn, Gln), and amino acids with low charge density (Lys and Asp) was observed in ycfD$_{RM}$, compared to mesophilic ycfD$_{EC}$ and psychrophilic ycfD$_{PA}$ (Figure 4.30a). This observation is consistent with an increase in the number of electrostatic interactions, such as salt bridges, which have been found to stabilise folded states of hyperthermophilic proteins (19). Salt-bridges enhance thermostability by reducing $\Delta C(p)$, leading to the up-shifting and broadening of the protein stability curves (32). Protein stability can further be enhanced by formation of hydrophobic cores formed by branched apolar residues (19). Consistent with this requirement for branched apolar residues, valine is the most highly abundant amino acid in ycfD$_{RM}$ (14% of total protein sequence, Figure 4.30b), followed by alanine (10%). However, these amino acids are also highly abundant in ycfD$_{EC}$ and ycfD$_{PA}$ (11 and 12% for valine, respectively; 6 and 8% for alanine, respectively), indicating that hydrophobic interactions may be common to all three homologues.

A greater number of 2OG oxygenase homologues from thermophiles, mesophiles and psychrophiles would be necessary to draw statistically significant conclusions about amino acid substitutions. As ycfD$_{RM}$ is the only known thermophilic 2OG oxygenase, such a study is not presently possible.

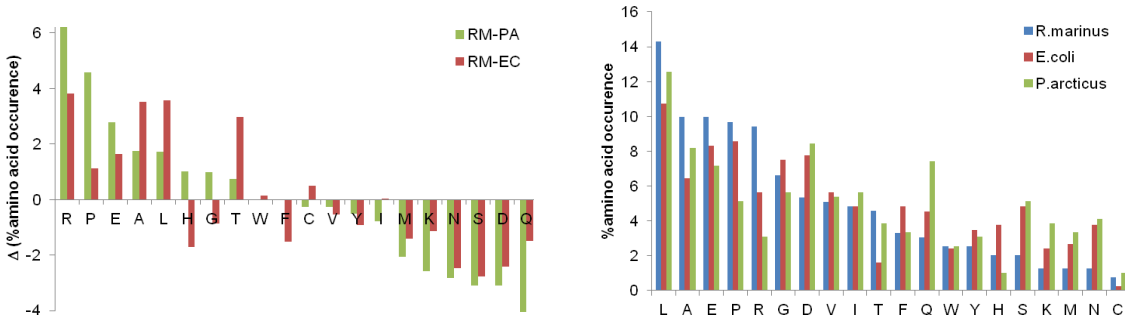


Figure 4.30: Thermal adaptation of 2OG oxygenase primary sequence, as exemplified by analysis of ycfD homologues from a thermophilic (*R. marinus*; RM), mesophilic (*E. coli*, EC) and psychrophilic (*P. arcticus*, PA) organisms. A) Difference in % amino acid occurrence between *R. marinus* and *P. arcticus* (green bars) and between *R. marinus* and *E. coli* (red bars). B) Percent occurrence of individual amino acids calculated from primary sequence and represented as percentage of total protein sequence.

4.10.2. Comparison of enzymatic activity and stability

Salt-bridge interactions require polar amino acids in their ionic states; the strength of ionic interactions should thus depend on the pH of solution. The dependence of catalytic activity of protein stability on pH was therefore investigated. A set of buffers was prepared, covering the pH range 6-10; buffers were used as close to their pI values as possible. Tris-HCl buffer was not used because it was found to negatively affect the activity of ycfD$_{EC}$, and because its pK_a is temperature-dependent (see Chapter 3).

A bell-shaped pH-dependence was observed, with maximum enzymatic activity achieved at pH 7.5 (Figure 4.31). The optimum pH is thus negatively shifted by 0.5 units relative to ycfD$_{EC}$ (optimum pH 8.0), perhaps reflecting a requirement for a more neutral pH value, at which most charged polar amino acids will exist in their ionic forms, thereby enabling salt-bridge formation.

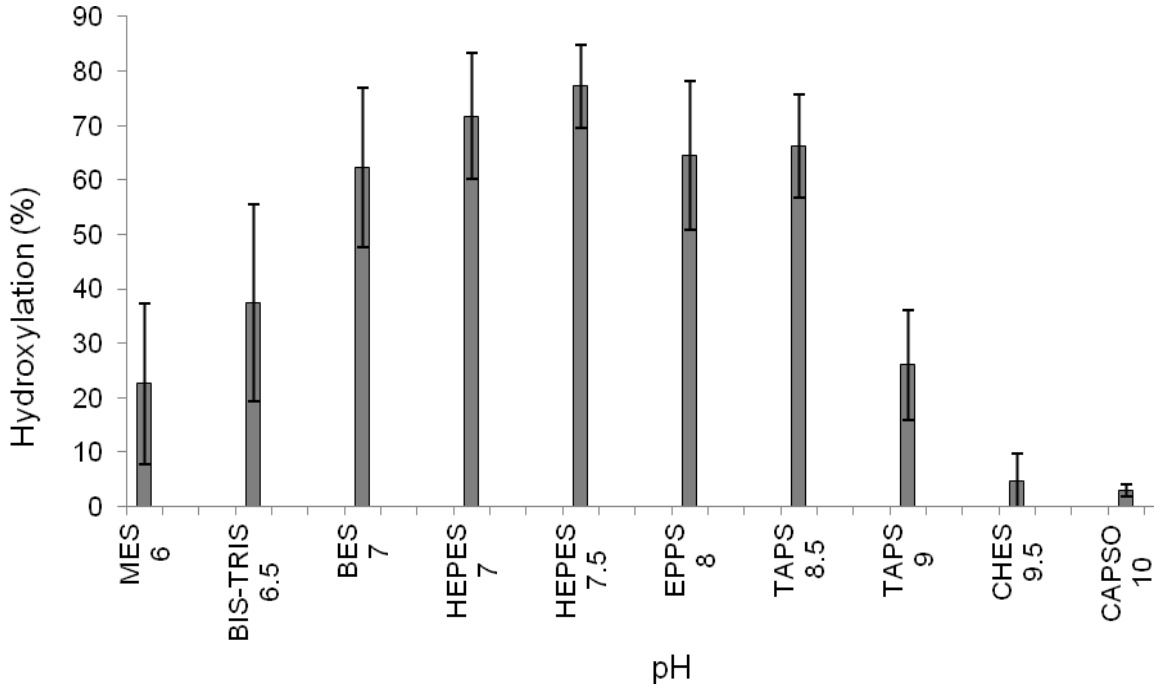


Figure 4.31: pH dependence of ycfD$_{RM}$ activity. The activity of ycfD$_{RM}$ at different pH values was measured by incubation of ycfD$_{RM}$ (1.5 μM) with L16$_{RM}$ substrate fragment, 2OG, Fe(II) and ascorbate (50 μM, 200 μM, 100 μM and 1 mM, respectively) diluted from initial stocks (500 mM for 2OG and ascorbate, 400 mM for Fe(II) and 1 mM for substrate peptide) in 50 mM HEPES pH 7.5, to final stock (10x final concentration) in selected buffer solutions. Substrate and enzyme mixes were prepared for each reaction, and allowed to react for 30 min at 65°C before quenching with equal volume of CF_3COOHaq (1% vol/vol). Reactions were carried out in triplicate and analysed by MALDI-MS. Error bars denote standard deviation from the mean.

The effect of pH on protein stability was addressed next. Differential scanning fluorimetry (DSF) was used to study protein denaturation. DSF measures the fluorescence intensity of a hydrophobic dye (Sypro®Orange) which fluoresces when bound to hydrophobic patches that are usually buried in the protein core. As thermal unfolding of a protein proceeds, hydrophobic pockets become solvent exposed, thus allowing SyproOrange™ to bind and emit a fluorescent signal (33).

YcfD$_{RM}$ showed three distinct profiles of unfolding: a major unfolding at 80-85°C (pH 5.0-7.0), ii) unfolding at 50-55°C followed by unfolding at 80-85°C (pH 7.5-8.5) and iii) single unfolding at 45°C (pH 9.0-10.0, Figure 4.32). Maximum fluorescence signal for pH range 5.0-8.5 was 0.2, markedly lower compared to other 2OG oxygenases, which had been studied by DSF at comparable concentration (Inga Pfeffer, personal communication). The maximum fluorescence intensity increased to 0.7 above pH 8.5, which suggested that a higher proportion of hydrophobic residues become exposed at high pH upon thermal denaturation.

For ycfD$_{RM}$ the Sypro®Orange signal was low below pH 8.5, where enzymatic activity was maximal (Figure 4.31), suggesting that a significant proportion of hydrophobic patches remained buried up to 98°C and that ycfD$_{RM}$ in this pH range does not undergo a complete unfolding. T_m in that pH range was 80-85°C (Figure 4.33). Between pH 7.5-8.5 two transitions of approximately equal intensity were observed, at 53 and 84°C. Above pH 8.5, the Sypro®Orange signal increased 4-fold, T_m decreased to 45°C, concomitant with a loss of enzymatic activity (<20%, Figure 4.31). This suggested that the exposure of hydrophobic patches was significantly increased, leading to loss of structural integrity and catalytic activity of ycfD$_{RM}$. Overall, the results suggested that hydrophobic patches are an important stabilising feature of ycfD$_{RM}$ which provide stabilisation up to 80°C (onset of unfolding) between pH 5.0-8.5. The low fluorescence intensity for ycfD$_{RM}$ suggested that complete unfolding does not take place at this pH. Above pH 8.5, the tertiary structure is disrupted so as to expose hydrophobic patches to a larger extent (a 3.5-fold increase, assuming a linear correlation between concentration of Sypro®Orange bound and fluorescence intensity), leading to a markedly lower T_m (45°C).

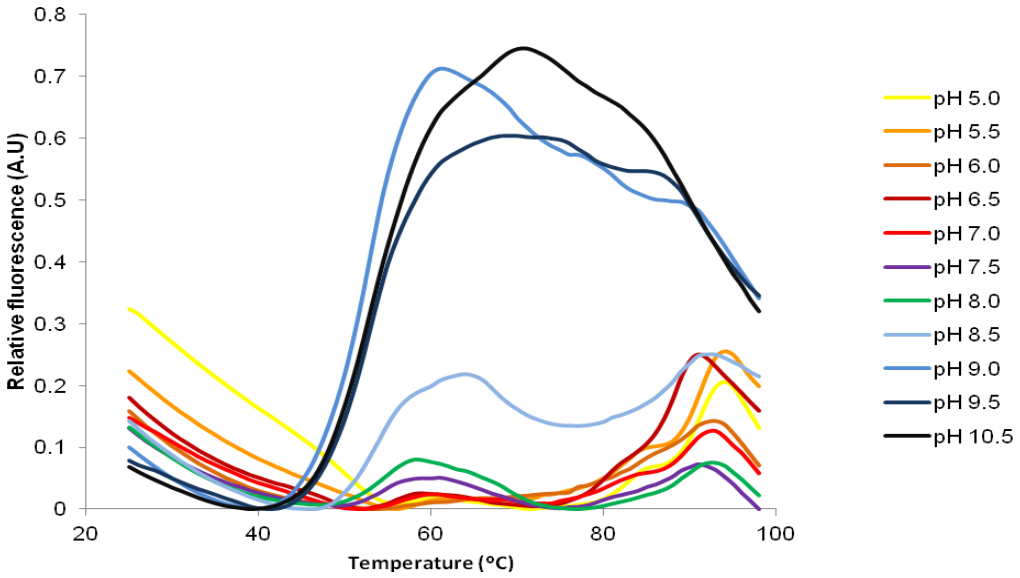


Figure 4.32: Differential scanning fluorimetry determination of ycfD$_{RM}$ thermal denaturation as a function of pH. YcfD$_{RM}$ (1 µM) was prepared in buffers at indicated pH values. It was incubated in the presence of Sypro®Orange and thermally denatured between 25-98°C, whilst the Sypro®Orange fluorescence was monitored.

The DSF profile was then compared to ycfD$_{EC}$ (Figure 4.33). Under the same conditions, ycfD$_{EC}$ showed a markedly lower T_m value (51°C compared to 84°C for ycfD$_{RM}$, Figure 4.33). The thermal

unfolding profile was similar to the profile observed for ycfD$_{RM}$ above pH 8.5. The result suggested that ycfD$_{EC}$ unfolds at a lower temperature compared to ycfD$_{RM}$ and at pH 8.0 (the optimal pH for activity) exposed a higher proportion of hydrophobic patches than ycfD$_{RM}$, leading to a higher absolute fluorescence intensity. For comparative purposes, the signals for ycfD$_{RM}$ and ycfD$_{EC}$ were therefore normalised to maximum fluorescence intensity observed for each protein (Figure 4.33).

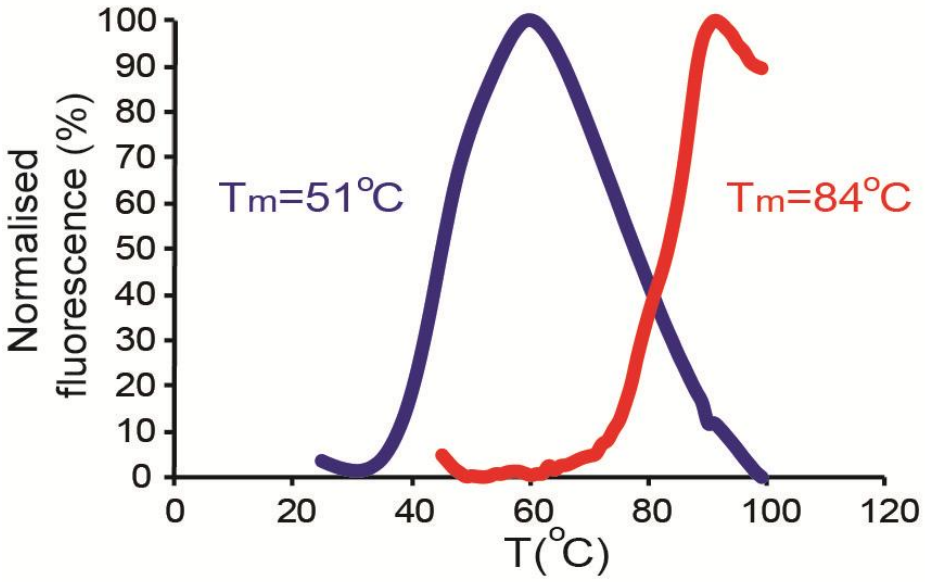


Figure 4.33: DSF spectra showing normalised fluorescence as a function of temperature for ycfD$_{EC}$ (blue) and ycfD$_{RM}$ (red). Enzymes (1 μM) were separately incubated in the presence of Sypro®Orange (1x final concentration) in 50 mM HEPES buffer (pH 8.0 and 7.5 for ycfD$_{EC}$ and ycfD$_{RM}$, respectively). Sypro®Orange fluorescence was measured in a Mini J PCR thermocycler (BioRad) using software Opticon Monitor 3.0. Fluorescence signals were normalised as a percentage of the maximum fluorescence value observed for each thermal denaturation.

4.10.3. Circular dichroism studies

DSF measures a fluorescence signal which depends on the presence and exposure of hydrophobic patches to the solvent (33). In order to gain insight into the nature of thermal unfolding of ycfD$_{RM}$ an alternative biophysical technique was required. Circular dichroism (CD) was chosen because the resulting signal carries information about relative unfolding, but also the relative proportion of secondary structure elements present in the protein at a particular temperature (34), thus providing more information about protein structure than other techniques.

CD measures the difference between the absorption of left-handed and right-handed circularly polarised light as it passes through a chiral biological sample (34). Folded proteins contain asymmetric structural elements, which emit characteristic CD spectra. α-helices and β-strands can be

distinguished by CD spectroscopy due to their different CD spectra. As proteins unfold, they lose the highly asymmetric ordered structures leading to a change in their CD spectra to a signature unfolded state (34).

All CD experiments were performed by Dr. Sarah Wilkins. A broad temperature range was chosen (20-94°C, sampled every 10°C or 2°C when approaching T_m) and CD spectra were collected at 0.5 nm intervals. The CD spectrum for ycfD$_{RM}$ demonstrated a major unfolding event in the temperature range 75-88°C (Figure 4.34). At 94°C, the minimum shifted to 207 nm, suggesting that a large degree of secondary structure was still present (unfolded protein signal shows a minimum at 194 nm (34)).

Data recorded in the 185-240 nm range were analysed using DichroWeb and the CONTIN deconvolution method was used to estimate secondary structural content using reference set 6 (35,36). YcfD$_{RM}$ retained 53% secondary structure at 90°C whereas ycfD$_{EC}$ retained only 46% secondary structure (Figure 4.34). Calculation of the loss of secondary structure between 20°C and 90°C revealed that loss of α-helical structure was 5-times, and of β-strand structure 9.5-times higher for ycfD$_{EC}$ than ycfD$_{RM}$, and the increase in disordered regions was 3-times higher for ycfD$_{EC}$ in the same temperature range (Table 4.6).

Table 4.6: Analysis of CD data analysed using DichroWeb and the CONTIN deconvolution method was used to estimate secondary structural content using reference set 6.

	T(°C)	Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
ycfD$_{EC}$	20	0.1	0.097	0.161	0.079	0.183	0.38	1
	50	0.094	0.082	0.145	0.082	0.161	0.437	1.001
	90	0.092	0.076	0.084	0.06	0.113	0.576	1.001
	Δ(20-90°C)	0.008	0.021	0.077	0.019	0.07	-0.196	-0.001
ycfD$_{RM}$	30	0.097	0.082	0.195	0.09	0.187	0.349	1
	70	0.089	0.079	0.19	0.089	0.175	0.379	1.001
	90	0.068	0.078	0.155	0.092	0.193	0.413	0.999
	Δ(20-90°C)	0.029	0.004	0.04	-0.002	-0.006	-0.064	0.001
	[fold loss EC/RM]	0.28	5.25	1.9	9.5	11.7	3.1	1

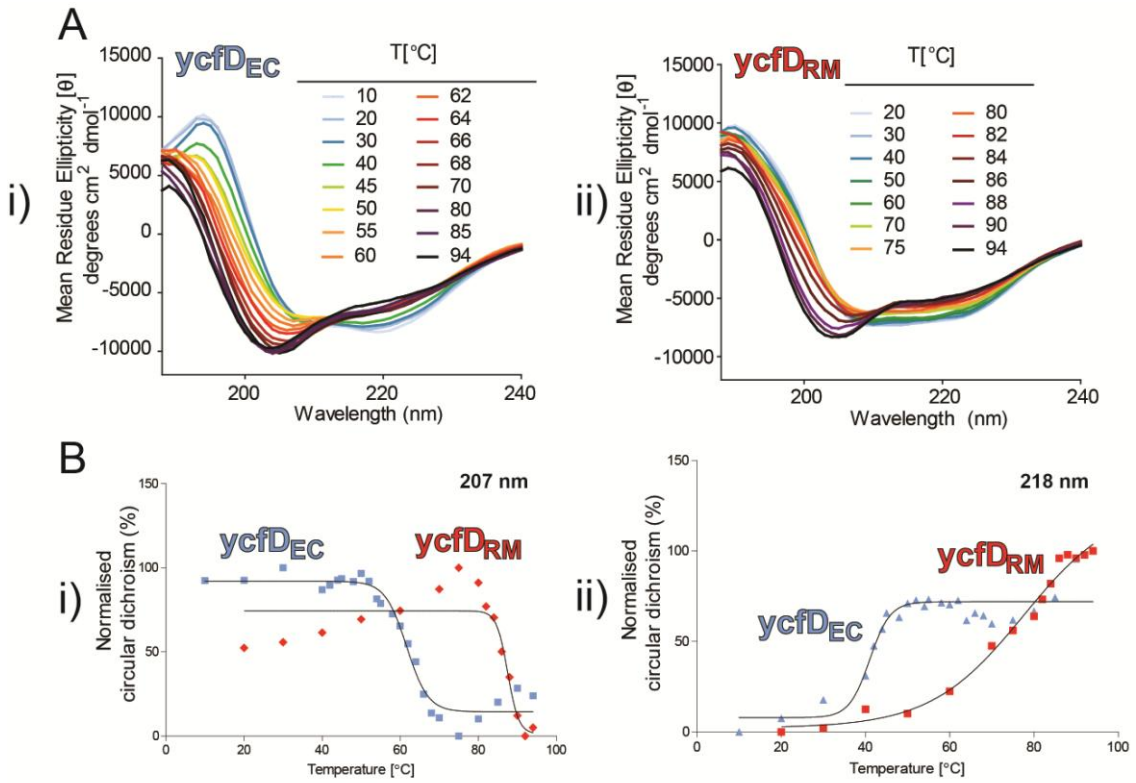


Figure 4.34: CD spectrum of ycfD$_{RM}$ and ycfD$_{EC}$ thermal denaturation. A) CD spectra of i) ycfD$_{EC}$ and ii) ycfD$_{RM}$ (6.7 μM in 10 mM phosphate buffer pH 8.0) were collected on a Chirascan spectropolarimeter (Photophysics) in a 1 cm-pathlength cuvette. Circular dichroism was measured between 190-260 nm (0.5 nm steps). Protein denaturation was measured between 30 and 95°C at 10°C intervals (2°C intervals between 80-94°C). Three measurements were taken for each wavelength and temperature. Automatic smoothing was performed by the software (window size 11). B) Circular dichroism signal measured at i) 207 nm and ii) 218 nm during thermal denaturation. Boltzmann sigmoidal function was used to fit a melting curve to the data (GraphPad Prism version 5.02).

To compare the denaturation spectra of ycfD$_{EC}$ and ycfD$_{RM}$, CD signals were compared at 207 nm and 218 nm (Figure 4.34B). Different secondary structure elements absorb circularly polarised light at different wavelengths, thus the melting profiles differed markedly between 207 nm and 218 nm. Fitting Boltzmann sigmoidal function to the signals revealed variation in T_m^{app} values that depended on the wavelength chosen.

At 207 nm, ycfD$_{EC}$ had a T_m^{app} of 61.6°C and ycfD$_{RM}$ had T_m^{app} of 87.5°C. At 218 nm, which is more frequently used to monitor thermal denaturation, T_m^{app} values were 41°C and 85.3°C for ycfD$_{EC}$ and ycfD$_{RM}$, respectively. These T_m^{app} values were consistent with the values derived from the DSF experiments. From this data it was possible to conclude that T_m^{app} for ycfD$_{RM}$ is at least 25.9°C higher than for ycfD$_{EC}$ (measured at 207 nm); at 218 nm, the difference in T_m^{app} is 44.8°C.

4.10.4. Temperature dependence of activity

Optimal growth temperature for *R. marinus* is 65-70°C and the fastest growth rate occurs at 65°C (6). Protein stability studies suggested that partial unfolding occured at 84°C, above the growth temperature of the organism. Thermophilic proteins are typically adapted for high stability, rather than activity at high temperatures and are frequently less active at low temperatures compared to mesophilic homologues (30). Optimal activity temperature (T_{opt}) of ycfD$_{RM}$ was therefore compared to ycfD$_{EC}$ and a human prolyl hydroxylase PHD2 as another example of a mesophilic protein hydroxylase. The experiment was conducted by Alexandra East under supervision of the author as follows: the enzyme and substrate were separately pre-incubated at the indicated temperature for 3 min prior to the reaction. In order to ensure maximum heat transfer, reactions were conducted in the Genius thermal cycler (Techne) using 500 µL Eppendorf tubes. Enzyme was then added to the substrate mix and the reaction allowed to proceed for 3 min before quenching. Each temperature point was measured in triplicate.

Temperature dependence of mesophilic enzymes ycfD$_{EC}$ and PHD2 was comparable; both enzymes reached T_{opt} at 45°C at which they showed 20-25% activity (Figure 4.35). On the contrary, ycfD$_{RM}$ had a T_{opt} of 55°C with 35-40% activity (Figure 4.35). An increase in T_{opt} was expected for a thermophilic enzyme; nonetheless, the T_{opt} for ycfD$_{RM}$ was 10-15°C below optimal growth temperature (65-70°C) and approximately 30°C below the T_m value of the protein as measured by DSF and CD. This suggested that another factor limits the activity of ycfD$_{RM}$ above 55°C. Likely reasons are thermal instability of co-factors (Fe^{2+}, 2OG, ascorbate) and decreased solubility of O_2 at increased temperatures. It should be noted that T_{opt} below 65°C for *R. marinus* enzymes is not unusual. The range of reported T_{opt} values spans 55-100°C (Table 4.7).

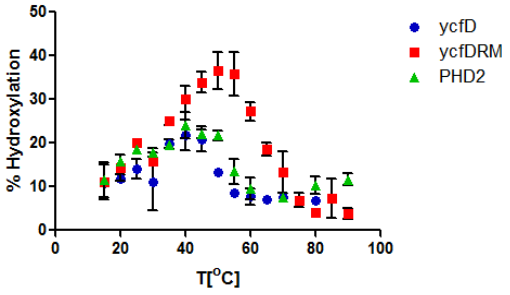


Figure 4.35: Determination of T_{opt} for ycfD$_{EC}$, ycfD$_{RM}$ and PHD2. L16$_{EC(71-90)}$, L16$_{RM\ (72-91)}$ and HIF CODD$_{(556-574)}$ peptides were used (at 10, 50 and 100 µM, respectively) to test the activity of ycfD$_{EC}$, ycfD$_{RM}$ and PHD2. After 3 min pre-incubation and 3 min reaction time, reactions were quenched and analysed by MALDI-MS. Reactions were carried out in triplicate with error bars denoting the standard deviation from the mean.

Table 4.7: Thermostable enzymes from *R. marinus* strains cloned and characterized biochemically.

Gene	Strain	Product	Function	T_{opt} (°C)
polA	ITI 518	DNA polymerase I	DNA synthesis	55°C (37)
ligA	DSM 4253	DNA ligase	DNA ligation	55°C (38)
glgB	JCM 9785	1,4-α-Glucan branching enzyme	Starch biosynthesis	65°C (39)
tdk	ITI 518	Thymidine kinase	Pyrimidine metabolism	65°C (40)
chiA	DSM 4253	Chitinase	Chitin hydrolysis	70°C (41)
mpgP	DSM 4252	Mannosyl-3-phospho-glycerate phosphatase	Mannosylglycerate biosynth.	70–80°C (42)
ppc	JCM 9785	Phosphoenolpyruvate carboxylase	Pyruvate metabolism	75°C (43)
mpgS	DSM 4252	Mannosyl-3-phospho-glycerate synthase	Mannosylglycerate biosynth.	80°C (42)
RmCE	JCM 9785	Cellobiose-2-epimerase	Epilactose biosynthesis	80°C (44)
RmebE	DSM 4252	1,4-α-amylase hydrolase	Starch hydrolysis	80°C (45)
bglA	DSM 4253	β-Glucanase	β-Glucan hydrolysis	85°C (46)
manA	DSM 4252	Mannanase	Mannan hydrolysis	85°C (47)
mgs	DSM 4252	Mannosylglycerate synthase	Mannosylglycerate biosynth.	85–90°C (48)
EC3.2.1.28	DSM 4252	Trehalase	Trehalase uptake	88°C (49)
lamR	DSM 4252	Laminarinase	β-Glucan hydrolysis	88°C (50)
celA	DSM 4253	Cellulase	β-Glucan hydrolysis	100°C (51)

4.10.5. Co-factor thermostability studies

It was important to establish what factors limit the activity of ycfD$_{RM}$ *in vitro*. The previous sections demonstrated that ycfD$_{RM}$ retained structural integrity above 70°C (T_m ~84°C). Its catalytic activity, however, was maximal at 55°C and decreased at higher temperatures. Structural integrity is therefore unlikely to limit ycfD$_{RM}$ activity between 55-80°C. In order to address the question of limiting factors for ycfD$_{RM}$ activity, the enzyme and co-factors/co-substrates were individually incubated at 70°C and 95°C for a prolonged period of time (Figure 4.36). The experiment was performed by Aayan Hussein, under the supervision of the author. Incubating ycfD$_{RM}$ at 70°C for up to 18 hours did not decrease the catalytic activity, whereas incubation at 95°C led to a loss of catalytic activity after 2 h (Figure 4.36a). Other cofactors were then individually incubated either below or above the T_m (70°C or 95°C), before adding them into the hydroxylation reaction with all other co-factors freshly prepared. The results suggested that the reducing factor ascorbate was the most thermo-labile cofactor, while 2OG and Fe(II) did not show thermal degradation (Figure 4.36 b,c,d). Since bacteria do not produce ascorbate (52), it is unlikely that this reducing agent would limit the activity of ycfD$_{RM}$ *in vivo*; other thermostable reducing agents are likely present in *R. marinus*, although no evidence currently exists to support this claim.

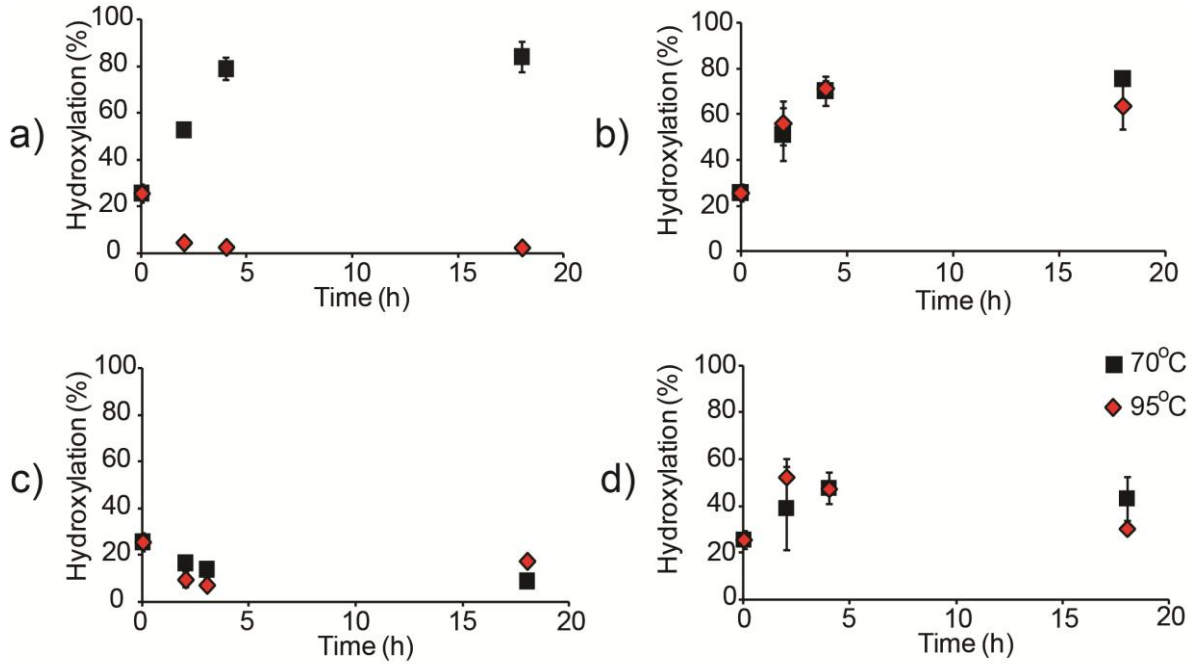


Figure 4.36 Thermal stability of ycfD$_{RM}$ and cofactors, and rate-dependence of ycfD$_{RM}$ on 2OG and Fe(II). Temperature-dependence of co-factor/co-substrate stability: a) ycfD$_{RM}$, b) Fe(II), c) ascorbate and d) 2OG were individually incubated at 70/95°C (denoted by black squares and red diamonds, respectively) for 0, 2, 4 and 18 h, then added to the enzymatic reaction mixture with other components freshly prepared; the conversion of the substrate at 55°C in HEPES buffer (50 mM, pH 7.5) was measured by MALDI-MS. Reactions were performed in triplicate, with standard deviation from the mean denoted by error bars.

4.11. Oxygen dependence of ycfD$_{RM}$ catalytic rate

The high temperature of the habitats of thermophiles exerts evolutionary pressure not only on protein structures, but also on the co-factors. Although most solutes become more soluble with high temperature, the solubility of gases in fact decreases at higher temperature (14). Standard oxygen solubility tables report data up to 50°C, at which solubility is 146 μM (at 37°C, the value is 176 μM; http://www.engineeringtoolbox.com/oxygen-solubility-water-d_841.html). It is possible, therefore, that oxygen supply is limiting for 2OG oxygenase activity at high temperatures. If ycfD$_{RM}$ activity was indeed limited by oxygen solubility, that would be the first example of a bacterial 2OG oxygenase with properties of an oxygen sensor.

YcfD$_{RM}$ is a distant homologue of a JmjC domain-containing asparaginyl hydroxylase Factor Inhibiting HIF (FIH), which is an important regulator of the hypoxic response (53). Kinetic studies on oxygen consumption determined $K_m^{app}(O_2)$ at 90-150 μM for FIH hydroxylation of a HIF-1α substrate fragment (35-mer) (54). FIH thus acts as an oxygen sensor at very low oxygen concentrations. A

comparable $K_m^{app}(O_2)$ value of ycfD$_{RM}$ may enable ycfD$_{RM}$ to act as an oxygen sensor at 70°C, at which O_2 levels are significantly reduced. A Michaelis-Menten kinetic study of ycfD$_{RM}$ dependence on oxygen revealed a K_m^{app} of 11±2% O_2 (100±20 μM, Figure 4.37).

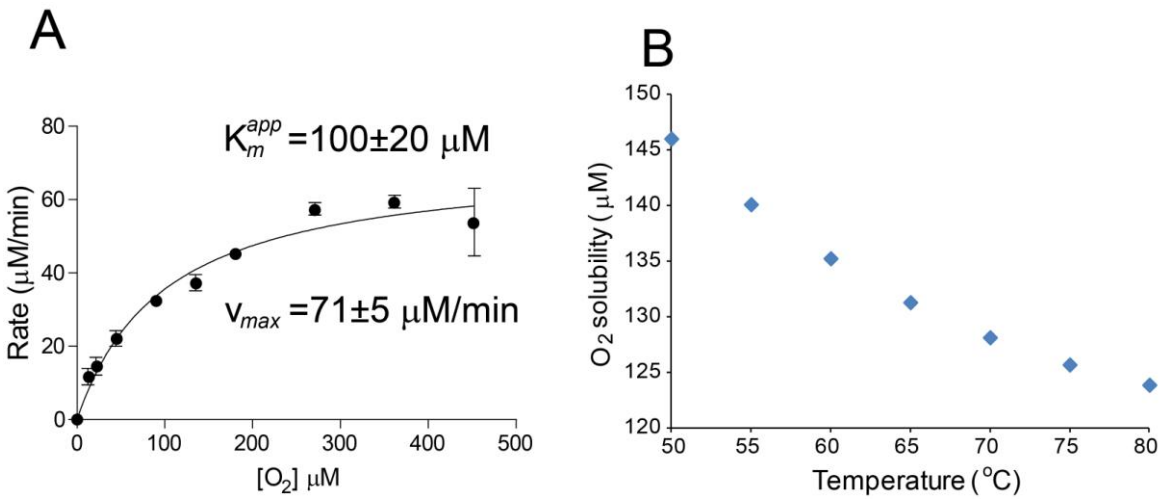


Figure 4.37: Steady-state kinetics of ycfD$_{RM}$ dependence on O_2 and temperature-dependence of O_2 solubility. A) Rate-dependence of ycfD$_{RM}$ on O_2 was determined by incubating ycfD$_{RM}$ (2 μM) with a solution of HEPES buffer (50 mM, pH 7.5), Fe(II), 2OG, ascorbate and L16$_{RM}$ peptide fragment (at 100 μM, 500 μM and 1 mM and 500 μM, respectively) separately pre-equilibrated at indicated concentrations of O_2, controlled by a mass-flow controller (operated by Hanna Tarhonskaya). Reactions were carried out by Hanna Tarhonskaya at 35°C and were quenched after 3 min with equal volume of CF_3COOH_{aq} (1% v/v); conversion was monitored by MALDI-MS. Experiments were conducted in triplicate, with average values shown and error bars denoting standard deviation. B) An O_2 solubility curve (35% salinity) was scaled to μM units by Dr Ian Clifton using previously described methods (14).

With the caveat that a L16$_{RM}$ fragment (20-mer) was used rather than intact protein, it is notable that the $K_m^{app}(O_2)$ for ycfD$_{RM}$ was 100±20 μM (Figure 4.37), which is within the range of oxygen solubility at 70°C (calculated solubility of O_2 at 70°C and 35% salinity is 128 μM (14), Figure 4.37B). The results with isolated ycfD$_{RM}$ are therefore consistent with the temperature-dependent decrease in L16$_{RM}$ hydroxylation levels observed between 60-75°C in *R. marinus* cells presented at the start of this chapter. Collectively, the results suggest an apparently direct relationship between oxygen availability and L16$_{RM}$ hydroxylation status in *R. marinus*, implicating ycfD$_{RM}$ in an oxygen sensing-role that presumably converges on the ribosome.

4.12. Conclusions

Intact protein MS analyses revealed that in cells L16$_{RM}$ hydroxylation levels decrease linearly with increasing temperature, and that the absolute extent of hydroxylation depends on oxygenation of growth medium, which is known to vary between unbaffled and baffled flasks (13). In baffled flasks, which provide better oxygenation of the growth media, relatively higher levels of L16$_{RM}$ hydroxylation were observed, supporting the proposal that the hydroxylation is limited by oxygen availability. The results of thermal denaturation experiments, including DSF and CD, as well as MALDI-MS based kinetic assays demonstrated that protein stability, as well as Fe(II) and 2OG stability are not major limiting factors for ycfD$_{RM}$ activity below 84°C. The $K_m^{app}(O_2)$ for ycfD$_{RM}$ was 100±20μM, which is in the range of oxygen solubility at 70°C, consistent with the temperature-dependent decrease in levels of L16$_{RM}$ hydroxylation observed between 60-75°C in *R. marinus* cells (Fig. 1C).

The identification of a thermophilic oxygenase whose activity in cells is limited by oxygen availability raises the possibility that oxygen sensing mechanisms in bacteria, in particular in thermophiles, extends beyond the currently established [4Fe-4S] cluster- and heme-containing oxygen sensors (55) to encompass members of the 2OG-dependent oxygenases. It is of interest that ycfD$_{RM}$ targets a ribosomal protein that is crucial for the process of translation, thereby potentially connecting translation, metabolism and oxygen sensing. It differs from the known oxygen sensors in that it could potentially regulate translation, rather than transcription, thereby providing a complementary level of gene transcription/translation control.

To date, *R. marinus* is the only thermophilic bacterium found to contain a functional 2OG oxygenase. It may be that the *R. marinus* habitat at ~70°C presents the extreme of environmental conditions that can sustain the catalytic function of a 2OG oxygenase, the activity of which depends on solubility of oxygen, as well as stability of cofactors/co-substrates Fe(II) and 2OG. Given the emerging role of L16 hydroxylation (and more generally hydroxylation of the translation machinery) in regulating global translation rates and growth rates under limiting nutrient conditions, future research should focus on the regulatory effects of O_2-sensitive ribosome hydroxylations, including of L16.

4.13. References

1. Yang, C. G., Yi, C., Duguid, E. M., Sullivan, C. T., Jian, X., Rice, P. A., and He, C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961-965
2. Feller, G. (2013) Psychrophilic enzymes: from folding to function and biotechnology. *Scientifica* **2013**, 512840
3. Oren, A., Heldal, M., Norland, S., and Galinski, E. A. (2002) Intracellular ion and organic solute concentrations of the extremely halophilic bacterium Salinibacter ruber. *Extremophiles : life under extreme conditions* **6**, 491-498
4. Karlsson, E. N., Hachem, M. A., Ramchuran, S., Costa, H., Holst, O., Svenningsen, S. F., and Hreggvidsson, G. O. (2004) The modular xylanase Xyn10A from Rhodothermus marinus is cell-attached, and its C-terminal domain has several putative homologues among cell-attached proteins within the phylum Bacteroidetes. *FEMS microbiology letters* **241**, 233-242
5. Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806
6. Alfredsson, G. A., Kristjansson, J. K., Hjorleifsdottir, S., and Stetter, K. O. (1988) Rhodothermus-Marinus, Gen-Nov, Sp-Nov, a Thermophilic, Halophilic Bacterium from Submarine Hot Springs in Iceland. *J Gen Microbiol* **134**, 299-306
7. Bjornsdottir, S. H., Blondal, T., Hreggvidsson, G. O., Eggertsson, G., Petursdottir, S., Hjorleifsdottir, S., Thorbjarnardottir, S. H., and Kristjansson, J. K. (2006) Rhodothermus marinus: physiology and molecular biology. *Extremophiles* **10**, 1-16
8. Chen, R., and Chen-Schmeisser, U. (1977) Isopeptide linkage between N-alpha-monomethylalanine and lysine in ribosomal protein S11 from Escherichia coli. *Proc Natl Acad Sci U S A* **74**, 4905-4908
9. Kowalak, J. A., and Walsh, K. A. (1996) beta-methylthio-aspartic acid: Identification of a novel posttranslational modification in ribosomal protein S12 from Escherichia coli. *Protein Sci* **5**, 1625-1632
10. Vanet, A., Plumbridge, J. A., Guerin, M. F., and Alix, J. H. (1994) Ribosomal protein methylation in Escherichia coli: the gene prmA, encoding the ribosomal protein L11 methyltransferase, is dispensable. *Mol Microbiol* **14**, 947-958
11. Chang, F. N. (1978) Temperature-dependent variation in the extent of methylation of ribosomal proteins L7 and L12 in Escherichia coli. *J Bacteriol* **135**, 1165-1166
12. Brosius, J., and Chen, R. (1976) The primary structure of protein L16 located at the peptidyltransferase center of Escherichia coli ribosomes. *FEBS Lett* **68**, 105-109
13. McDaniel, L. E., Bailey, E. G., and Zimmerli, A. (1965) Effect of Oxygen Supply Rates on Growth of Escherichia Coli .I. Studies in Unbaffled and Baffled Shake Flasks. *Appl Microbiol* **13**, 115-119
14. Tromans, D. (1998) Temperature and pressure dependent solubility of oxygen in water: a thermodynamic analysis. *Hydrometallurgy* **48**, 327-342
15. Ge, W., Wolf, A., Feng, T., Ho, C. H., Sekirnik, R., Zayer, A., Granatino, N., Cockman, M. E., Loenarz, C., Loik, N. D., Hardy, A. P., Claridge, T. D., Hamed, R. B., Chowdhury, R., Gong, L., Robinson, C. V., Trudgian, D. C., Jiang, M., Mackeen, M. M., McCullagh, J. S., Gordiyenko, Y., Thalhammer, A., Yamamoto, A., Yang, M., Liu-Yi, P., Zhang, Z., Schmidt-Zachmann, M., Kessler, B. M., Ratcliffe, P. J., Preston, G. M., Coleman, M. L., and Schofield, C. J. (2012) Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol* **8**, 960-962
16. Liang, X., Kaya, A., Zhang, Y., Le, D. T., Hua, D., and Gladyshev, V. N. (2012) Characterization of methionine oxidation and methionine sulfoxide reduction using methionine-rich cysteine-free proteins. *BMC biochemistry* **13**, 21
17. Hopkinson, R. J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K. H., Li, X. S., Kristensen, J. B. L., King, O. N. F., Chan, M. C., Yeoh, K. K., Choi, H., Walport, L. J.,

Thinnes, C. C., Bush, J. T., Lejeune, C., Rydzik, A. M., Rose, N. R., Bagg, E. A., McDonough, M. A., Krojer, T. J., Yue, W. W., Ng, S. S., Olsen, L., Brennan, P. E., Oppermann, U., Muller, S., Klose, R. J., Ratcliffe, P. J., Schofield, C. J., and Kawamura, A. (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci* **4**, 3110-3117

18. Radestock, S., and Gohlke, H. (2011) Protein rigidity and thermophilic adaptation. *Proteins* **79**, 1089-1108
19. Kumar, S., and Nussinov, R. (2001) How do thermophilic proteins deal with heat? *Cell Mol Life Sci* **58**, 1216-1233
20. Ng, S. S., Kavanagh, K. L., McDonough, M. A., Butler, D., Pilka, E. S., Lienard, B. M., Bray, J. E., Savitsky, P., Gileadi, O., von Delft, F., Rose, N. R., Offer, J., Scheinost, J. C., Borowski, T., Sundstrom, M., Schofield, C. J., and Oppermann, U. (2007) Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87-91
21. Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806
22. Schlippe, Y. V., Hartman, M. C., Josephson, K., and Szostak, J. W. (2012) In vitro selection of highly modified cyclic peptides that act as tight binding inhibitors. *J Am Chem Soc* **134**, 10469-10477
23. Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C., and Ramakrishnan, V. (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* **16**, 528-533
24. Sako, Y., Goto, Y., Murakami, H., and Suga, H. (2008) Ribosomal synthesis of peptidase-resistant peptides closed by a nonreducible inter-side-chain bond. *ACS Chem Biol* **3**, 241-249
25. Amann, E., Ochs, B., and Abel, K. J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene* **69**, 301-315
26. Schofield, C. J., and Zhang, Z. (1999) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol* **9**, 722-731
27. Crawford, L., Stepan, A. M., McAda, P. C., Rambosek, J. A., Conder, M. J., Vinci, V. A., and Reeves, C. D. (1995) Production of cephalosporin intermediates by feeding adipic acid to recombinant Penicillium chrysogenum strains expressing ring expansion activity. *Biotechnology (N Y)* **13**, 58-62
28. Wei, C. L., Yang, Y. B., Wang, W. C., Liu, W. C., Hsu, J. S., and Tsai, Y. C. (2003) Engineering Streptomyces clavuligerus deacetoxycephalosporin C synthase for optimal ring expansion activity toward penicillin G. *Appl Environ Microbiol* **69**, 2306-2312
29. Shibasaki, T., Mori, H., and Ozaki, A. (1999) Development of the regio- and stereospecific proline hydroxylases and their application. *J Syn Org Chem Japan* **57**, 523-531
30. Lam, S. Y., Yeung, R. C., Yu, T. H., Sze, K. H., and Wong, K. B. (2011) A rigidifying salt-bridge favors the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity. *PLoS Biol* **9**, e1001027
31. Szilagyi, A., and Zavodszky, P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493-504
32. Chan, C. H., Yu, T. H., and Wong, K. B. (2011) Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *PLoS One* **6**, e21624
33. Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-2221
34. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* **1**, 2876-2890
35. Provencher, S. W., and Glockner, J. (1981) Estimation of Globular Protein Secondary Structure from Circular-Dichroism. *Biochemistry-Us* **20**, 33-37

36. Vanstokkum, I. H. M., Spoelder, H. J. W., Bloemendal, M., Vangrondelle, R., and Groen, F. C. A. (1990) Estimation of Protein Secondary Structure and Error Analysis from Circular-Dichroism Spectra. *Anal Biochem* **191**, 110-118
37. Blondal, T., Thorbjarnardottir, S. H., Kieleczawa, J., Hjorleifsdottir, S., Kristjansson, J. K., Einarsson, J. M., and Eggertsson, G. (2001) Cloning, sequence analysis and functional characterization of DNA polymerase I from the thermophilic eubacterium Rhodothermus marinus. *Biotechnol Appl Biochem* **34**, 37-45
38. Thorbjarnardottir, S. H., Jonsson, Z. O., Andresson, O. S., Kristjansson, J. K., Eggertsson, G., and Palsdottir, A. (1995) Cloning and sequence analysis of the DNA ligase-encoding gene of Rhodothermus marinus, and overproduction, purification and characterization of two thermophilic DNA ligases. *Gene* **161**, 1-6
39. Shinohara, M. L., Ihara, M., Abo, M., Hashida, M., Takagi, S., and Beck, T. C. (2001) A novel thermostable branching enzyme from an extremely thermophilic bacterial species, Rhodothermus obamensis. *Appl Microbiol Biot* **57**, 653-659
40. Blondal, T., Thorbjarnardottir, S. H., Kieleczawa, J., Einarsson, J. M., Hjorleifsdottir, S., Kristjansson, J. K., and Eggertsson, G. (1999) Cloning, sequence analysis and overexpression of a rhodothermus marinus gene encoding a thermostable thymidine kinase. *FEMS microbiology letters* **179**, 311-316
41. Hobel, C. F., Hreggvidsson, G. O., Marteinsson, V. T., Bahrani-Mougeot, F., Einarsson, J. M., and Kristjansson, J. K. (2005) Cloning, expression, and characterization of a highly thermostable family 18 chitinase from Rhodothermus marinus. *Extremophiles : life under extreme conditions* **9**, 53-64
42. Borges, N., Marugg, J. D., Empadinhas, N., da Costa, M. S., and Santos, H. (2004) Specialized roles of the two pathways for the synthesis of mannosylglycerate in osmoadaptation and thermoadaptation of Rhodothermus marinus. *J Biol Chem* **279**, 9892-9898
43. Takai, K., Sako, Y., and Uchida, A. (1998) ppc, the gene for phosphoenolpyruvate carboxylase from an extremely thermophilic bacterium, Rhodothermus obamensis: cloning, sequencing and overexpression in Escherichia coli. *Microbiol-Sgm* **144**, 1423-1434
44. Ojima, T., Saburi, W., Sato, H., Yamamoto, T., Mori, H., and Matsui, H. (2011) Biochemical characterization of a thermophilic cellobiose 2-epimerase from a thermohalophilic bacterium, Rhodothermus marinus JCM9785. *Biosci Biotechnol Biochem* **75**, 2162-2168
45. Yoon, S. A., Ryu, S. I., Lee, S. B., and Moon, T. W. (2008) Purification and characterization of branching specificity of a novel extracellular amylolytic enzyme from marine hyperthermophilic Rhodothermus marinus. *J Microbiol Biotechnol* **18**, 457-464
46. Spilliaert, R., Hreggvidsson, G. O., Kristjansson, J. K., Eggertsson, G., and Palsdottir, A. (1994) Cloning and sequencing of a Rhodothermus marinus gene, bglA, coding for a thermostable beta-glucanase and its expression in Escherichia coli. *Eur J Biochem* **224**, 923-930
47. Politz, O., Krah, M., Thomsen, K. K., and Borriss, R. (2000) A highly thermostable endo-(1,4)-beta-mannanase from the marine bacterium Rhodothermus marinus. *Appl Microbiol Biotechnol* **53**, 715-721
48. Martins, L. O., Empadinhas, N., Marugg, J. D., Miguel, C., Ferreira, C., da Costa, M. S., and Santos, H. (1999) Biosynthesis of mannosylglycerate in the thermophilic bacterium Rhodothermus marinus. Biochemical and genetic characterization of a mannosylglycerate synthase. *J Biol Chem* **274**, 35407-35414
49. Jorge, C. D., Sampaio, M. M., Hreggvidsson, G. O., Kristjanson, J. K., and Santos, H. (2007) A highly thermostable trehalase from the thermophilic bacterium Rhodothermus marinus. *Extremophiles : life under extreme conditions* **11**, 115-122
50. Krah, M., Misselwitz, R., Politz, O., Thomsen, K. K., Welfle, H., and Borriss, R. (1998) The laminarinase from thermophilic eubacterium Rhodothermus marinus--conformation, stability, and identification of active site carboxylic residues by site-directed mutagenesis. *Eur J Biochem* **257**, 101-111

51. Halldorsdottir, S., Thorolfsdottir, E. T., Spilliaert, R., Johansson, M., Thorbjarnardottir, S. H., Palsdottir, A., Hreggvidsson, G. O., Kristjansson, J. K., Holst, O., and Eggertsson, G. (1998) Cloning, sequencing and overexpression of a Rhodothermus marinus gene encoding a thermostable cellulase of glycosyl hydrolase family 12. *Appl Microbiol Biotechnol* **49**, 277-284
52. Lee, B. H., Huh, W. K., Kim, S. T., Lee, J. S., and Kang, S. O. (1999) Bacterial production of D-erythroascorbic acid and L-ascorbic acid through functional expression of Saccharomyces cerevisiae D-arabinono-1,4-lactone oxidase in Escherichia coli. *Appl Environ Microbiol* **65**, 4685-4687
53. Schofield, C. J., and Ratcliffe, P. J. (2004) Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Biol* **5**, 343-354
54. Ehrismann, D., Flashman, E., Genn, D. N., Mathioudakis, N., Hewitson, K. S., Ratcliffe, P. J., and Schofield, C. J. (2007) Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem J* **401**, 227-234
55. Green, J., Crack, J. C., Thomson, A. J., and LeBrun, N. E. (2009) Bacterial sensors of oxygen. *Curr Opin Microbiol* **12**, 145-151

5. Structural studies on ycfD$_{RM}$

Contents

5. Structural studies on ycfD$_{RM}$ 181

5.1. Aims 183

5.2. Stabilisation and inhibition studies on ycfD$_{RM}$ 183

5.2.1. The DSF thermal stabilisation assay identifies ycfD$_{RM}$ stabilisers 184

5.2.2. TCA cycle intermediate inhibition of ycfD$_{RM}$ 185

5.2.3. TCA cycle intermediate inhibition of ycfD$_{EC}$ 187

5.3. Crystallization trials with ycfD$_{RM}$ 188

5.3.1. Anion Exchange Chromatography 188

5.3.2. Crystallisation of ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$ complex 191

5.3.3. Optimisation of the JCSG_A11 crystallisation condition 191

5.3.4. Crystallisation of a ycfD$_{RM}$.Mn(II).FG2216 complex 193

5.3.5. Data collection and processing 194

5.4. Crystal structure of ycfD$_{RM}$ 195

5.4.1. Overall fold 195

5.4.2. Topology analysis of ycfD$_{RM}$ 197

5.4.3. Salt bridge analysis 198

5.4.4. Cation-π interactions 199

5.4.5. Hydrophobic interactions 199

5.5. The DSBH and substrate binding of ycfD$_{RM}$ 199

5.5.1. DSBH strand βI 200

5.5.2. DSBH strands βII-βVIII 202

5.6. Comparison of VioC and ycfD$_{RM}$ substrate complex structures 206

5.7. NOG binding 208

5.7.1. Substitution of the basic 2OG-binding residue on βVIII 210

5.8. Induced fit upon L16 Binding 212

5.8.1. Comparison of ycfD$_{RM}$ and ycfD$_{EC}$ substrate-binding residue conformations 213

5.9. A hydrophobic pocket accommodates the substrate residue V81 215

5.10. The dimerisation domain 215

5.10.1. Testing the effect of salt-bridge interactions in the ycfD dimerisation domain 217

5.10.2. Disruption of dimerisation leads to loss of catalytic activity 218

5.11. The *C*-terminal ('winged-helix') domain 221

5.12. Towards the structure of a ycfD.L16 protein-protein complex 223

5.13. Conclusions 227

5.14. References 229

5.1. Aims

The work described in the previous chapter identified ycfD$_{RM}$ as a thermophilic 2OG oxygenase catalysing hydroxylation of Arg82 on the ribosomal protein L16$_{RM}$. Biochemical aspects of L16 hydroxylation were described, including its co-factor and co-substrate dependence, thermal stability and temperature dependence of activity.

The aim of the work described in this chapter was to derive a structural understanding of L16 hydroxylation by the ycfD family of proteins, including ycfD$_{EC}$ and ycfD$_{RM}$. T_m shift assays were conducted in the presence of known 2OG oxygenase inhibitors, which identified small molecules with potent stabilising effect on ycfD$_{RM}$ and ycfD$_{EC}$ proteins. This was followed by crystallisation studies, which identified optimal crystallisation conditions for ycfD$_{RM}$. A ycfD$_{RM}$ crystal structure in the presence and absence of L16$_{RM}$ substrate fragment is described. The crystal structure of ycfD$_{RM}$ is analysed in comparison with that of ycfD$_{EC}$ in order to identify structural features of thermal stabilisation. L16$_{RM}$ binding to ycfD$_{RM}$ is discussed in detail; mutational studies were undertaken on substrate-binding residues in ycfD$_{EC}$ to better understand structural observations in the ycfD$_{RM}$.L16$_{RM}$ complex. The structure of ycfD$_{RM}$ in complex with a substrate fragment was used to design a ycfD$_{EC}$ variant for cross-linking to L16 protein. Expression and purification of a ycfD.L16 protein-protein complex is described and native mass spectrometry studies on the complex were undertaken in order to study the stoichiometry of binding. Finally, the understanding of the substrate binding mode for ycfD$_{RM}$ was applied to other ROX, including NO66 and MINA53, to model ribosomal protein binding by ROX. Ultimately, the understanding of the mechanism of ROX substrate binding may help to rationalise the findings that mutations in substrate-binding loop of cancer-associated ROX MINA53 affect the *in vivo* activity of MINA53, as described in Chapter 6.

5.2. Stabilisation and inhibition studies on ycfD$_{RM}$

In order to increase the chances of successful protein crystallisation, a study of small molecule-induced protein thermal stabilisation is useful. An in-house library of small molecules containing >200 compounds that were shown to inhibit at least one enzyme from the 2OG

oxygenase/demethylase superfamily (1) was tested for ycfD$_{RM}$ stabilisation. The library contains a range of distinct chemotypes, including TCA cycle intermediates, iron chelators, *N*-oxalyl tyrosine derivatives, pyridine dicarboxylic acids, and natural products.

Differential scanning fluorimetry (DSF) has successfully been used to identify potent stabilisers of a wide range of 2OG oxygenases (for review see Rose *et al*.(1)). The DSF technique was discussed in detail in the previous chapter. Briefly, upon thermal unfolding of a protein the hydrophobic patches become solvent-exposed, leading to binding of the dye and an increased quenching of the fluorescence signal. Small molecule binders stabilise the folding of a protein, resulting in an increase in the temperature required to denature the protein. Small molecules can therefore be assessed by the ΔT_m; and it is frequently found that an increase in ΔT_m provides a good indication of potency of inhibition (2). Due to the relative simplicity, low cost and high throughput of the DSF assay, it is a useful technique for an initial screen of inhibitors.

5.2.1. The DSF thermal stabilisation assay identifies ycfD$_{RM}$ stabilisers

The previous chapter described initial DSF studies on ycfD$_{RM}$ which determined the T_m^{app} of ycfD$_{RM}$ to be 84°C. The finding was corroborated by CD studies, which independently determined T_m^{app} to be 85°C. In this chapter, the use of DSF is described as a preliminary stabiliser screen.

The conditions of the DSF experiment described in Chapter 4 were used for screening a library of compounds. Raw data were visually inspected for local minima and maxima values (> 60^{o}C) and fitted with the Boltzmann sigmoidal function (GraphPad Prism, version 5.02). A compound was considered as a stabiliser if it led to ΔT_m >+2°C relative to DMSO control (Figure 5.1). The experiment was conducted three times to minimise the rate of false-positive signals. As shown in Table 5.1, three compounds consistently increased ΔT_m (the complete dataset is presented in the Appendix). Two TCA cycle intermediates, citrate and succinate, increased T_m by 4°C and 5°C, respectively, whilst the inactive 2OG analogue, *N*-oxalyl-glycine (NOG) did not lead to a large increase in T_m. FG2216, originally identified as a selective inhibitor of collagen prolyl hydroxylase, led to largest ΔT_m of 6°C.

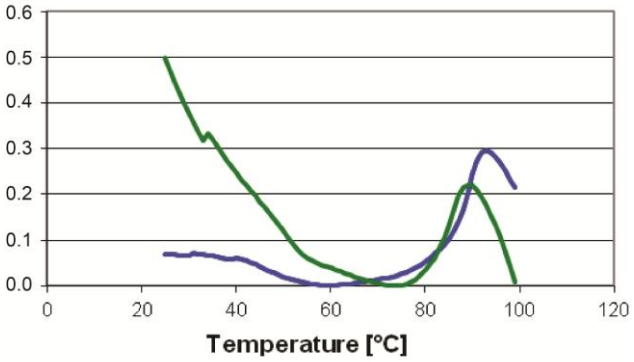
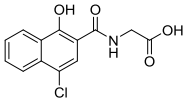
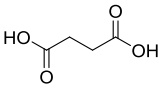
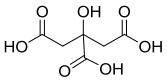


Figure 5.1: DSF experiment identifies succinate as a potent stabiliser of ycfD$_{RM}$. YcfD$_{RM}$ (2 μM) was incubated with Mn(II) (50 μM) and succinate (100 μM) in the presence of SyproOrange^{TM}. The thermal denaturation was measured by monitoring the relative fluorescence of SyproOrange^{TM} as a function of temperature between 25-98^{o}C. The resulting curves are fitted with a Boltzmann sigmoidal function (GraphPad Prism, version 5.02) between the minimum and maximum of the sigmoidal part of the curve. The T_m was then calculated as 84.7°C for the control (green), and T_m 89.5°C for the 2OG analogue (blue), indicating a ΔT_m of 4.8^{o}C.

Table 5.1: DSF determination of ΔT_m for ycfD$_{RM}$ stabilisers

	Chemical structure	ΔT_m (°C)
FG2216		6°C
Succinate		5°C
Citrate		4°C

5.2.2. TCA cycle intermediate inhibition of ycfD$_{RM}$

The large ΔT_m for the TCA cycle intermediates citrate and succinate led to a more thorough investigation of their inhibitory effects on ycfD$_{RM}$ activity. IC$_{50}$ values were determined for eight TCA cycle intermediates (citrate, cis-aconitate, D-isocitrate, *N*-oxalyl glycine, succinate, fumarate, malate and oxaloacetate). The concentration of 2OG was below the K_m^{app} (20 μM) to allow competitive inhibition to be studied. The results revealed succinate as the most potent inhibitor, with an IC$_{50}$ of 45±1.5 μM, consistent with its observed increase in ΔT_m (Figure 5.2). NOG inhibited ycfD$_{RM}$ relatively poorly, with an IC$_{50}$ of 128±2μM, a 2.8-fold increase relative to succinate. Consistent with its relatively poor inhibition, NOG did not show a significant increase in T_m. Fumarate, a desaturated analogue of succinate, showed inhibition with IC$_{50}$ 90±1.3 μM, and citrate inhibited ycfD$_{RM}$ with IC$_{50}$ 112±1.3 μM (Figure 5.2). *cis*-aconitate, D-isocitrate, malate and oxaloacetate showed no inhibition.

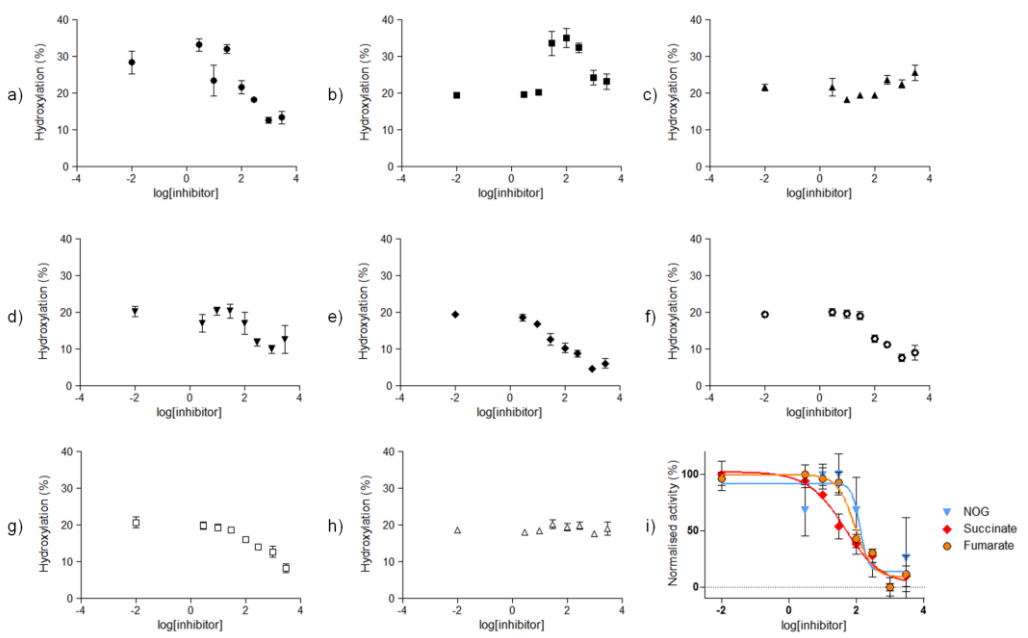


Figure 5.2: TCA cycle intermediate inhibition of ycfD$_{RM}$. ycfD$_{RM}$ (2 μM) was incubated with Fe^{2+}, 2OG, ascorbate and a 20-mer L16$_{RM}$ substrate fragment (at 100 μM, 20 μM, 1 mM and 100 μM , respectively) in the presence of a serial dilution of TCA cycle intermediates. Reactions were allowed to proceed for 15 min at 55°C before quenching with CF_3COOH_{aq} (1%) and results were analysed by MALDI-MS. The average value of three experiments is shown, with error bars denoting s.d. from the mean. a) citrate, b) cis-aconitate, c) D-isocitrate, d) *N*-oxalyl glycine, e) succinate, f) fumarate, g) malate, h) oxaloacetate, i) normalised activity of ycfD$_{RM}$ in presence of NOG (cyan), succinate (red) and fumarate, fitted with log(inhibitor) vs. Response function (GraphPad Prism, version 5.02).

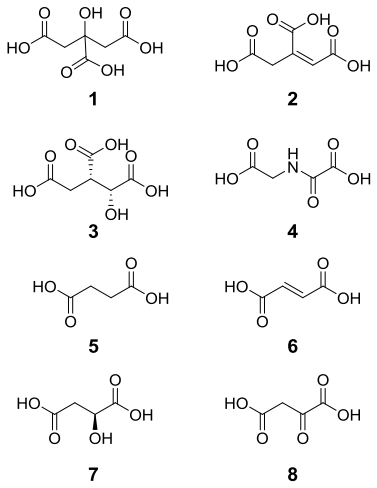
The low IC_{50} value for succinate in combination with a high IC_{50} for NOG suggested a potential feedback mechanism based on product inhibition. The experiment was independently replicated twice to confirm the validity of the result. The high IC_{50} value for NOG was surprising given that NOG is the closest analogue of the ycfD$_{RM}$ co-substrate 2OG and is frequently found to be a potent inhibitor of 2OG oxygenases (1). The relatively greater inhibition of a 2OG oxygenase by the catalytic by-product succinate compared to substrate analogue NOG may suggest structural differences of the active site, which may be not be optimal for the binding of 2OG; this is consistent with the relatively high K_m^{app} observed for 2OG (70 μM, Chapter 4) compared to other 2OG oxygenases (typically ~10 μM) (3). Furthermore, the result, if true for ycfD$_{EC}$ as well, could explain the product inhibition observed for ycfD$_{EC}$ hydroxylation of the L16$_{EC}$ fragment peptide (Chapter 4): succinate produced during the catalysis of L16$_{EC}$ hydroxylation competes with 2OG for binding of the catalytic Fe(II) and thereby decreases the rate of hydroxylation.

5.2.3. TCA cycle intermediate inhibition of ycfD$_{EC}$

For comparison, TCA cycle intermediates were tested as inhibitors of ycfD$_{EC}$. The results again revealed succinate as the most potent inhibitor, with an IC_{50} value comparable to that for ycfD$_{RM}$ (53±1.7 μM, Figure 5.4). NOG and fumarate also inhibited ycfD, but their IC_{50} values were higher, 455±2.2 and 426±2.2 μM, respectively. Although isocitrate showed partial inhibition at high concentrations, IC_{50} could not be fitted to the data; concentrations above 10 mM were deemed biologically irrelevant and were not tested. The TCA cycle intermediates were tested for thermal stabilisation of ycfD$_{EC}$. Consistent with the data presented above, succinate was found to be the most potent stabiliser, leading to ΔT_m of 6°C (Figure 5.3).

Table 5.2: IC_{50} values for ycfD$_{RM}$ with NOG and the TCA cycle intermediates.

IC_{50} (μM)		ycfD$_{RM}$	ycfD$_{EC}$
Citrate	**1**	112±1.3	No inhibition
cis-Aconitate	**2**	No inhibition	No inhibition
D-Isocitrate	**3**	No inhibition	No inhibition
NOG	**4**	128±2	455±2.2
Succinate	**5**	45±1.5	53±1.7
Fumarate	**6**	90±1.3	426±2.2
Malate	**7**	No inhibition	No inhibition
Oxaloacetate	**8**	No inhibition	No inhibition



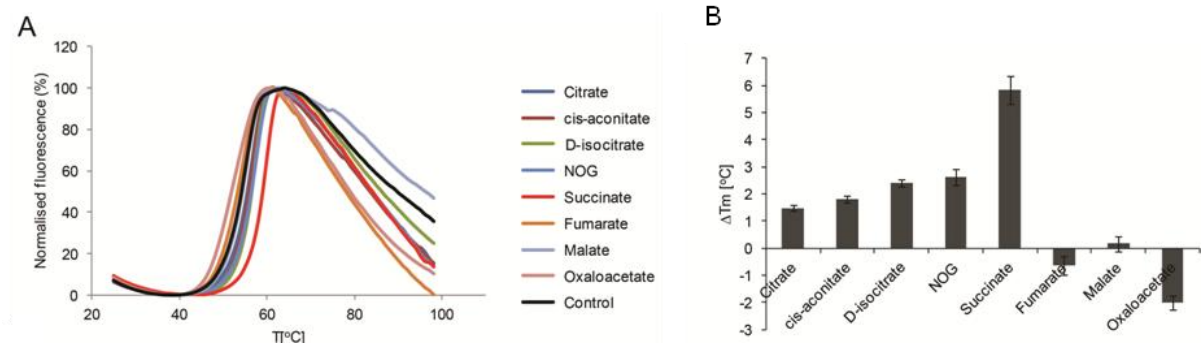


Figure 5.3: DSF measurements of TCA cycle intermediate stabilisation of ycfD$_{EC}$. A) SyproOrange^{TM} fluorescence of ycfD$_{EC}$ incubated with TCA cycle intermediates; b) ΔT_m analysis of melting curves presented in a).

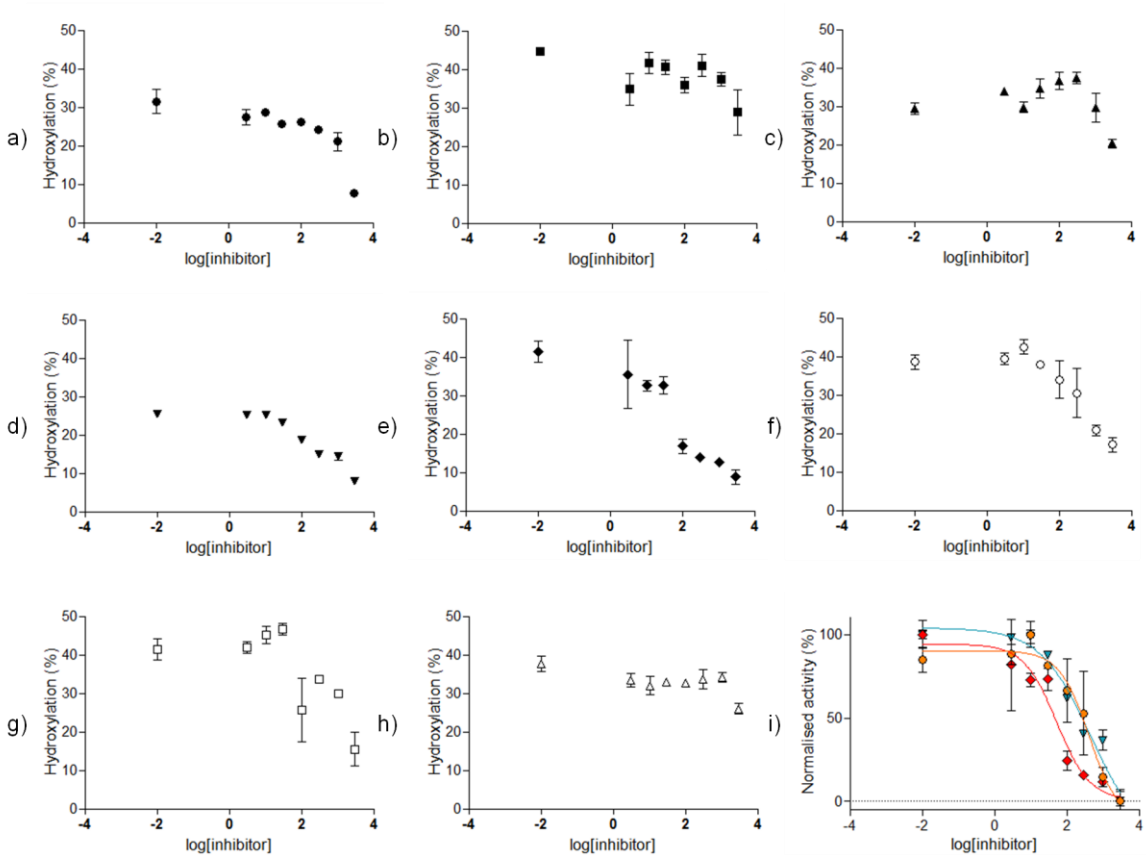


Figure 5.4: TCA cycle intermediate inhibition of ycfD$_{EC}$. YcfD$_{EC}$ (2 μM) was incubated with Fe^{2+}, 2OG, ascorbate and a 20-mer L16$_{EC}$ substrate fragment (at 100 μM, 20 μM, 1 mM and 100 μM , respectively) in the presence of a serial dilution of TCA cycle intermediates. Reaction mixtures were incubated for 15 min at 37°C before quenching with CF_3COOH_{aq} (1% vol/vol) and the results were analysed by MALDI-MS. The average value of three experiments is shown, with error bars denoting s.d. from the mean. a) citrate, b) cis-aconitate, c) D-isocitrate, d) *N*-oxalyl glycine, e) succinate, f) fumarate, g) malate, h) oxaloacetate, i) normalised activity of ycfD in the presence of NOG (cyan), succinate (red) and fumarate, fitted with 'log(inhibitor) vs. Response' function (GraphPad Prism, version 5.02).

5.3. Crystallisation trials with ycfD$_{RM}$

Initially, crystallization trials were conducted with ycfD$_{RM}$ prepared as described in Chapter 4. A broad screen of conditions (8×96×3 in total) was tested for crystallisation at 4°C and 20°C. No protein crystals were observed after 1 month, suggesting that the protein sample is not homogeneous and may, despite relatively high purity as determined by SDS-PAGE, contain a mixture of oligomerisation/folding states.

5.3.1. Anion Exchange Chromatography

Several alternative purification options were considered. Aside from size exclusion chromatography, ion exchange chromatography is the most common approach to purify proteins, relying on proteins carrying a net charge at a particular pH value.

The pKa calculated for the 6xHis-ycfD$_{RM}$ construct was 5.3 (ProtParam software; http://web.expasy.org/protparam/). Assuming an equal distribution of charged residues over the protein surface, the overall charge of the protein at pH 7.5 should thus be negative and protein should bind to an anion-exchange resin. In order to test for heterogeneity of the folding states in the sample, a strong anion exchange (quaternary amine) monolithic column was used for analysis of ycfD$_{RM}$. YcfD$_{RM}$ was purified by HisTrap Ni(II) affinity column as described in Chapter 4. The protein sample was buffer exchanged into 50 mM HEPES pH 7.5 by gel filtration chromatography (S200, 300 mL). The protein sample was then applied onto a strong anion-exchange monolithic column (CIM QA-1, 1 mL, BIA Separations). Initially, a shallow linear gradient of 0-100% buffer B: (1 M NaCl in 50 mM HEPES pH 7.5) was used over 45 column volumes at high flow-rate (10 mL/min). A sharp peak with an unresolved shoulder was observed at ~30% of buffer B (Figure 5.5), suggesting that the sample purified by Ni(II) affinity and gel filtration columns is not uniform. The eluate was concentrated and re-injected (2 mg total), then eluted with a stepped gradient (Figure 5.5b). No protein was observed up to 35% buffer B. At 35% B elution of a sharp protein peak was observed; as %B was increased in steps of 5% up to 50% B, two more major protein peaks were observed, suggesting that ycfD$_{RM}$ purified by Ni(II)-affinity chromatography is composed of at least three major forms.

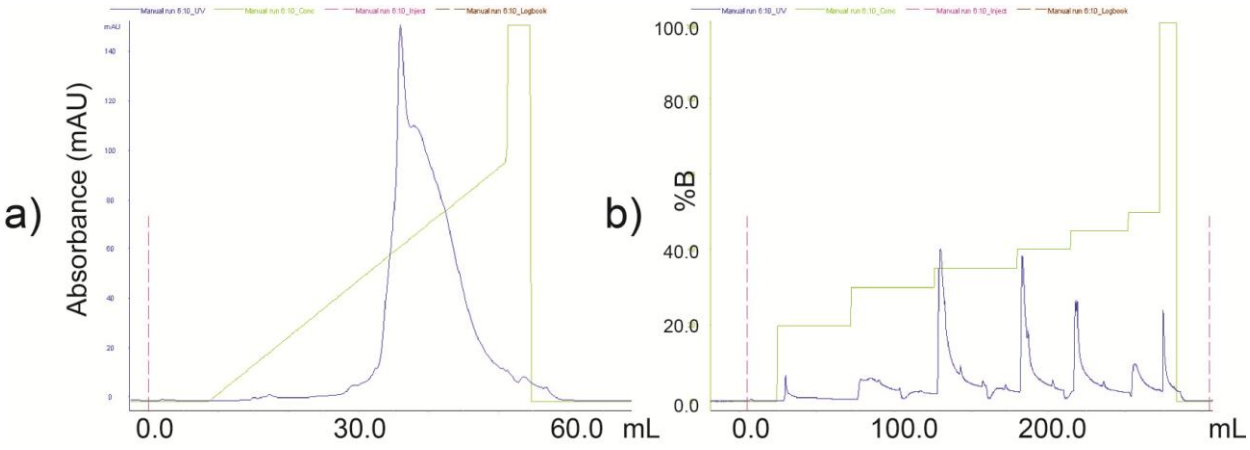


Figure 5.5: YcfD$_{RM}$ purified by Ni(II)-affinity chromatography followed by strong ion exchange monolithic column (CIM QA-1). a) An aliquot of YcfD$_{RM}$ (2 mg) was injected onto CIM QA-1 and eluted with a linear gradient 0-100% B over 40 CV at flow rate 10 mg/mL. b) Concentrated sample purified by linear gradient was re-injected and purified by stepped gradient. Elution of protein peaks was observed at 35% B; after elution of a major peak the column was washed with 35% B for 40 CV; concentration of B was then increased to 40%, resulting in elution of another peak of equally intensity; column was washed with 40 CV of 40% B, then the concentration was increased to 45% B, leading to elution of a third major peak.

In order to further purify ycfD$_{RM}$, a preparative Source^{TM} 15Q strong anion exchange column (GE Healthcare) was then used. After a Ni(II)-affinity chromatography step (HisTrap, GE Healthcare), ycfD$_{RM}$ was buffer exchanged into 50 mM HEPES pH 7.5, the sample was applied onto Source^{TM} 15Q. Elution with a linear gradient of NaCl (0 – 0.75 M in 25 mM Tris-HCl, pH 7.5 over 50 CV) resulted in partial separation of two peaks eluting at 25% B (0.188 M NaCl), labelled peaks A and B (Figure 5.6a).

Increasing the pH of buffer to 7.8 and 8.0, resulted in improved resolution of peaks (Figure 5.6b-c). Higher pH values were not tested because the buffer screen described in Chapter 4 demonstrated a sharp decrease in catalytic activity at pH >8.0.

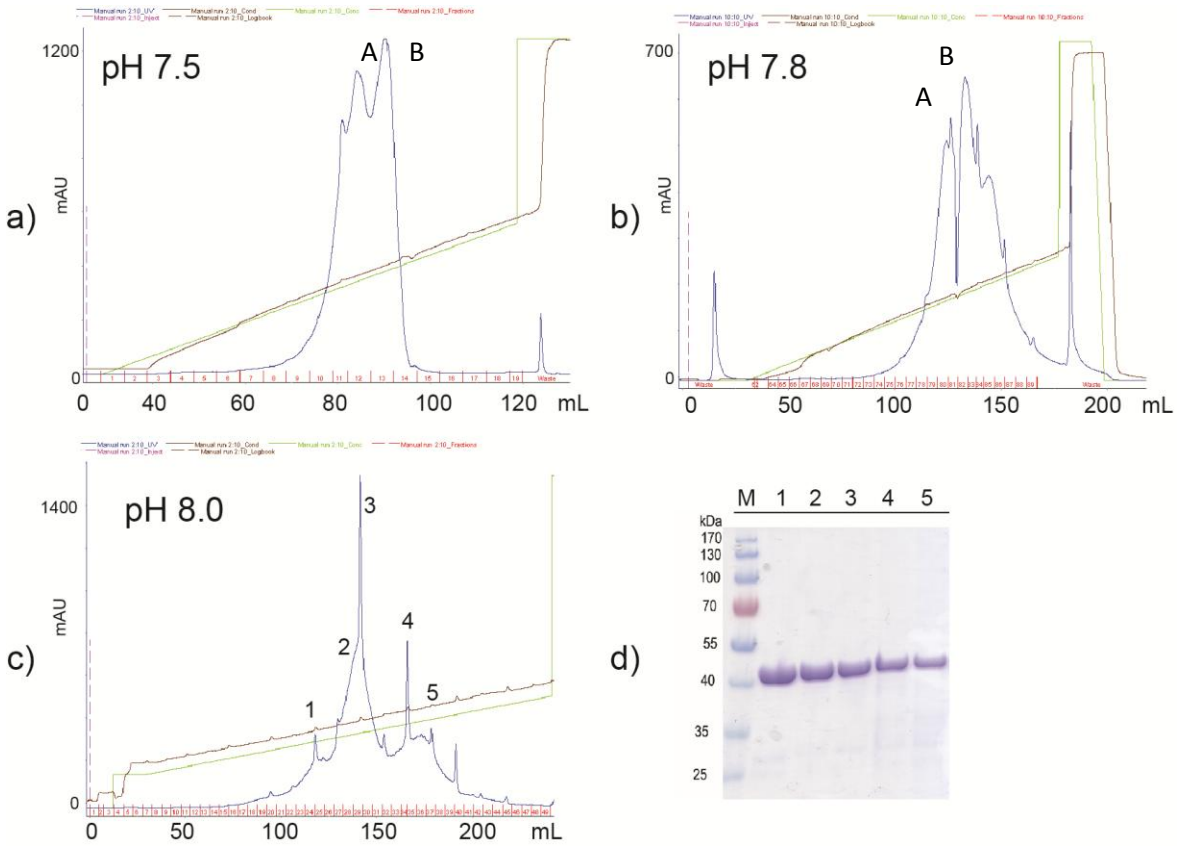


Figure 5.6: The effect of pH on the separation of sub-populations of ycfD$_{RM}$ by anion-exchange chromatography (Source^{TM} 15Q). The pH was increased from a) pH 7.5, b) 7.8 to c) 8.0. Separation of ycfD$_{RM}$ subpopulations was greatest at pH 8.0. Higher pH values were not tested due to decreased activity of ycfD$_{RM}$ observed with activity studies. D) SDS-PAGE gel demonstrates that peaks 1-5 observed in chromatogram c) correspond to species with different ionic properties of the same protein (>90% pure).

Fractions A and B from the Source^{TM} 15Q column were concentrated separately and were used separately for crystallisation trials. An aliquot of ycfD_{RM} (25 μL at 30 mg/mL) was mixed with $MnCl_2$ (5 μL at 10 mM in 50 mM HEPES pH 7.5, mixed with 5 μL of buffer), NOG (1 μL of 100 mM stock in 50 mM Tris-HCl, pH 7.5) and $L16_{RM}$ substrate fragment (20-mer, 2 μL of 50 mM stock mixed with 12 μL buffer). It was crucial to adjust the pH value of the peptide solution to ~7-8 to prevent precipitation upon mixing with the enzyme solution.

5.3.2. Crystallisation of ycfD_{RM}.Mn(II).NOG.L16$_{RM}$ complex

Initially, a screen was conducted using commercial crystallisation screens; JCSG, INDEX, MIDAS, PACT Premier, PEG ION, NATRIX and SALT-RX (obtained from Hampton Research and Molecular Dimensions) at 20°C. JCSG and INDEX screens were also set up and incubated at 4°C. After 6 days, conditions JCSG_E8 (0.1 M sodium acetate and 1.0 M di-ammonium hydrogen phosphate, pH 4.5) and JCSG_A11 (0.2 M ammonium dihydrogen phosphate, 0.1 M Tris pH 8.5, 50% MPD) yielded crystals for both fractions A and B at both 20°C and 4°C; PEG ION D8 (0.2 M ammonium phosphate dibasic, 20% PEG 3,350) resulted in a single crystal at 20°C. Crystals were harvested using nylon loops of appropriate size and screened in-house by Dr Michael McDonough. The crystals grown in 50% MPD were cryoprotected by transferring to well solution before cyro-cooling in liquid N_2 (the high MPD levels served as cryo-protectant), crystals from condition PEG ION_D8 were cryoprotected in 25% glycerol diluted in well solution. Crystals from fraction A grown in condition JCSG_A11 showed diffraction patterns characteristic of protein.

5.3.3. Optimisation of the JCSG_A11 crystallisation condition

Condition JCSG_A11 was optimised in a matrix approach. In a 96-well format, the pH/[$NH_4(H_2PO_4)$]/[MPD] matrix was varied systematically; the pH was increased between 7.5 – 9.0 in 0.5 unit increments; ammonium dihydrogen phosphate was varied from 0.1-0.4 M in 0.1 M increments; MPD was varied from 35-60% in 5% increments (Figure 5.7). Well A1 was used for the original condition, JCSG A11, as a positive control.

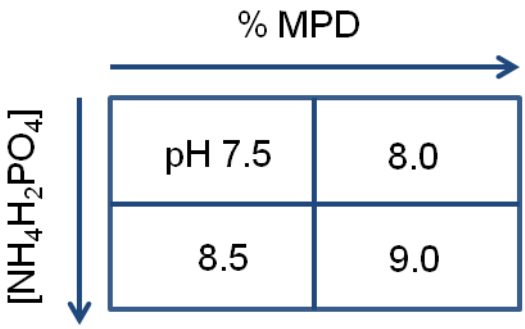


Figure 5.7: Optimisation plate layout for crystallisation of ycfD$_{RM}$.L16$_{RM}$.Mn(II).NOG complex.

After 7 days, crystals of ycfD$_{RM}$ from fraction A were observed in several wells (Figure 5.8). Rhombohedral rods (500 μm×30 μm) were observed in wells containing 0.1 M $NH_4H_2PO_4$ (lowest concentration tested) pH 8.5-9.0, and 50-55% MPD (Figure 5.8). YcfD$_{RM}$ from Fraction B yielded phase separation appearing as oily droplets. The results were reproducible using these conditions and were replicated with several batches of protein purified on separate days; ycfDRM from fraction A consistently generated large rhombohedral crystals; ycfD$_{RM}$ from fraction B consistently yielded oily droplets.

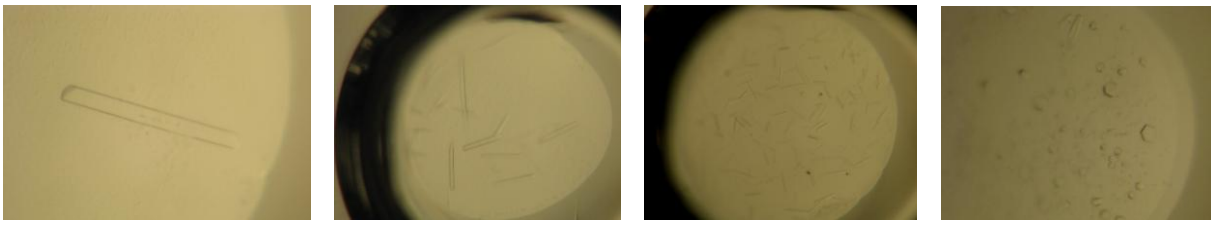


Figure 5.8: First round of optimisation of ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$ crystallisation condition. a) 0.1 M $NH_4(H_2PO_4)$, 0.1 M Tris pH 7.5, 45% MPD; b) 0.1 M $NH_4(H_2PO_4)$, 0.1 M Tris pH 8.5, 50% MPD; c) 0.1 M $NH_4(H_2PO_4)$, 0.1 M Tris pH 8.0, 45% MPD; d) 0.1 M $NH_4(H_2PO_4)$, 0.1 M Tris pH 9.0, 60% MPD.

A second round of optimisation using a fine grid approach was carried out at constant pH (pH 8.5) with varying MPD (40-50% 1 % increments) along the x-axis (12 conditions) and $NH_4H_2PO_4$ concentration (0-0.14 M in 0.02 M increments) along the y-axis (8 conditions). No improvement in crystal size/shape was observed. During in-house screening of crystals, a crystal from condition E3 (0.1 M $NH_4H_2PO_4$, 0.1 M Tris pH 8.5, 45% MPD) diffracted to 3.0 Å. This condition was therefore progressed into a third optimisation round, which included detergent and additive screens (Hampton Research). The additive screen resulted in 6 distinct crystal hits with unique morphologies ($BaCl_2$,

CsCl, hexammine Co(III)Cl$_3$, MgCl$_2$, spermine and TCEP). The crystals were harvested by transfer to well solution and flash-cooled in liquid N_2 followed by in-house screening, which indicated that they were salt crystals.

Fourteen detergent screen conditions yielded protein crystals. Several large crystals were observed and harvested as described previously (lauryl maltose neopentyl glycol (A2), 250×150 μm; anapoe®-80 (B2), 350×90 μm, n-tetradecyl-β-D-maltoside (B1), 200×200 μm; n-nonyl-β-D-thioglucoside (D3), 200×125 μm; hexaethylene glycol monooctyl ether (E3), 200×90 μm, FOS-MEA®-10 (H1), 175×100 μm), whilst others generated smaller crystals (50×50 μm), which were not screened. A selection of different crystal morphologies observed in the detergent screen is shown in Figure 5.9.

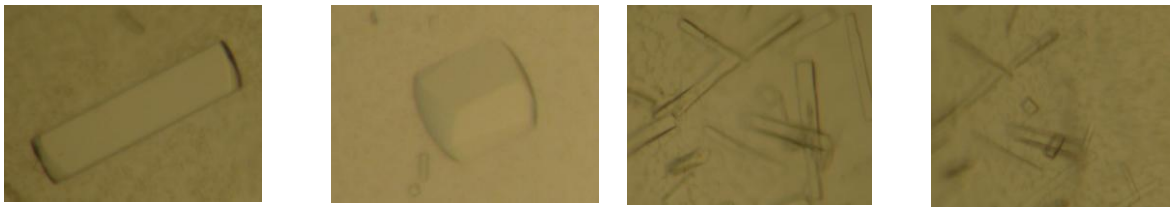


Figure 5.9: Detergent screening with ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$ of JCSG A11.1.1 (0.1 M $NH_4H_2PO_4$), 0.1 M Tris pH 8.5, 45% MPD). a) Anapoe®-80, b) n-tetradecyl-b-D-maltoside, c) Anapoe®-C12E10, d) NDSB-195.

Optimisation of the PEG IONscreen (well D8) condition was also conducted. Concentration of PEG was increased 0-30% in 2.5% steps along the x-axis (12 conditions) and ammonium phosphate dibasic was varied 0-0.35 M in 0.05 M increments along the y-axis (8 conditions). Although crystals were observed in several wells, in-house diffraction screening indicated diffraction to a lower resolution than that obtained previously using this condition (> 4.0 Å), therefore the condition was not pursued further.

5.3.4. Crystallisation of a ycfD$_{RM}$.Mn(II).FG2216 complex

Crystallization of ycfD$_{RM}$ was also attempted in the absence of the 20-mer L16$_{RM}$ fragment peptide. In order to facilitate crystallisation, a stabilising compound, FG2216 (IOX3, ΔT_m of 6°C) identified using the DSF assay (Table 5.1). YcfD$_{RM}$ (25 μL, 20 mg/mL) was mixed with $MnCl_2$ (5 μL of 10 mM, mixed with 5 μL of 50 mM HEPES, pH 7.5) and FG2216 (4 μL of 100 mM in 50 mM Tris-HCl, pH 7.5). The ycfD$_{RM}$ Mn(II) FG2216 mixture was screened using the INDEX screen (Hampton Research)

and incubated at 20°C. Two hits were observed within 1 week (Figure 5.10): condition C4 (35% Tacsimate, pH 7.0) and F3 (5% Tacsimate, pH 7.0, 0.1 M HEPES, pH 7.0, 10% PEG 5,000). Despite the relatively small size (150×50 μm), crystals from INDEX_C4 diffracted to sub-3Å resolution in-house.

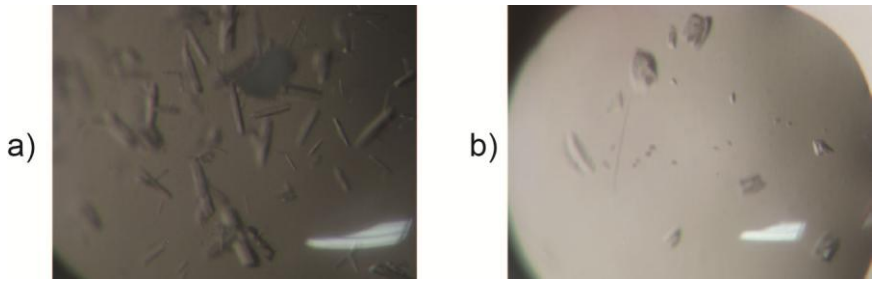


Figure 5.10: ycfD$_{RM}$.Mn(II).FG2216 crystal hits. A) INDEX screen (Hampton research) condition C4 (35% Tacsimate, pH 7.0); b) INDEX condition F3 (5% Tacsimate, pH 7.0, 0.1 M HEPES, pH 7.0, 10% PEG 5,000).

5.3.5. Data collection and processing

Crystals obtained from optimised crystallisation conditions (Table 5.3) were studied by single-crystal X-ray crystallography. Diffraction data were collected at 100K using synchrotron radiation at the Diamond Light Source (DLS). The data were processed as outlined in the Appendix. All processing was carried out by Dr. Michael A. McDonough.

Table 5.3: Crystallisation conditions for ycfD$_{RM}$ complexes used for structural determination

Protein complex	Sample	Additive	Crystallisation condition
ycfD$_{RM}$.Mn(II).FG2216	15 mg/mL ycfD$_{RM}$ 1 mM $MnCl_2$ 2 mM FG2216	50% (v/v) Tacsimate pH 7.0	Sitting drop (200 nl), protein-to-well ratio 1:1, 293K
ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$	15 mg/mL ycfD$_{RM}$ 1 mM $MnCl_2$ 1 mM NOG 2 mM L16$_{RM}$	0.1 M Tris.HCl pH 8.5 0.1M ammonium dihydrogen phosphate 45% (v/v) methyl-2,4-pentanediol	Sitting drop (200 nl), protein-to-well ratio 1:1 293K

Structures of ycfD$_{RM}$.Mn(II).FG2216 and ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$ were solved by molecular replacement using the ycfD$_{EC}$ structure as the search model (unpublished data, Schofield Lab). Structural refinement was carried out with PHENIX(4) with iterative rebuilding of the models using COOT(5) until R_{work}/R_{free} converged to final values 0.152/0.212 and 0.184/0.221 for ycfD$_{RM}$.Mn(II).FG2216 and ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$, respectively. Refinement statistics are detailed in the Appendix.

5.4. Crystal structure of ycfD$_{RM}$

-the first crystal structure of a ROX substrate complex -

5.4.1. Overall fold

To investigate the molecular details of how ycfD$_{RM}$ interacts with L16$_{RM}$, structures of both apo-ycfD$_{RM}$ (Space Group $C2_1$) and ycfD$_{RM}$ in complex with a substrate fragment (L16$_{RM}^{72-91}$), NOG, and Mn(II) (substituting for the catalytic Fe(II)) (ycfD$_{RM}$.Mn.NOG.L16$_{RM}$, Space Group $P2_12_12_1$) were solved to 2.8 Å and 3.0 Å, respectively.

The crystal structures of ycfD$_{RM}$ revealed that, like ycfD$_{EC}$, it forms a homodimer (Figure 5.11). The overall fold of ycfD$_{RM}$ highly resembles the fold recently reported for other ROX, including human NO66 (6) and MINA53 (Schofield Lab, unpublished data) (Figure 5.11).

YcfD$_{RM}$ consists of a JmjC-like domain (residues 1-212) containing a core double-stranded beta-helix fold (DSBH) (β-strands I-VIII) made up of two β-sheets, the major and the minor β-sheet, characteristic of 2OG oxygenases. The 2OG and Fe(II) binding pocket is sandwiched between the major and minor β-sheets. The JmjC is followed by a dimerisation domain (residues 214-280) comprising three α-helices that interlink with the other protomer in the dimer, and a *C*-terminal domain (residues 296-392) of unknown function (Figure 5.12).

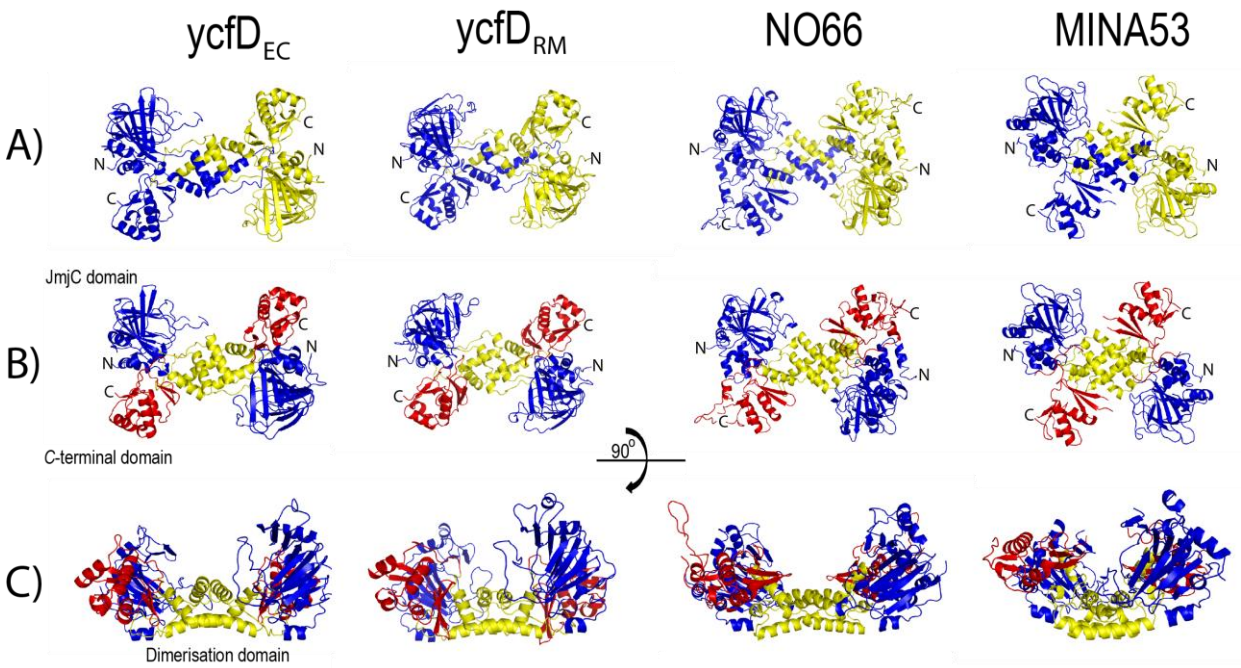


Figure 5.11: Comparison of tertiary structures of bacterial and human ROX. A) A top-down view on the tertiary structure of ycfD$_{EC}$, ycfD$_{RM}$, NO66 and MINA53. Two monomers are coloured yellow and blue. *N*- and *C*- termini are indicated; B) domain architecture of ROX is comprised of a JmjC (blue), dimerisation (yellow) and *C*-terminal (red) domains which have been conserved throughout evolution. C) A 'side-on' view of the dimerisation interface of the different ROX structures reveals subtle differences in the arrangement between monomeric units as well as location and length of loop extensions some of which protrude from the JmjC and *C*-terminal domains.

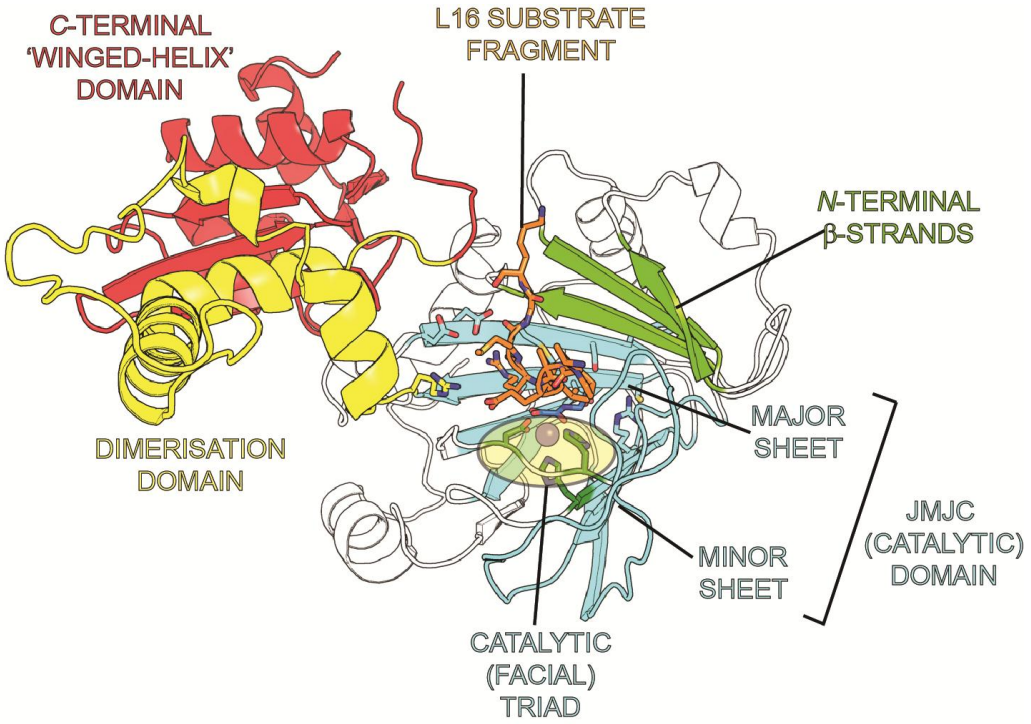


Figure 5.12: The tertiary structure of ycfD$_{RM}$. a) YcfD$_{RM}$ is a homodimer of two units comprising a JmjC domain (blue), dimerisation domain (yellow) and a *C*-terminal domain. A schematic representation of ycfD$_{RM}$ tertiary structure showing the key structural features: jmjC domain (cyan) contains the facial triad comprising H133, D135 and H196 (dark green), which octahedrally coordinate the active site metal (purple), NOG (dark blue) and the substrate chain (orange). The *N*-terminal β-strands (light green) provide the 'lip' facilitating the binding of substrate to the active site. Dimerisation domain (yellow) participates in binding of L16$_{RM}$, as well as dimerisation with another molecule of ycfD$_{RM}$. The function of C-terminal domain (red) is unknown. Figure was created using PyMol software (Schrodinger).

5.4.2. Topology analysis of ycfD$_{RM}$

A topological analysis of the ycfD$_{RM}$ structure indicated minor differences to ycfD$_{EC}$ in the secondary structure of the *N*-terminal extension (Figure 5.13), namely an α-helix present in ycfD$_{RM}$ (residues 74-79) is absent in ycfD$_{EC}$, whereas an α-helix present in the *C*-terminus of ycfD$_{EC}$ (residues 343-353) is absent in ycfD$_{RM}$. These differences are unlikely to explain the differences in biochemical properties observed between ycfD$_{EC}$/ycfD$_{RM}$.

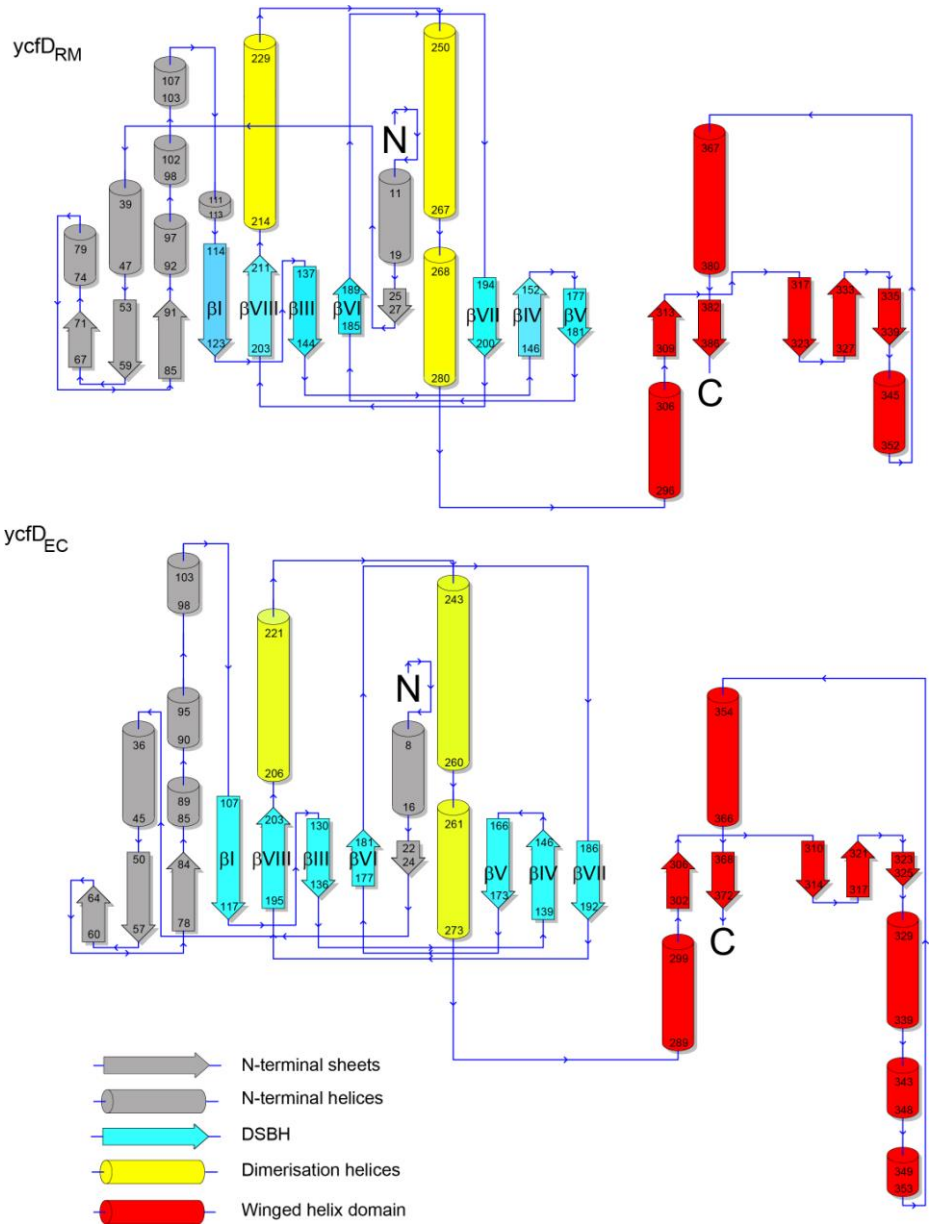


Figure 5.13: Topology diagrams of ycfD$_{RM}$ and ycfD$_{EC}$. Topology diagrams were generated using PDBsum software available on the EMBL website (https://www.ebi.ac.uk/pdbsum). Note that strand βII is not fully formed in neither ycfD$_{RM}$ nor ycfD$_{EC}$.

5.4.3. Salt bridge analysis

The PISA online server (http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) was used to analyse the ycfD$_{RM}$ structure for salt-bridge interactions for comparison with ycfD$_{EC}$. This analysis identified 25 intermolecular salt-bridge pairs formed between two protomers of the ycfD$_{RM}$ homodimer. In contrast, only 13 salt-bridge pairs were identified in the ycfD$_{EC}$ homodimer (see Appendix for a full list). YcfD$_{RM}$ therefore forms twice as many salt-bridge interactions as ycfD$_{EC}$. This is consistent with previous reports that identified increased salt-bridge content as a dominant contributor towards thermostability (7,8). The number of hydrogen bonds formed between the two protomers is 35 and 34 for ycfD$_{RM}$ and ycfD$_{EC}$, respectively (see Appendix for a full list). Hydrogen bonds are therefore less likely to contribute importantly towards the differential thermostability of ycfD$_{RM}$ as compared to ycfD$_{EC}$. Furthermore, the analysis of salt-bridge content identified 23 intramolecular salt-bridge interactions within a ycfD$_{RM}$ monomer (see Appendix for a full list); salt-bridges were found to span predominantly between the domains of the protein, either between the JmjC and the dimerisation domain or the dimerisation domain and the *C*-terminal domain (Figure 5.14). The highest number of salt-bridge interactions was found at the interface between the jmjC- and dimerisation domains (Figure 5.14).

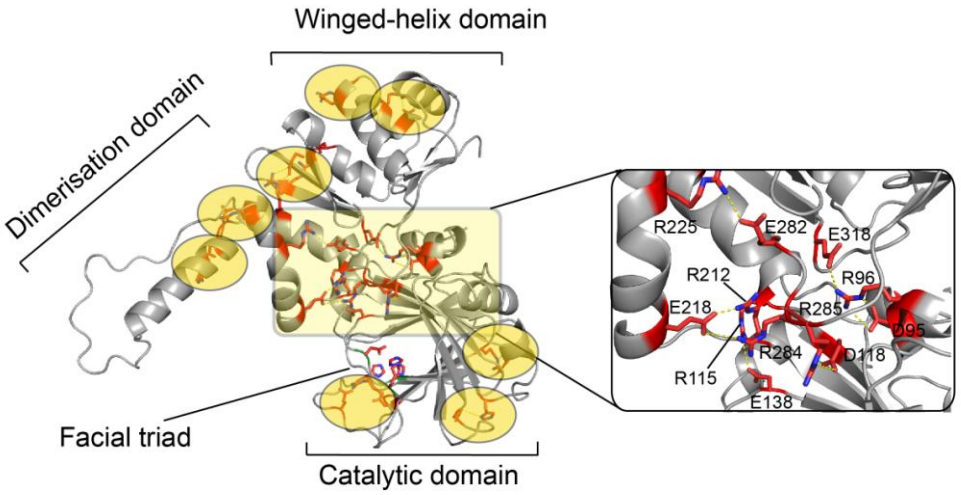


Figure 5.14: Intramolecular salt-bridge interactions within a ycfD$_{RM}$ monomer. Residues involved in salt-bridge interactions are coloured red. 23 intramolecular salt-bridge interactions are observed within a ycfD$_{RM}$ monomer; salt-bridges span between domains of the protein, either between the JmjC and dimerisation domains, or the dimerisation domain and the *C*-terminal domain. Inset: the highest density of salt-bridge interactions was observed at the interface between the JmjC and the dimerisation domain.

5.4.4. Cation-π interactions

A number of cation-π interactions were identified in ycfD$_{RM}$: R115 forms an apparent cation-π interaction with W114, R225 with W229 and R24 with Y20. Interestingly, upon substrate binding the R115-W114 interaction remains, whereas the others are apparently lost; new cation-π interactions form between R148-W150 and R212-Y137. As will be discussed later, Y137 forms a cation-π interaction with the side chain of L16$_{RM}$ R82, thus forming an unusual cation-π-cation sandwich complex (*vide infra*).

5.4.5. Hydrophobic interactions

An analysis of hydrophobic interactions (defined by proximity of hydrophobic residues within 5 Å; performed using the 'Protein Interactions Calculator available on http://pic.mbu.iisc.ernet.in/job.html) within a monomer of ycfD$_{RM}$ revealed 401 hydrophobic contacts (complete list is detailed in the Appendix). In ycfD$_{EC}$, 305 hydrophobic interactions were identified. YcfD$_{RM}$ therefore contains 30% more hydrophobic interactions than ycfD$_{EC}$, indicating that hydrophobic interactions contribute to the increased thermal stability of ycfD$_{RM}$ compared to ycfD$_{EC}$.

5.5. The DSBH and substrate binding of ycfD$_{RM}$

Electron density for nine of the 20 amino acid residues of the L16 peptide (K77-K85) was observed in the ycfD$_{RM}$.Mn.NOG.L16 structure (Figure 5.15). The interaction of the substrate with individual strands of the DSBH were analysed according to the methodology described in Hangansky (9), allowing the comparison with other studied 2OG oxygenases (Figure 5.12).

The L16$_{RM}$ fragment was found to form a large number of interactions with ycfD$_{RM}$ side-chains, in particular from strands βI and βVIII. L16$_{RM}$ R82 is anchored in the active site with its pro-*R* hydrogen projecting towards the metal (metal-Cβ 4.7 Å). It occupies the substrate binding pocket formed by Y137$_{RM}$ and M120$_{RM}$ (Y129$_{EC}$ and M112$_{EC}$ in *E. coli* ycfD), apparently including cation-π and hydrophobic interactions.

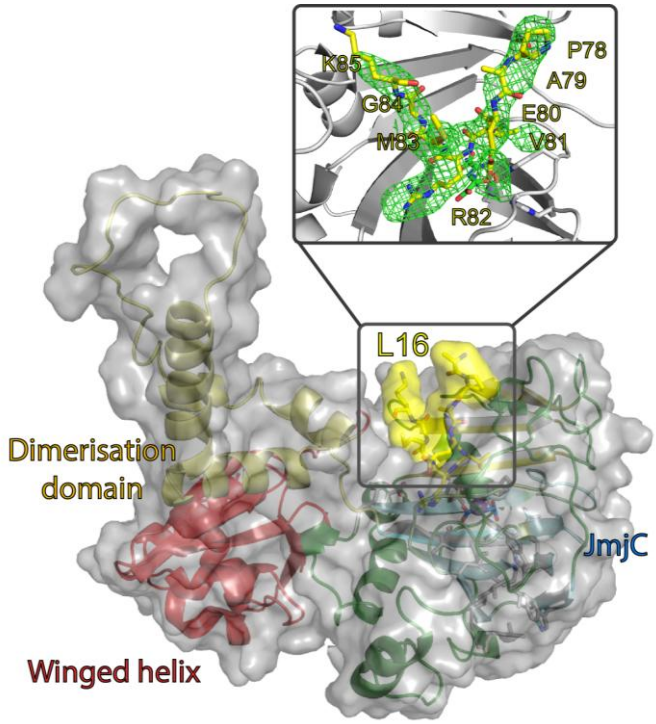


Figure 5.15: A 2Fo-Fc OMIT map the L16$_{RM}$ peptide fragment observed crystallographically in the active site of ycfD$_{RM}$. The map was contoured at a level of 1σ. Three domains of ycfD$_{RM}$ (JmjC, dimerisation domain and the *C*-terminal 'winged helix' domain) are coloured blue, yellow and red, respectively. The L16$_{RM}$ fragment is observed to protrude into the JmjC active site. The inset shows the electron density of the fragment.

5.5.1. DSBH strand βI

YcfD$_{RM}$ strand βI is one of four strands forming the major β-sheet. βI contains four functionally important residues, D117, D118, M120 and S122 (Figure 5.16). The side chain of D117 forms a hydrogen bond (3.2 Å) with the side chain hydroxyl of Y137 upon substrate binding, providing an additional interaction for positioning the substrate in the active site (Figure 5.17). Additionally, it forms a hydrogen bonding interaction with R285 from the *C*-terminal domain, which is also involved in substrate binding. The side chain of residue D118 (D110 in ycfD$_{EC}$) forms a salt-bridge interaction with the guanidinium side chain of substrate R82 (O-NH 2.8 Å). Furthermore, it is part of a hydrogen bonding network spanning four residues: it forms a hydrogen-bonding interaction with the side chain of Q91 from β-1 (3.6Å), which hydrogen bonds to the side chain of R55 (β-2, 3.2 Å), which in turn binds to the backbone carbonyl oxygen of V81 of substrate L16$_{RM}$. Residue M119 lines one side of hydrophobic pocket to accommodate R82. Finally, S121 forms an apparent hydrogen bonding interaction with NOG C5-carboxylate (2.8 Å), potentially providing an explanation for the unusual twist in C-COO$^-$ bond angle which 'pulls' the NOG C5 carboxylate away from R148.

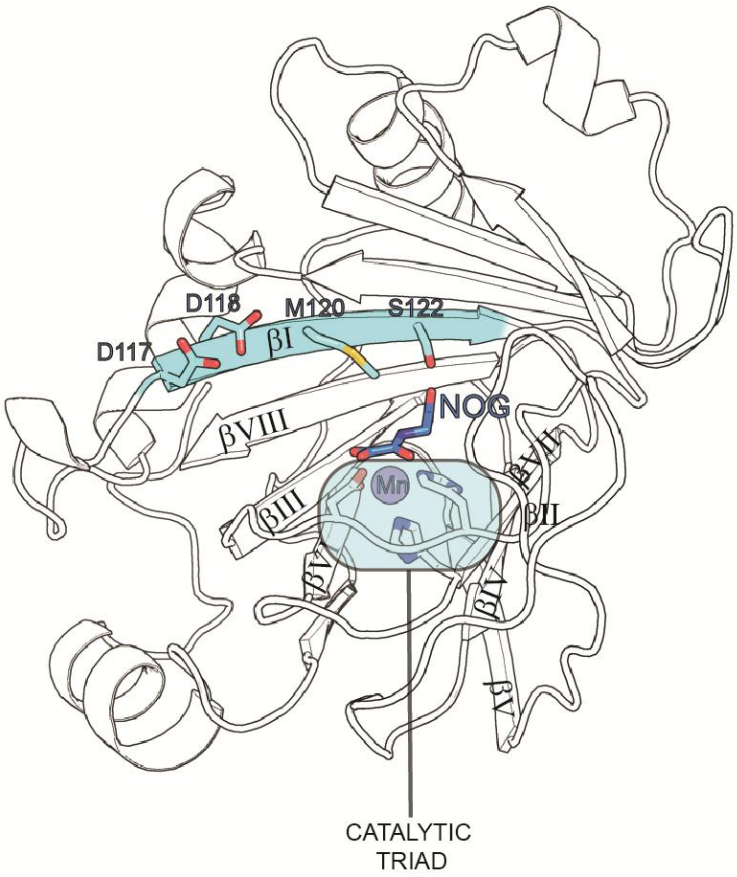


Figure 5.16: A cartoon representation of the ycfD$_{RM}$ DSBH with the strand βI coloured blue. Residues D117, D118, M120 and S122 are shown in stick representations. D117 forms a hydrogen bond with Y137 upon substrate binding (3.2Å), as well as a hydrogen bonding interaction with R285 from the *C*-terminal domain, which is also involved in substrate-binding. Residue D118 forms a salt-bridge interaction with the guanidinium side-chain of the L16$_{RM}$ substrate R82 (O-NH 2.8 Å). Residue M119 lines one side of hydrophobic pocket to accommodate R82. S121 forms a hydrogen bonding interaction with the NOG C5-carboxylate (2.8A).

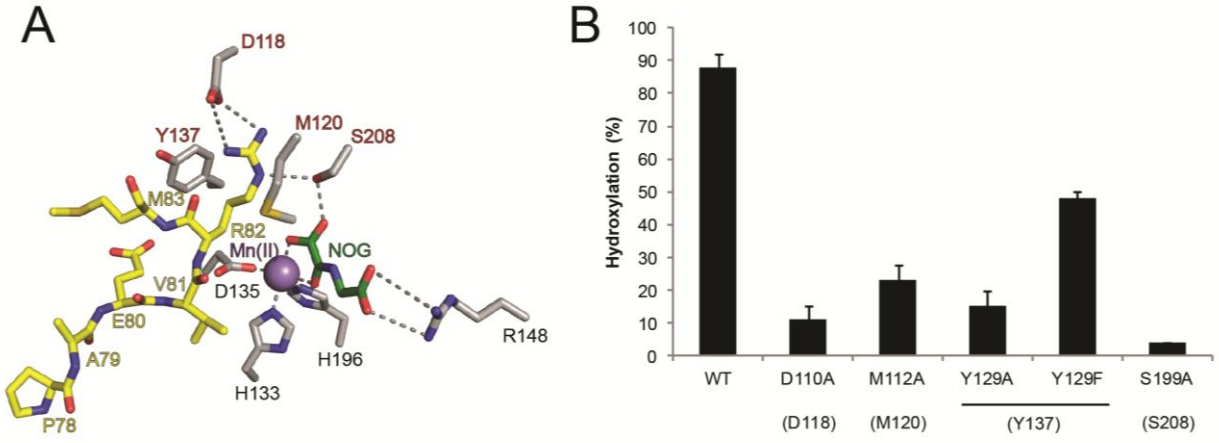


Figure 5.17: A view of the L16$_{RM}$ substrate fragment binding to ycfD$_{RM}$ active site and MALDI-MS analysis of activity of ycfD$_{EC}$ variants. A) L16$_{RM}$ (yellow) is observed to make several interactions with the side-chains of ycfD$_{RM}$ active site residues (blue). NOG (green), Mn(II): purple. B) YcfD$_{EC}$ variants (1 μM) were incubated with Fe^{2+}, 2OG and ascorbate (100, 200 and 1000 μM, respectively) in the presence of L16$_{EC}$ substrate fragment (100 μM). The reaction was incubated for 30 min at 37°C before quenching with CF_3COOH_{aq} (1%). The reactions were conducted in triplicate, with error bars denoting s.d. from the mean.

In order to study the importance of the observed substrate binding residues, and to confirm the generality of results for ycfD enzymes, variants of ycfD$_{EC}$ at selected residues were prepared by site-directed mutagenesis. The ycfD$_{EC}$ D110A variant, corresponding to ycfD$_{RM}$ residue D118, retained <10% activity (Figure 5.17B). Therefore, the structurally observed salt-bridge interaction between D118 and L16 R82 is important for catalytic activity and supports the validity of the ycfD$_{RM}$.Mn.NOG.L16$_{RM}$ crystal structure representing a catalytically relevant complex. Interestingly, D110E ycfD$_{EC}$ retained 30% activity relative to WT (data not shown), indicating that the observed distance between the L16$_{RM}$ R82 guanidinium and the ycfD$_{RM}$ D118 carboxylate (2.8 Å) is optimal and extending the carbon chain of the carboxylic acid backbone by one extra carbon atom contributes to lowering enzymatic activity. An attempt was made to reverse the charge of the substrate-enzyme interaction, i.e. swap the L16 R82 residue with a carboxylic acid (Asp/Glu), concomitant with a replacement of the enzyme based acid with a basic residue (Arg/Lys). A DSF measurement of T_m revealed that D110R ycfD$_{EC}$ variant was significantly de-stabilised relative to WT (T_m of 32^{o}C compared to 48°C for WT, data not shown), therefore the experiment was terminated since it would be difficult to distinguish the destabilising effects from change in activity.

The ycfD$_{EC}$ Y129A variant, corresponding to ycfD$_{RM}$ Y137, which was observed to interact closely with residues from β-strand βI, showed a marked decrease in activity (15% hydroxylation observed, Figure 5.17B), suggesting that the aromatic ring of Y129$_{EC}$/Y137$_{RM}$ is important for catalytic activity. Variant Y129F retained 50% activity, supporting the role of Y129$_{EC}$/Y137$_{RM}$ in cation-π interaction, but also suggesting a mechanistic role for the tyrosine hydroxyl group.

5.5.2. DSBH strands βII-βVIII

Strand βII in ycfD$_{RM}$ (and ycfD$_{EC}$) is not fully formed. It harbours the residue V129, which lines the hydrophobic pocket providing hydrophobic interaction for accomodating substrate residue V81 from L16$_{RM}$. The strand βIII contains two hydrophobic residues, V138, L140, which are part of a hydrophobic pocket at the narrow end of the DSBH barrel. The side chain of Q142 may form a hydrogen bond with the side chain of T205 (strand βVIII); T205 is in turn involved with binding

NOG. The β-strand βIV is important for 2OG binding; the side chain of residue R148 forms a salt-bridge interaction with the C5-carboxylate of NOG (Nε – NOG O4 3.7 Å; NH2 – NOG O3 4.3 Å, Figure 5.19). Furthermore, the side chain of R148 also forms a hydrogen bond with the hydroxyl group of T206 (strand βVIII, Nε–OH 3.2 Å). 2OG/NOG is therefore involved in a salt bridge/hydrogen bonding network involving R148-T206-NOG-S122 (see below). The βV strand forms the minor sheet of the DSBH barrel. It contains hydrophobic residues I152, A178, W180 and L182, which line the narrow end of the DSBH. βVI provides further two residues for the hydrophobic lining of the barrel: L188, L190. βVII harbours the facial triad distal histidine H196 (H_{n+m}). Moreover, βVII residues G198 and A200 are a part of the hydrophobic core.

Finally, βVIII (Figure 5.18) contains C203, the function of which is unclear; this residue is not conserved in ycfD$_{EC}$. T206 is involved in an interaction with the NOG/2OG C5 carboxylate (discussed below); S208 (S199 in ycfD$_{EC}$) is positioned within 2.7 Å from the NOG oxalyl carboxylate and 3.5 Å from the substrate arginine Nε, indicating a potential (weak) hydrogen bonding interaction with the side chain of substrate R82 (Figure 5.19).

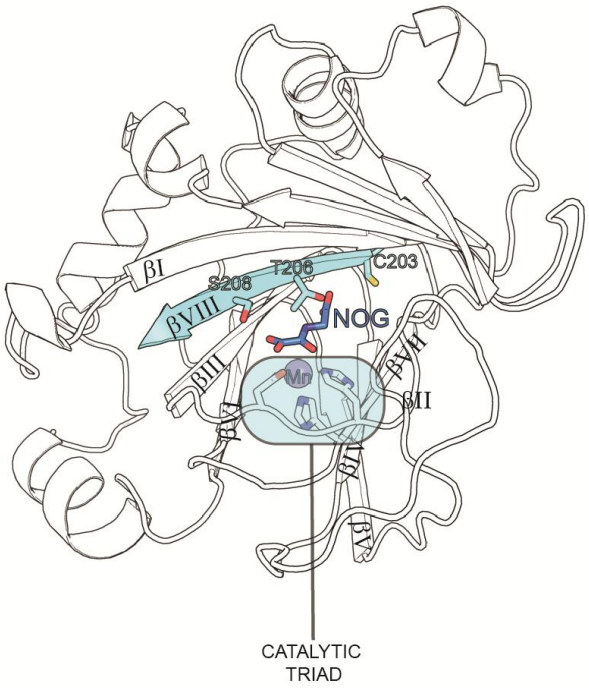


Figure 5.18: A cartoon representation of the ycfD$_{RM}$ DSBH strand βVIII containing residues C203, T206 and S208. The function of C203 is unclear; the residue is not conserved in ycfD$_{EC}$; T206 is involved in an interaction with NOG/2OG C5 carboxylate (discussed below); S208 is positioned within 2.7 Å from the NOG oxalyl carboxylate and 3.5 Å from substrate R82 Nε.

Consistent with the proposed role of S208 ycfD$_{RM}$ in binding 2OG and the substrate R82, the S199A ycfD$_{EC}$ variant had <5% activity compared to WT as analysed by MALDI-MS (Figure 5.17B), suggesting its catalytic importance for interacting with NOG carboxylate and potentially interacting with R82$_{L16}$. This residue may promote coupled turnover of L16 and 2OG (Figure 5.19). According to the classification of Hangansky *et al*., S208 is the residue B2-6 (minor sheet no. 2, residue 6) of the DSBH. According to their analysis, the B2-6 residue forms a close contact with with C-1 carboxylate of 2OG/NOG in every studied 2OG oxygenase structure (9). The position is usually occupied by a hydrogen bond donor (S/T/N) or a basic residue (R). In NO66/MINA53, the B2-6 residue is T419/T255; in FIH it is N294. Mutation of N294A in FIH resulted in a 25-fold decrease in k_{cat}/K_m for 2OG (10).

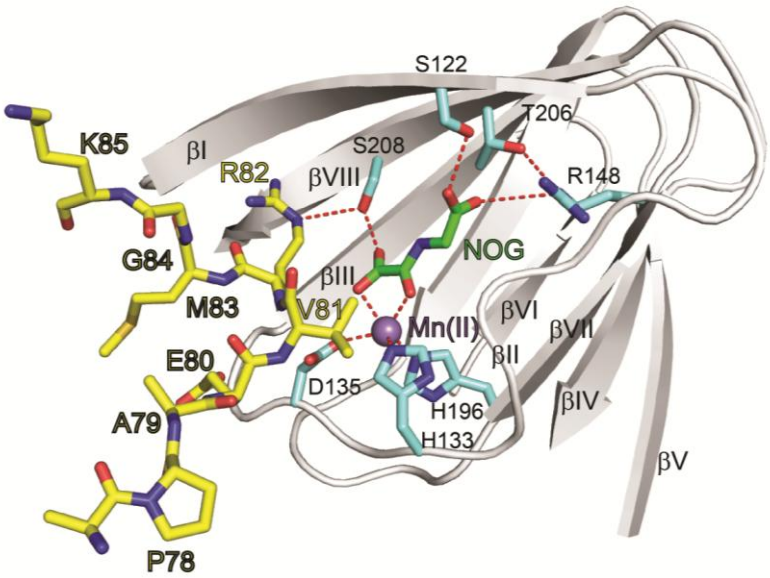


Figure 5.19: A view of the ycfD$_{RM}$ DSBH with L16$_{RM}$ substrate fragment and NOG bound. The C5 carboxylate of NOG forms an electrostatic interaction with R148 and a hydrogen bonding interaction with S122. The NOG oxalyl carboxylate forms a hydrogen bond with S208, which in turn forms a hydrogen bond with the Nε of the substrate R82.

Interactions of L16$_{RM}$ with other ycfD$_{RM}$ residues were identified by LigPlot analysis (Figure 5.20). The side chain guanidinium group of ycfD$_{RM}$ residue R55 forms a hydrogen-bonding interaction with the backbone carbonyl oxygen of V81 L16$_{RM}$; the side chain amide group of Q91 ycfD$_{RM}$ forms a hydrogen bond with the backbone carbonyl oxygen of M83 L16$_{RM}$, which also forms potential hydrogen bonding interactions with the guanidinium group of R285 ycfD$_{RM}$. This interaction is notable because R285 protrudes into the active site of the JmjC domain from the *C*-terminal domain, the function of which is unclear (see below).

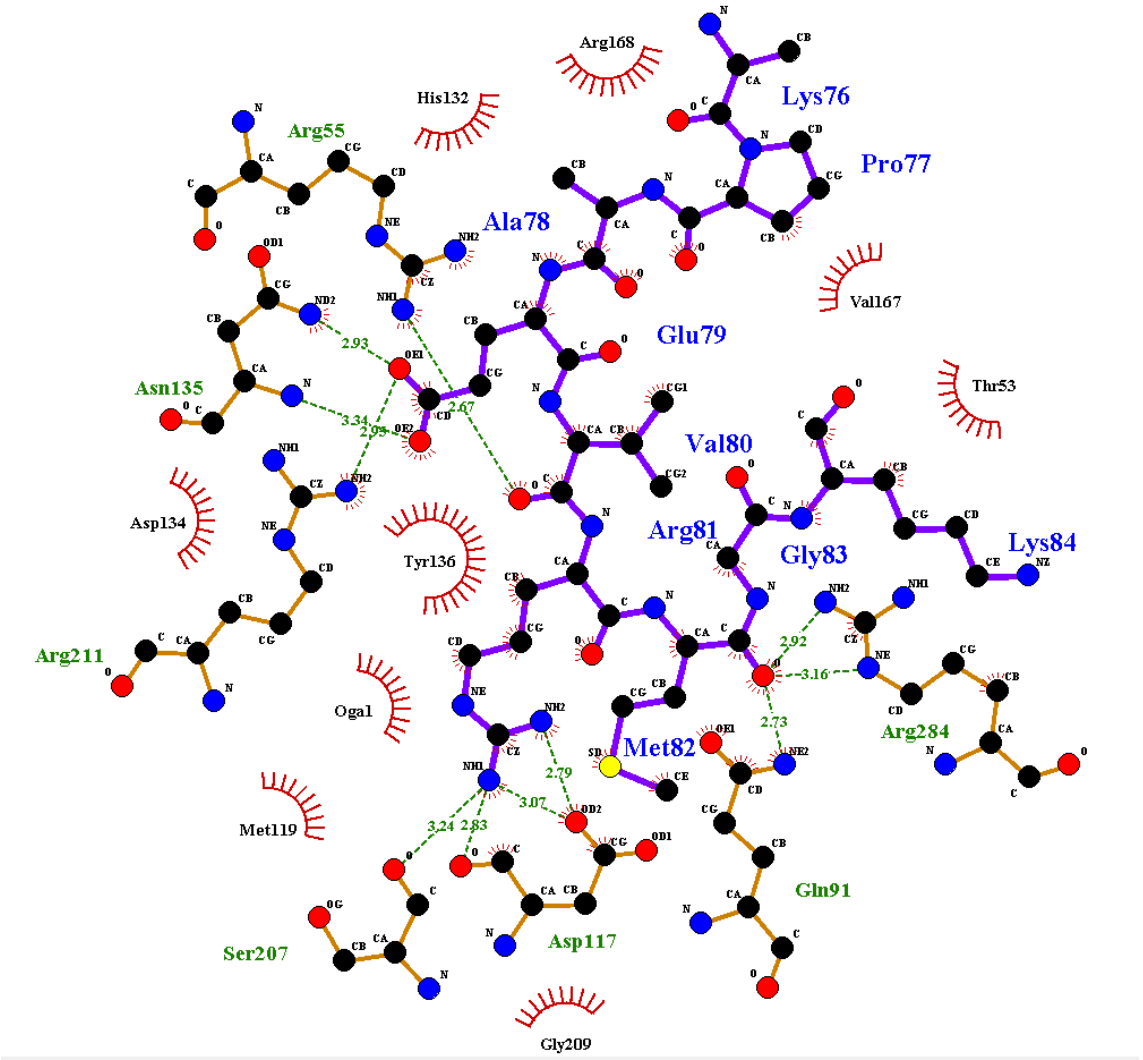


Figure 5.20: LigPlot analysis of ycfD$_{RM}$-L16$_{RM}$ interactions. The L16$_{RM}$ chain is coloured purple. Electrostatic interactions are indicated with green dashes. Hydrophobic interactions are indicated with red semi-circles with 'rays'. The analysis was performed using 'LigPlot' software (https://www.ebi.ac.uk/thornton-srv/software/LIGPLOT).

Overall, it can be concluded that the major sheet strands βI and βVIII harbour residues that are crucial for substrate recognition, whereas other strands form the hydrophobic core. The minor sheet strands likewise contain residues that form the hydrophobic core of the DSBH, as well as the catalytically important residues R148 (βIV) and H195 (βVII). Since the bottom of the is narrower than the top, the DSBH could be represented as a cup into which the substrate is 'dipped' akin to a coffee-cup (with coffee replaced by 2OG and Fe^{2+}) into which a 'biscuit' (representing L16) is dipped in order to change its chemical composition (Figure 5.21).

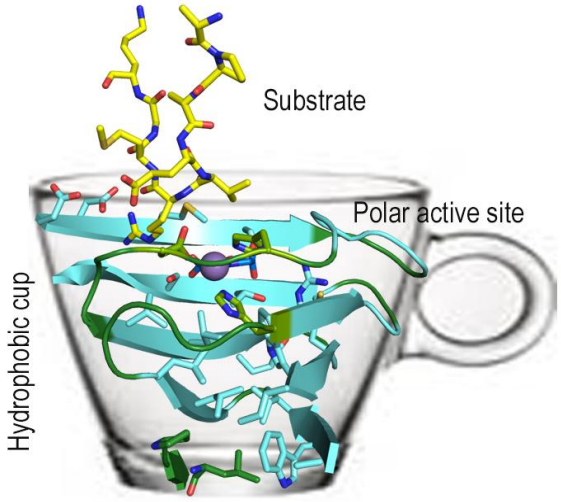


Figure 5.21: Due to the compressed hydrophobic bottom side of the jelly roll, the JmjC-domain of ycfD$_{RM}$ is cup-like. The catalytically important residues are located on strands βI and βVIII only, further strengthening the metaphor of a coffee-cup which holds an active ingredient that modifies an incoming substrate.

5.6. Comparison of VioC and ycfD$_{RM}$ substrate complex structures

Elucidation of the binding mode for the L16$_{RM}$ arginyl residue required to produce the 2*S*,3*R*-hydroxyarginyl diastereomer allows structural comparison with the structure of VioC (*Streptomyces vinaceus*), a distantly related arginine hydroxylase (11). VioC is a 2OG-dependent hydroxylase from the CAS family of 2OG- and Fe(II)-dependent dioxygenases, containing a HxE…H facial triad (12). VioC catalyses *L*-arginine hydroxylation to produce the (2*S*, 3*S*)-hydroxyarginine, which is a precursor to capreomycidine, itself a precursor in the biosynthesis of viomycin (11) (Figure 5.22b).

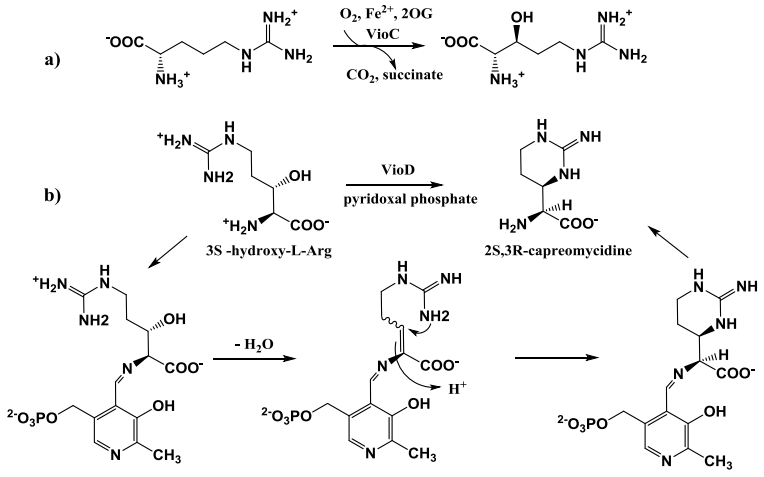


Figure 5.22: VioC and VioD-catalysed synthesis of 2*S*,3*R*-capreomycidine. a) VioC catalyses arginine hydryoxylation to 2*S*,3*S*-hydroxy-L-Arginine; b) VioD catalyses the cyclisation of 2*S*,3*S*-hydroxy-L-Arginine to 2*S*,3*R*-capreomycidine.

Similar to ycfD$_{RM}$ (and most other 2OG oxygenases), the structure of VioC comprises of a major (5 β-strands) and a minor sheet (3 β-strands). It contains an extensive insert between strands βIV and βV (βIV-βV insert) of 97 residues, including 3 α-helices (α5-α7) (13). The βIV-βV insert in ycfD$_{RM}$ comprises 24 residues.

Substrate structures of VioC contain tartrate (replacing 2OG) and Fe(II). Similar to the L16$_{RM}$ substrate structure complex in which the L16$_{RM}$ R82 side chain conformation is not favourable, the substrate arginine in the VioC substrate complex structure exists in a strained conformation. Similarly to ycfD$_{RM}$, the Cβ of L-arginine in the VioC active site is 4.2 Å from Fe(II). In contrast to L16$_{RM}$ R82, which is observed to project its pro-*R* hydrogen towards the active site metal, L-arginine is found to project its pro-*S* hydrogen towards the catalytic iron in the VioC substrate complex. The different diastereoselectivity achieved during catalysis occurs by swapping the locations of arginine-binding residues on the core DSBH scaffold, relative to the facial triad (Figure 5.23). In ycfD$_{RM}$ the guanidinium side chain of R82 forms a salt-bridge with D118 from strand βI. In contrast, the guanidinium group of L-arginine in VioC forms two hydrogen-bonding interactions with the carboxylates of D268 and D270, diagonally opposite strand βI (Figure 5.23). The substrate L-arginine carboxylate forms a salt bridge with VioC R334 side chain and a hydrogen bond to the backbone amide nitrogen of S158. There is no evidence of a cation-π interaction with substrate, as was observed in the ycfD$_{RM}$ substrate structure complex. The crystal structures of ycfD$_{RM}$ and VioC therefore demonstrate how Nature can utilise the DSBH scaffold to perform chemistry to achieve stereoselective hydroxylation producing several diastereomers of the arginine/arginyl moiety.

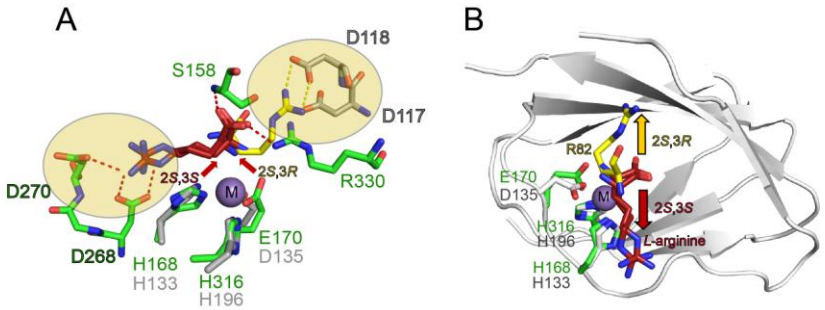


Figure 5.23: Superposition of ycfD$_{RM}$ and VioC substrate structures. A) YcfD$_{RM}$ is coloured grey with R82 in yellow. VioC (PDB: 2WBO) is coloured green, with L-arginine (found in two conformations) in red. Residue numbering for VioC is shown in green, with grey symbols denoting ycfD$_{RM}$ residues. B) The facial triad residues of ycfD$_{RM}$ and VioC were aligned to show substrate orientation relative to the DSBH (only ycfD$_{RM}$ DSBH is shown for clarity).

5.7. *N*-oxalyl glycine (NOG) binding

The Mn^{2+} ion in the active site of ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$ is octahedrally co-ordinated by the 2-His-1-carboxylate facial triad (HxD…H) comprised of H133 and D135 from βII and H196 from βVII of the DSBH. NOG occupies two co-ordination sites of the metal; the NOG carbonyl binds trans to D135 and the oxalyl-carboxylate binds trans to the proximal histidine (H133) (Figure 5.24), commonly found in JmjC-containing 2OG oxygenases (9). Additionally, the NOG C-5 carboxylate forms a salt bridge with the side chain of ycfDRM R148 located on βIV of the DSBH. This location of a basic 2OG-binding residue is characteristic of the JmjC domain oxygenases/demethylases, but differs from PHD2-like oxygenases in which the 2OG-binding residue (most commonly arginine) is located on either βVII or βVIII (14). As discussed above, the NOG C5-carboxylate forms an apparent hydrogen bond with the side chain hydroxyl of S121 (2.8 Å), whilst the NOG oxalyl carboxylate is positioned to form a hydrogen-bonding interaction (2.7 Å) with the side-chain hydroxyl of S208.

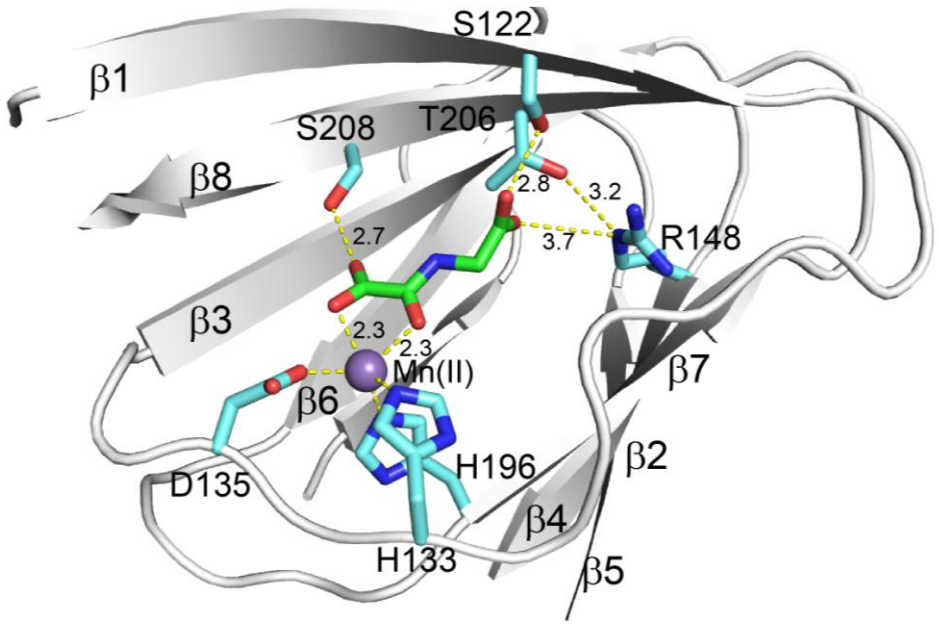


Figure 5.24: NOG interactions with DSBH residues. NOG occupies two co-ordination sites of the Mn^{2+} ion, which is octahedrally co-ordinated by the 2-His-1-carboxylate facial triad (HxD…H) comprised of H133 and D135 from βII and H196 from βVII of the DSBH. The NOG carbonyl binds *trans* to D135 and the oxalyl-carboxylate binds *trans* to the proximal histidine (H133). The NOG C-5 carboxylate forms a salt bridge with the side chain of ycfD$_{RM}$ R148 located on βIV of the DSBH. The NOG C5-carboxylate forms an apparent hydrogen bonding interaction with the side chain of S121 (2.8 Å), whilst the NOG oxalyl carboxylate is positioned to form a hydrogen-bonding interaction (2.7 Å) with the side-chain hydroxyl of S208. NOG is coloured green.

Since the outset of this project, crystal structures of ycfD.Mn^{2+}.2OG and ycfD.Mn^{2+}.succinate became available (in-house data by Schofield group). A comparison with the ycfD$_{RM}$.Mn^{2+}.NOG.L16$_{RM}$ crystal structure led to the proposal of a tentative model of the ycfD catalytic cycle with L16 R81/82 hydroxylation.

According to the classification by Hangasky *et al*., the 2OG oxalyl carboxylate can exist in the 'clock' (opposite His$_{n+m}$) and 'anti' (opposite His$_n$) positions (9) (Figure 5.25). In the ycfD$_{EC}$.2OG crystal structure, the 2OG oxalyl carboxylate is observed in the 'clock' position (Figure 5.25). The coordination site immediately adjacent to side-chain guanidine of R82 is thus occupied by a 2OG oxalyl carboxylate oxygen, possibly preventing the binding of molecular oxygen. This may correspond to an 'inactive' position of 2OG. In contrast, in the presence of the L16$_{RM}$ substrate fragment, the 2OG oxalyl carboxylate is observed coordinated to the metal in the 'anti' position, and the coordination site *anti* to His$_{n+m}$ is likely available for binding of molecular oxygen at the coordination site in closest proximity to both the metal and the target residue (Figure 5.25). In the ycfD$_{EC}$.Mn^{2+}.succinate structure one of the succinate carboxylate oxygen atoms binds anti to Asp$_{n+2}$ whilst the second oxygen atom binds anti to His$_{n+m}$, consistent with oxygen attack from position *anti* to His$_{n+m}$ (Figure 5.25). Only a rearrangement of 2OG coordination from 'clock' to 'anti' would enable the oxygen attack from the coordination position *anti* to the His$_{n+m}$ that would result in the observed position of the succinate carboxylate (assuming that the succinate carboxylate oxygen atoms occupy distinct coordination sites on the metal centre and the crystallographic data is not over-interpreted). It is possible that the 'clock' position of 2OG prevents or at least slows down uncoupled turnover of 2OG in the absence of primary substrate. This may be required at the high temperatures (55–75°C) at which ycfD$_{RM}$ operates. To further investigate this hypothesis, a complementary set of crystal structures would be useful, including ycfD$_{EC}$.Mn^{2+}.NOG.L16 or ycfD$_{RM}$.Mn^{2+}.NOG. Despite a sustained effort, it was impossible to obtain crystals of either complex.

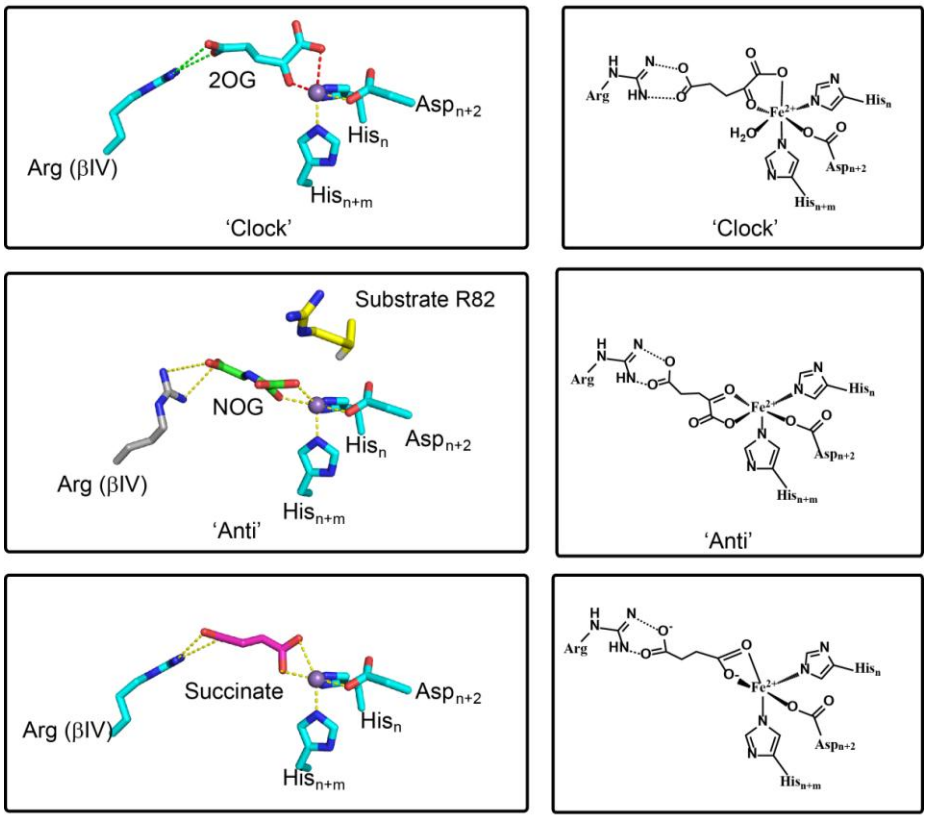


Figure 5.25: Changes in the position of the NOG/2OG oxalyl carboxylate between ycfD$_{EC}$.Mn^{2+}.2OG, ycfD$_{RM}$.Mn^{2+}.NOG.L16$_{RM}$ and ycfD$_{EC}$.Mn^{2+}.succinate structures. Top: ycfD$_{EC}$.Mn^{2+}.2OG, middle: ycfD$_{RM}$.Mn^{2+}.NOG.L16$_{RM}$, botto: ycfD$_{EC}$.Mn^{2+}.succinate. A schematic representation of metal coordination for each structure is presented on the right hand side. In the ycfD$_{EC}$.2OG crystal structure (top), the 2OG oxalyl carboxylate is observed in the 'clock' position (*anti* to His$_{n+m}$). The coordination site immediately adjacent to substrate R82 is thus occupied by the 2OG oxalyl carboxylate, possibly preventing the binding of molecular oxygen. In the presence of the L16$_{RM}$ substrate fragment (middle), the 2OG oxalyl carboxylate is in the 'anti' position (*anti* to His$_n$), and the coordination site *anti* to His$_{n+m}$ is likely available for binding of molecular oxygen at the coordination site in the closest proximity to both the metal and the target residue. In the ycfD$_{EC}$.Mn^{2+}.succinate structure (bottom) one of the succinate carboxylate oxygen atoms binds *anti* to Asp$_{n+2}$ whilst the second oxygen atom binds *anti* to His$_{n+m}$, consistent with a nucleophilic attack by oxygen from position *anti* to His$_{n+m}$.

5.7.1. Substitution of the basic 2OG-binding residue on βVIII

Protein sequence alignment of JmjC domain-containing proteins from prokaryotes and eukaryotes reveals that R148, the basic residue on βIV, which interacts with the 2OG C5-carboxylate, is substituted by a lysine in the majority of other JmjC domain-containing proteins, including NO66 homologues from bacteria (see Chapter 4, Figure 4.1), and all human JmjC domain-containing 2OG oxygenases, with the exception of JMJD3, UTX and UTY (Figure 5.26).

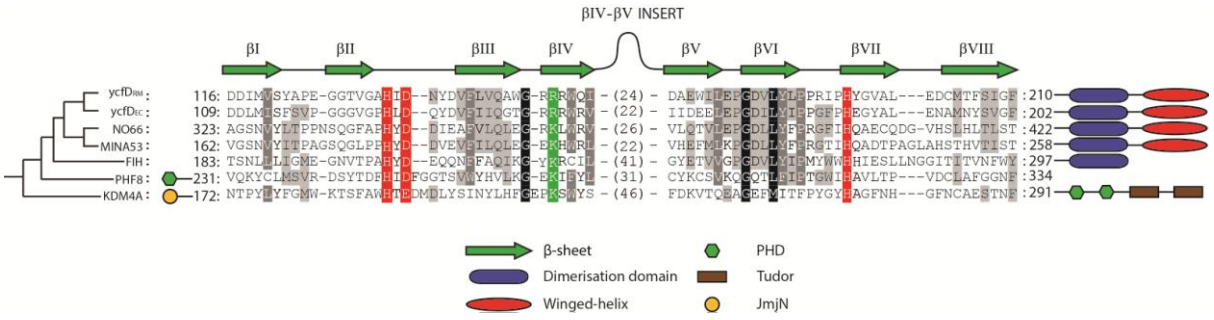
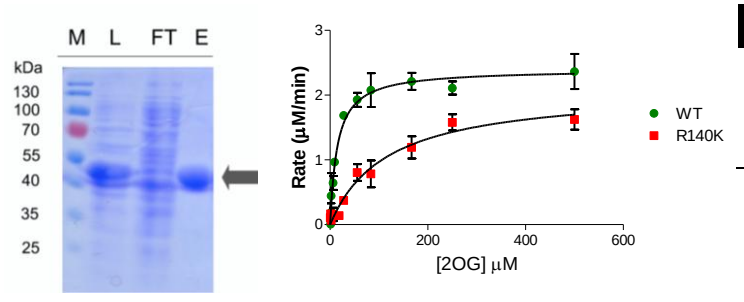


Figure 5.26: Protein sequence alignment of the DSBH core from several JmjC-domain containing 2OG oxygenases, including ribosomal oxygenases (ycfD$_{RM}$, ycfD$_{EC}$, NO66 and MINA53), HIF-1a asparagine hydroxylase FIH, and histone demethylases PHF8 and KDM4A (JMJD2A)) reveal that lysine at strand βIV is the most common residue binding 2OG. The structure-based alignment was generated using Strap software, and edited in GeneDoc. The dendrogram indicates likely evolutionary relationships between the proteins; nodes represent the last common ancestor of two diverging sequences. The DSBH secondary structure is indicated with green arrows above the alignment (Roman numerals indicate the strand identity within the DSBH; assignment was based on the ycfD$_{RM}$ crystal structure). Extra domains are indicated with blue cartouche (dimerisation domain), red ellipse (winged-helix), green hexagon (PHD), brown rectangle (Tudor) and yellow circle (JmjN). The catalytic triad residues are highlighted in red; the basic residue on strand βIV that interacts with 2OG is coloured red.

The higher occurrency of lysine as the basic 2OG-binding residue in JmjC-domain containing 2OG oxygenases suggests that arginine-to-lysine substitution may have conferred an evolutionary advantage, possibly due to a higher catalytic activity of the variant. In order to test this hypothesis, a ycfD$_{EC}$ variant R140K was prepared by site-directed mutagenesis. The rate-dependence of ycfD$_{EC}$ WT and R140K variants on 2OG was studied by Michaelis-Menten kinetic analysis. The results demonstrated that v_{max} values for WT and R140K are identical within experimental error (2.4±0.1 and 2.1±0.2 μM/min, respectively, Figure 5.27). In contrast, the K_m^{app} value for R140K increased to 116±27 μM from 15±2.5 μM for WT. This 7-fold increase in K_m^{app} value suggested that binding of 2OG is perturbed in the R140K variant.



	WT	R140K
V_{max} (μM/min)	2.4±0.1	2.1±0.2
K_m^{app} (μM)	15±2.5	116±27
R^2	0.93	0.90

Figure 5.27: Rate dependence of ycfD$_{EC}$ WT and R140K variants on 2OG. A) SDPAGE gel of the ycfD$_{EC}$ variant R140K prepared by site-directed mutagenesis and purified by Ni(II)-affinity chromatography; b) ycfD$_{EC}$ variants (2 μM) were incubated with Fe(II), ascorbate and L16$_{EC}$ substrate fragment (100 μM, 1 mM and 100 μM, respectively) and a serial dilution of 2OG. Reaction mixtures were incubated at 37°C for 3 min before quenching with CF_3COOH_{aq} (1% vol/vol). Reactions were performed in triplicate and analysed by MALDI-MS. Data were fitted with the Michaelis-Menten function (GraphPad Prism, version 5.02) to derive the kinetic parameters. The error bars denote the s.d. from the mean.

Another possibility is that more than a single substitution is required to rationalise the effect of R-K substitution. Comparing the microenvironment of ycfD$_{RM}$/NO66 NOG binding sites reveals that no residues in the immediate proximity of the basic residue on βIV are conserved (Figure 5.28). In particular, Y328 (S122 in ycfD$_{RM}$) provides a steric constraint for NOG binding in NO66, which appears to 'flatten' the NOG chain; S122 ycfD$_{RM}$ instead provides a potential hydrogen bonding interaction with the C5 carboxylate, which results in ~25° tilt in the of the C5 carboxylate relative to NOG in NO66 (whilst the oxalyl carboxylates approximately overlap), presumably weakening the interaction with the basic residue at βIV; lysine provides a weaker hydrogen-bonding interaction compared to arginine, which can thus explain the increase observed K_m^{app}(2OG) in the R140K variant.

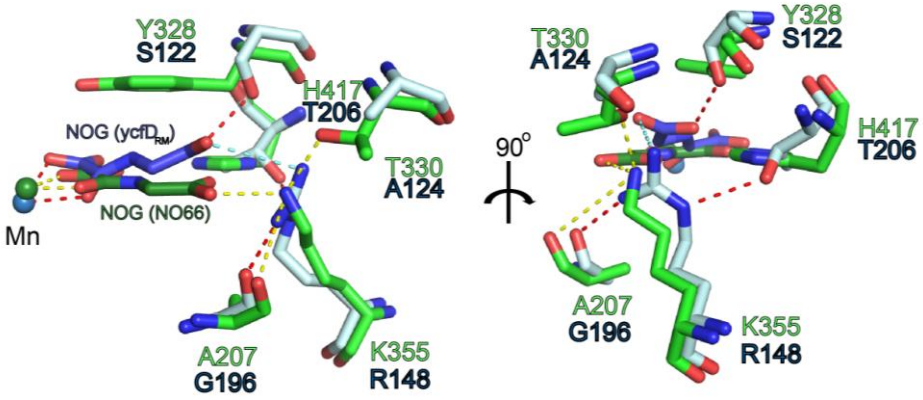


Figure 5.28: Comparison of NOG binding residues on ycfD$_{RM}$ (light blue) and NO66 (green). NOG bound to ycfD$_{RM}$ is dark blue, NOG bound to NO66 is dark green.

5.8. Induced fit upon L16 Binding

Substantial rearrangement of tertiary structure upon substrate binding has been reported for several 2OG oxygenases. For example, the β2-β3 loop on the HIF-1α hydroxylase PHD2, which is found in 'open' conformation in the absence of HIF-1α substrate, is closed in substrate bound form, forming several important electrostatic interactions with the substrate backbone (15).

Comparison of the ycfD$_{RM}$ *apo* and substrate-bound structures reveals two important features: i) no major rearrangements of the tertiary structure upon substrate binding were observed (i.e. there are no domain movements relative to each other, Figure 5.29); ii) a substantial conformational

rearrangement of the substrate-binding residue side chains was observed upon substrate binding, suggesting an induced-fit mechanism (Figure 5.29). In the absence of the substrate, the side chain of R212 in ycfD$_{RM}$ forms a salt-bridge interaction with the side chain carboxylate of E218 (2.3 Å); in the substrate-bound crystal structure, the R212 side chain is poised to form a salt-bridge interaction with the L16$_{RM}$ residue E80 (2.6 Å, Figure 5.29), ~8.2 Å away from its position observed in the *apo* structure. This R212$_{ycfDRM}$-E80$_{L16}$ salt bridge appears to be crucial for the activity of ycfD$_{RM}$ against L16$_{RM}$, as demonstrated by an alanine scanning experiment which revealed that the E80A substitution in L16$_{RM}$ abrogates the activity of ycfD$_{RM}$ against variant L16$_{RM}$ (see Chapter 4).

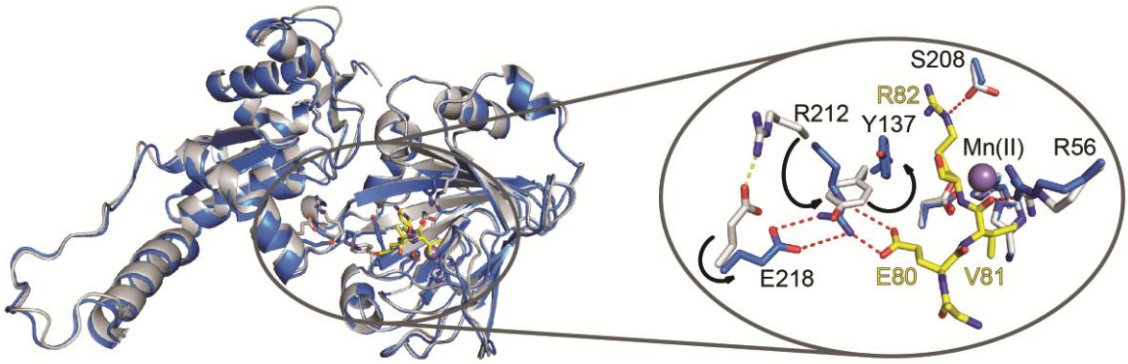


Figure 5.29: Induced-fit mechanism of YcfD$_{RM}$ substrate binding. In the absence of the substrate, ycfD$_{RM}$ R212 forms a salt-bridge interaction with E218 carboxylate (2.3 Å). In the substrate-bound structure, ycfD$_{RM}$ R212 is located 8.2Å away from its position in the *apo* structure, forming a salt-bridge interaction with L16$_{RM}$ residue E80 (2.6 Å). Grey: *apo* ycfD$_{RM}$, aquamarine: ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$.

In the substrate-bound form of ycfD$_{RM}$, the side chain of R212 forms an apparent cation-π interaction with the side chain of Y137 (Figure 5.29). Interestingly, Y137 is also involved in an apparent cation-π interaction with L16$_{RM}$ R82, thereby forming a R212 $_{ycfDRM}$ -Y137$_{ycfDRM}$-R82$_{L16RM}$ cation-π-cation sandwich complex.

5.8.1. Comparison of ycfD$_{RM}$ and ycfD$_{EC}$ substrate-binding residue conformations

Comparison of the substrate-bound structure of ycfD$_{RM}$ with the ycfD$_{EC}$ *apo* structure reveals a difference in the conformation of the βIV-βV insert, which is likely involved in substrate binding. The amino acid conservation between ycfD$_{RM}$ with ycfD$_{EC}$ is very low in this region (ycfD$_{RM}$ residues 151-174); see Figure 4.2 in Chapter 4), therefore the conformational differences between the βIV-βV inserts may reflect inter-species differences rather than a structural rearrangement with functional

importance for the enzyme. YcfD$_{EC}$ substrate-binding residues D110, Y129 and R203 are in an '*apo*' conformation (Figure 5.30). It is notable that the ycfD$_{EC}$ residue R203 is conserved (R212 in ycfD$_{RM}$); in Chapter 4 it was demonstrated that ycfD$_{EC}$ has a higher affinity for the L16$_{EC}$ variant A80E, as demonstrated by kinetics showing a decrease in K_m^{app} and loss of product inhibition. It is likely that the A80E variant is a better substrate for ycfD$_{EC}$ due to the interaction with the conserved R203 residue.

The side chain hydroxyl group of S199 points away from the proposed substrate binding site in the ycfD$_{EC}$ structure. Overall, the comparison of the *apo* ycfD$_{EC}$ with the *apo*- and the substrate-bound ycfD$_{RM}$ crystal structures suggests that it is possible that a conformational rearrangement of the ycfD$_{EC}$ residues as observed in the ycfD$_{RM}$ structures could take place in order to accommodate the substrate (i.e. residues D110, Y129 and R203 likely undergo analogous conformational changes upon substrate binding).

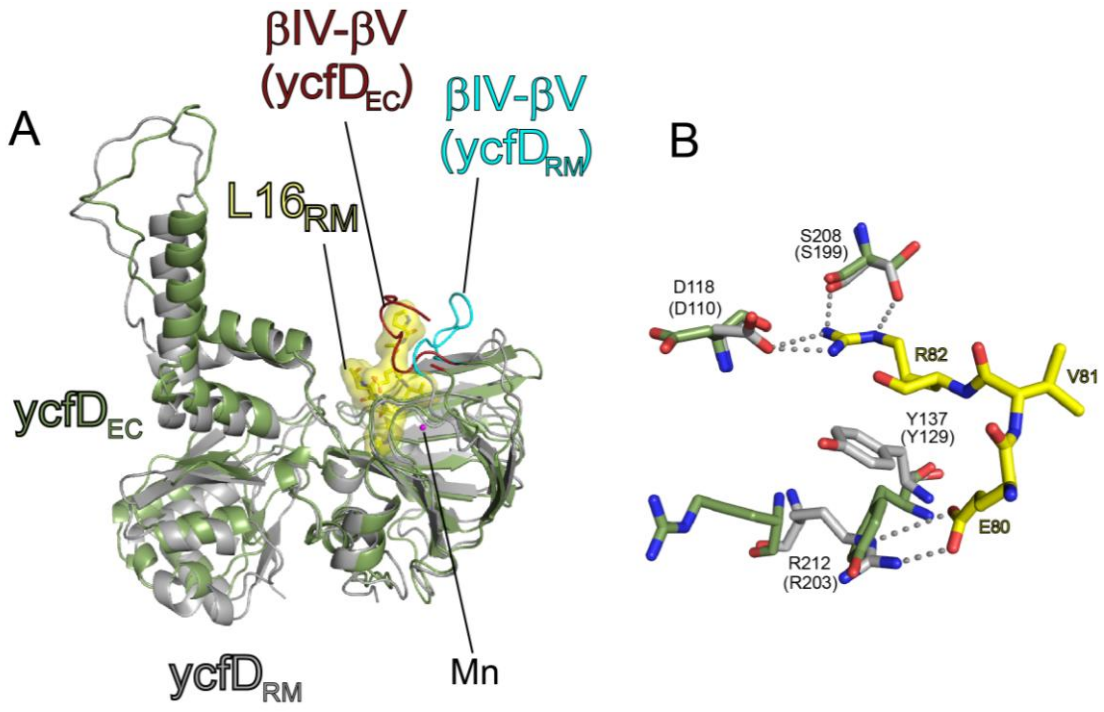


Figure 5.30: Overlay of ycfD$_{EC}$/ycfD$_{RM}$ structures (left) shows a difference in tertiary structures of two enzymes. A) Alignment of the JmjC domains shows an approximately 10° tilt of dimerisation domain helices, as well as a difference in the conformation of the βIV-βV insert (highlighted red for ycfD$_{EC}$ and cyan for ycfD$_{RM}$). B) Conserved residues involved in substrate binding are found in the same position for ycfD$_{RM}$ and ycfD$_{EC}$, but exist in different conformations when compared to the substrate-bound ycfD$_{RM}$. YcfD$_{EC}$ residues (forest green) are found in the 'inactive' conformation, whereas ycfD$_{RM}$ residues (grey) are in the 'active', substrate bound conformation.

5.9. A hydrophobic pocket accommodates the substrate residue V81

L16$_{RM}$ residue V81 was found to protrude into a conserved hydrophobic pocket, lined by residues I58, V69, and L90, located on the *N*-terminal β-strands 1,2,3, respectively; V30, located on βII, and V168 and L170, located on the βIV-βV insert (Figure 5.31). The *N*-terminal β-strands thus likely participate in hydrophobic interactions with the substrate; furthermore, the residues on the βIV-βV insert, which has been found to participate in substrate binding (14) are also predicted to interact with the substrate through hydrophobic interactions.

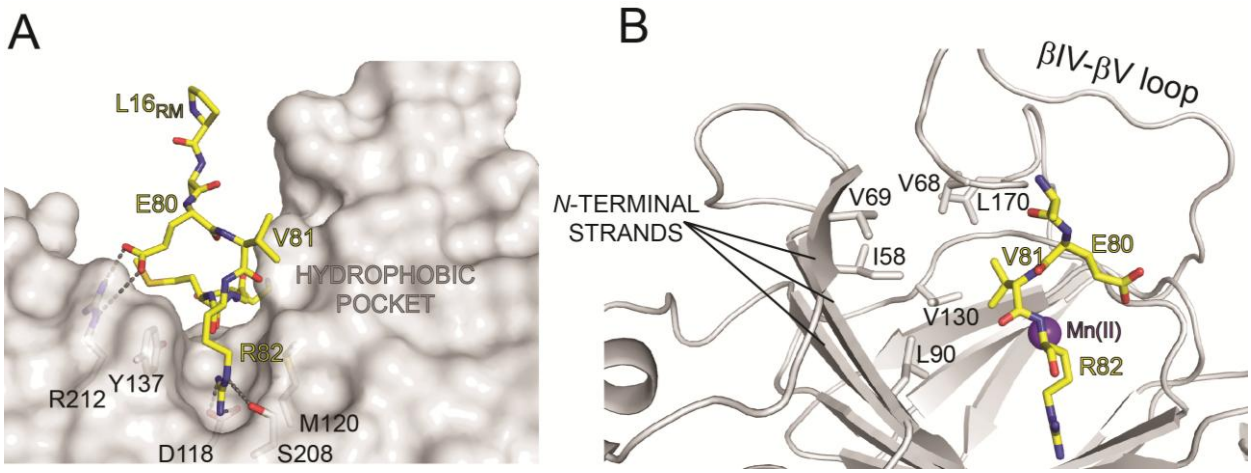


Figure 5.31: A view of the substrate binding pocket showing the location of the hydrophobic pocket relative to the R82 binding pocket. A) L16$_{RM}$ R82 (yellow) protrudes into a deep pocket lined by D118, M120, S208 and Y137. L16$_{RM}$ E80 is involved in a salt-bridge interaction with R212 and V81 protrudes into a hydrophobic pocket. The ycfD$_{RM}$ surface representation was designed using the BP Surface plugin in Pymol (Schrödinger). B) A cartoon representation of the hydrophobic substrate binding pocket showing interactions between L16$_{RM}$ V81 and ycfD$_{RM}$ residues I58, V69, L90 from *N*-terminal helices 1-3 and V168 and L170 from the βIV-βV insert.

5.10. The dimerisation domain

The dimerisation domain of ycfD$_{RM}$ (residues 214-280) comprises an orthogonal bundle of 3 α-helices separated by a loop which protrudes into the active site of the other protomer and possesses a C_2 centre of symmetry (Figure 5.32a). The dimerisation domain of ycfD$_{RM}$ is highly homologous to the NO66/MINA53 dimerisation domain, which contain also expansion elements compared to ycfD$_{RM}$ (Figure 5.32b). The dimer domain of ROX (ycfD, NO66, MINA53) is reminiscent of, but different from, the *C*-terminal dimerisation domain of FIH (16), a JmjC-domain containing 2OG oxygenase (Figure 5.32c).

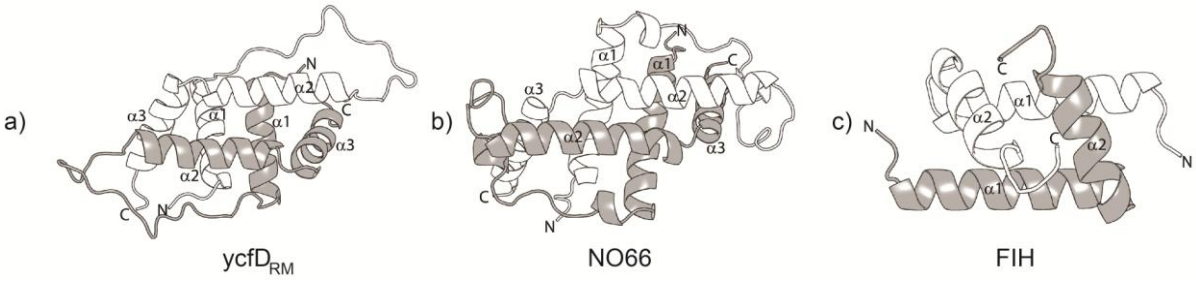


Figure 5.32: Schematic representations of dimerisation domains from a) ycfD$_{RM}$, b) NO66 and c) FIH. The dimerisation domain of ycfD$_{RM}$ is highly homologous to NO66/MINA53 dimerisation domain. The dimeric interface of ROX (ycfD, NO66, MINA53) is reminiscent of, but different from, the dimerisation domain of FIH.

Three main contributing interactions contribute to ROX dimer stability: i) inter-helical hydrophobic interactions, ii) hydrogen bonding interactions and iii) formation of aromatic cages at the dimer interface.

i) Hydrophobic interactions in the dimerisation domain of ycfD$_{RM/EC}$

Hydrophobic interactions between α-helices are the predominant type of interaction between the ROX monomers. In ycfD$_{EC}$ and ycfD$_{RM}$ the chemical nature of residues involved in hydrophobic interactions is well-conserved (ycfD$_{EC}$ residues are in parentheses): M222 (F214), I249 (V242), L254 (M247), I257 (L250), L260 (M253), L261 (M254), L264 (L257), I256 (I258). These residues are predominantly located on the helix α-2, which interacts with α-2 of the other monomer.

ii) Salt-bridge interactions in the dimerisation domain of ycfD$_{RM}$/$_{EC}$

Hydrogen bonding/salt-bridge formation is apparently more important for dimerisation in ycfD$_{RM}$, than ycfD$_{EC}$; in ycfD$_{RM}$ the network of hydrogen bonds between the two monomers A and B includes D255$_A$-R268$_B$-Q258$_A$-D266$_B$-R262$_A$ which, due to two-fold symmetry, makes a total of 8 hydrogen bonds/salt-bridges (Figure 5.33). In ycfD$_{EC}$ this network is disrupted by L255, which is positioned in the centre of the network (equivalent to R262 in ycfD$_{RM}$, Figure 5.33).

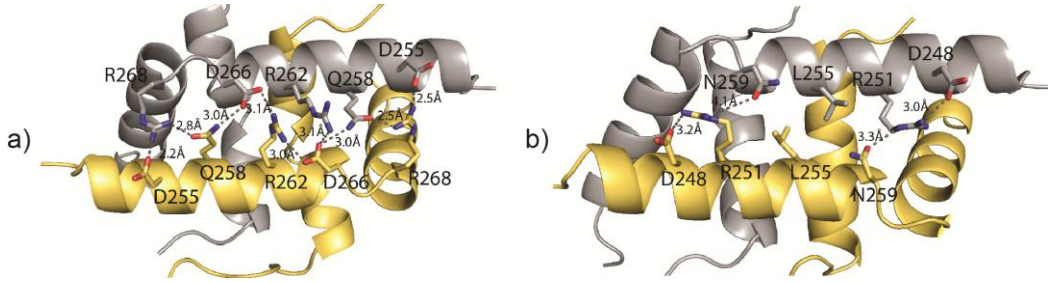


Figure 5.33: Dimerisation interface of a) ycfD$_{RM}$ and b) ycfD$_{EC}$. A salt-bridge-hydrogen bonding network spanning D255$_A$-R268$_B$-Q258$_A$-D266$_B$-R262$_A$ is found in ycfD$_{RM}$, but not in ycfD$_{EC}$ where the central residue is hydrophobic (L255).

It has been reported that salt-bridges increase the thermostability of thermophilic proteins (8). Since the salt-bridge forming residues found in ycfD$_{RM}$ are not conserved in ycfD$_{EC}$, they may be a contributing factor to the high thermostability observed for ycfD$_{RM}$ relative to ycfD$_{EC}$ (discussed in Chapter 4). In order to test whether the network of hydrogen bonds and salt bridges in the dimerisation domain contributes to the thermal stability, a ycfD$_{EC}$ variant with an interaction network, like that observed in ycfDRM, was prepared (Figure 5.34).

5.10.1. Testing the effect of salt-bridge interactions in the ycfD dimerisation domain

Three residues on the dimerisation helix α-2 (R251, L225 and N259) and a residue from helix α-3 (P261) of ycfD$_{EC}$ were replaced with analogous residues from the ycfD$_{RM}$ dimerisation domain to create a ycfD$_{EC}$ R251Q_L225R_N259D_P261R mimetic. The R251Q_L225R_N259D_P261R variant was prepared by successive site-directed mutagenesis steps and expressed in *E. coli* BL21 (DE3). Denaturing ESI-MS of the variant showed a major single peak at 44997 Da, whereas the mass of WT ycfD$_{EC}$ was 44923 Da; the mass difference was consistent with correct replacement of the four residues (expected +75 Da; observed +74 Da, Figure 5.34B). A DSF investigation then measured the thermal stabilisation properties of the mimetic protein (Figure 5.34c). YcfD$_{EC}$ WT was tested alongside, no co-factors or inhibitors were added. The result demonstrated that the mimetic ycfD$_{EC}$ is not thermally stabilised relative to WT; destabilisation by 0.5^{o}C relative to WT was observed instead (Figure 5.34C).

It was therefore hypothesised that the variant may contribute towards stability at high salt concentration. A native PAGE experiment with ycfD$_{EC}$ WT and R251Q_L225R_N259D_P261R was conducted in the presence of an increasing concentration of NaCl (0-1 M). Two bands were observed for ycfD$_{WT}$, presumably corresponding to the monomeric and dimeric forms of ycfD$_{EC}$ in solution. The band with lower mobility, predicted to correspond to the dimeric form of ycfD$_{EC}$, increased in intensity at high concentration of NaCl, suggesting that salt stabilises the dimeric form of WT ycfD$_{EC}$. (Figure 5.34D). Three strong bands were observed for ycfD$_{EC}$ R251Q_L225R_N259D_P261R, as well as two weaker bands at lower mobility and a degree of smearing below the highest mobility band. In contrast to ycfD$_{EC}$ WT, the lower mobility bands decreased in intensity with increasing

concentration of NaCl, suggesting that salt decreases the stability of oligomers. The presence of multiple bands observed by native PAGE for the ycfD$_{EC}$ mimetic suggested partial protein degradation; this was excluded by analysing the denaturing mass spectrum between 10-50 kDa; only the peak corresponding to nominal mass was observed (data not shown). It is likely, therefore, that ycfD$_{EC}$ R251Q_L225R_N259D_P261R forms oligomers of higher order compared to WT (e.g. trimer, tetramer) which are disrupted by NaCl.

At least two possible explanations of the results exist: (i) the replacement of WT residues in ycfD$_{EC}$ at structurally important sites leads to structural destabilisation due to unfavourable interactions with neighbouring residues (e.g. proximity of nearby negative charges) which were not replaced by site-directed mutagenesis, and the result is artefactual; or, (ii) the hydrophobic surface of ycfD$_{EC}$ exerts more stabilisation of the dimeric form of ycfD than the electrostatic interactions that were predicted to exist in the mimetic protein. Consistent with this interpretation, a high concentration of salt would shield the electrostatic potential of charged residues, thus preventing the formation of salt-bridges, whilst stabilising hydrophobic interactions.

5.10.2. Disruption of dimerisation leads to loss of catalytic activity

In order to assess the importance of the dimerisation domain for the catalytic activity of ycfD$_{EC}$, the variant I211R was designed to disrupt the hydrophobic interaction between two α-helices on opposite protomers (Figure 5.35A). The I211R ycfD$_{EC}$ variant was produced in *E. coli* BL21(DE3) (Figure 5.35B) and its catalytic activity compared to WT ycfD$_{EC}$ in a MALDI-MS based activity assay. Whilst the WT showed 87% activity, the I211R variant was only 7% active (Figure 5.35C).

The variant was then analysed by native gel electrophoresis. BSA was used as a standard, because it resolves into a series of oligomers in solution (Figure 5.35D). YcfD$_{EC}$ WT resolved into two major peaks and a minor peak. Variant I211R resolved in a single peak with lower mobility than the WT peak, suggesting that surface charge of the species has changed, consistent with introduction of a

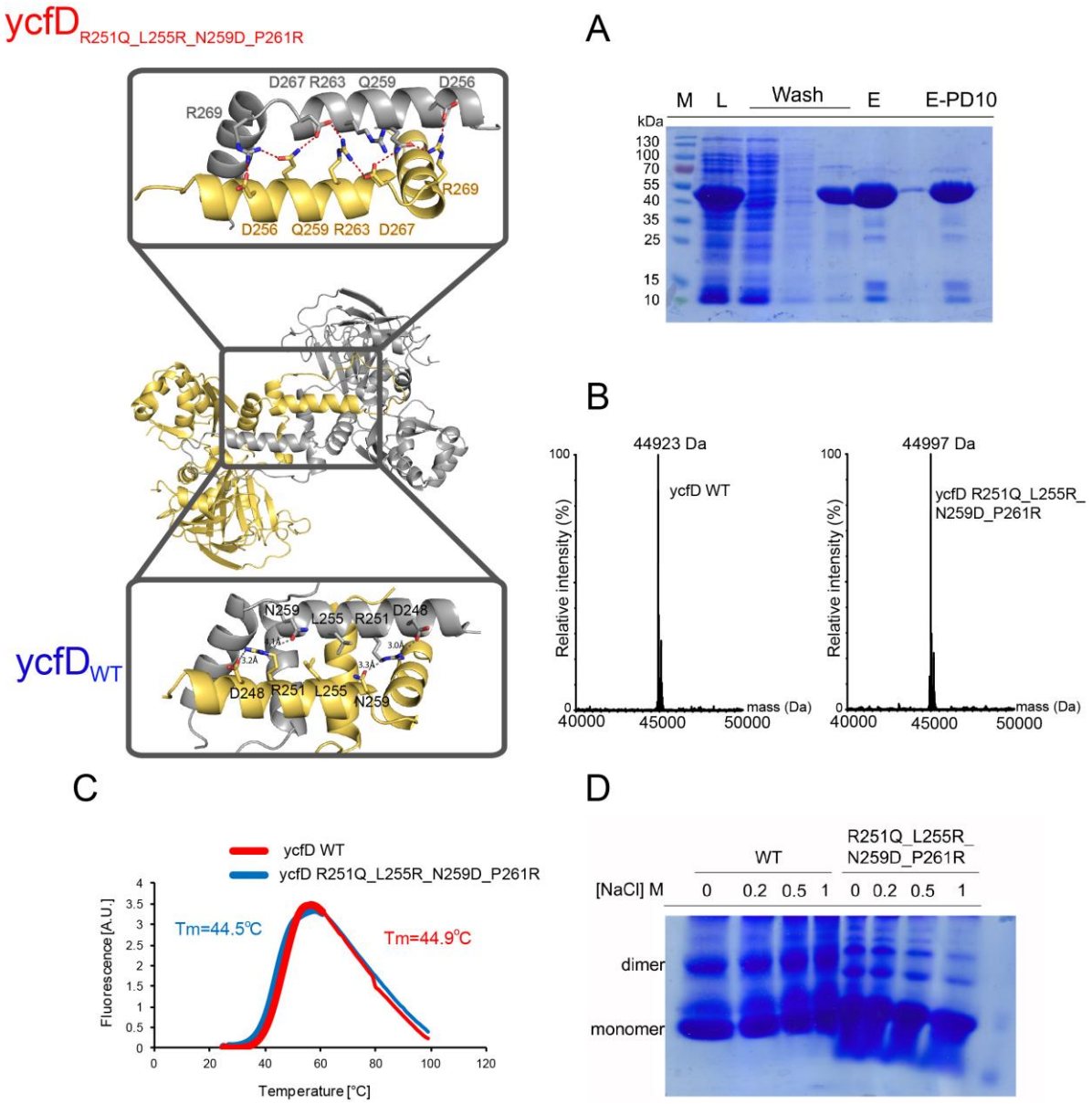


Figure 5.34: Thermal stability and salt-dependence of oligomerisation of a ycfD$_{EC}$ variant with features of ycfD$_{RM}$ dimerisation domain. A) Dimerisation interface of ycfD$_{EC}$ was converted into one more like that observed for ycfD$_{RM}$ by four rounds of site-directed mutagenesis. The resulting mutant was recombinantly expressed *E. coli* BL21 (DE3) and purified by HisTrap Ni(II)-affinity chromatography; B) Deconvoluted ESI-MS spectra of ycfD$_{EC}$ WT and ycfD$_{EC}$ R251Q_L225R_N259D_P261R confirm the correct mass of the mimetic protein (expected: 44998 Da, observed: 44997 Da); C) DSF study of protein thermal stability reveals destabiliation of ycfD$_{EC}$ mimetic by 0.4°C; D) native gel of WT and ycfD$_{EC}$ R251Q_L225R_N259D_P261R in presence of 0-1 M NaCl show an increase in the intensity of dimeric form of WT with increasing concentration of salt; multimeric forms of the mimetic variant decrease in intensity with increasing salt concentration.

surface-exposed charged residue. The second strong band present in the WT sample was not observed, suggesting that I221R loses the capacity to dimerise. Variant L160R was used as a control. The type of substitution was chemically identical, i.e. a branched nonpolar residue was substituted with a strongly positively charged residue; however, L160R is buried in the proximity of the active site, therefore the positive charge is not solvent-exposed. Indeed, the L160R variant resolved into a

major peak corresponding to the major peak observed in WT, and two minor peaks which migrated the same distance as WT peaks. L160R is catalytically active (data not shown). Overall, the results suggested that disruption of dimerisation by placing a positively charged residue into a hydrophobic dimerisation region abrogates catalytic activity ycfD$_{EC}$, and by inference, that dimerisation is crucial for the catalytic activity. Consistent with this observation, a distant eukaryotic homologue of ycfD, Factor Inhibiting HIF (FIH) had also been reported to lose activity upon disruption of dimerisation by a single point substitution (hydrophobic for basic) (17).

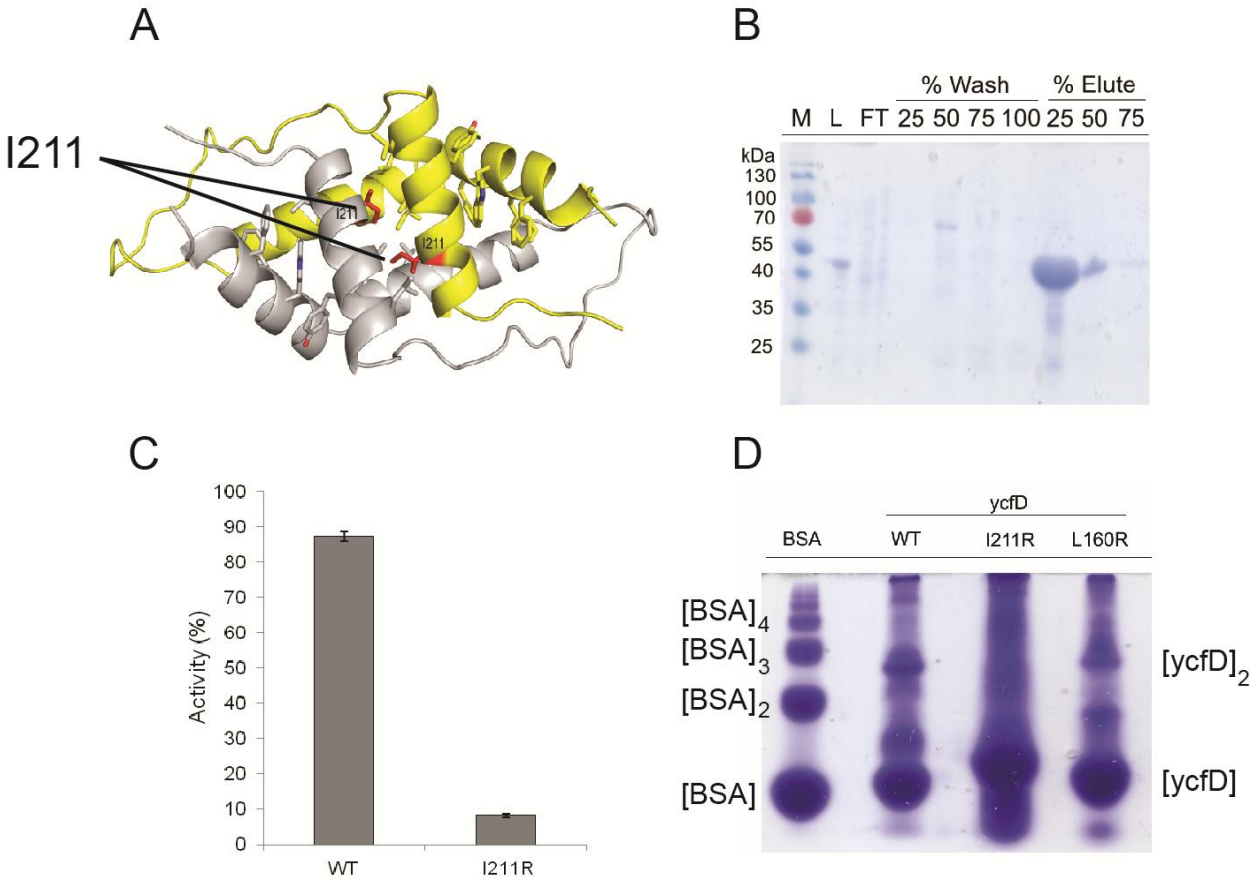


Figure 5.35: Studies on dimerisation domain of ycfD. A) Variant I211R was designed in order to place a positively charged residue at the dimer interface. B) YcfD$_{EC}$ I211R variant was produced in *E .coli* BL21 (DE3). SDS-PAGE gel demonstrates that >90% purity was achieved during purification. (M: molecular weight marker; L: lysate; FT: flow-through: % Wash: percentage of wash buffer used; % Elute: percentage of elution buffer used. C) Activity of ycfD WT and I211R was measured by MALDI-MS. Enzyme (1 μM) was incubated with 2OG, Fe^{2+} and ascorbate (200, 100 and 1000 μM respectively) in the presence of L16EC substrate fragment (20-mer, 50 μM). Reaction was allowed to proceed for 30 min at 37°C before quenching with equal amount of CF_3COOH_{aq} (1%). Reactions were performed in triplicate with error bars denoting s.d. from the mean. D) Native gel electrophoresis. Samples of BSA (10 mg), ycfD WT, I211R and L160R (all 10 mg) were loaded onto a 10% native gel and resolved at 80V for 4 h at 4°C.

5.11. The *C*-terminal ('winged-helix') domain

The dimerisation domain of ROX, including ycfD$_{EC}$, ycfD$_{RM}$, NO66 and MINA53, (but not FIH) continues into a *C*-terminal domain (residues 296-392), which comprises of 2 α-helices and 5 β-strands, which fold into two antiparallel β-sheets. Previous work reported that a *C*-terminal deletion construct of ycfD$_{EC}$ showed a complete loss of catalytic activity, as measured by the radioactive 2OG turnover assay (Dr Chiahua Ho, DPhil Thesis). This result suggested that the *C*-terminal domain is crucial for catalytic activity. The crystal structure of the substrate complex revealed that residue R285 from the *C*-terminus extends deeply into the active site of ycfD$_{RM}$, making a potential hydrogen bonding interaction with the backbone amide of substrate M83, as well a hydrogen bonding interaction with the side-chain carboxylate of D118, which is involved in salt-bridge interaction with the substrate residue R82 (Figure 5.36). The role of the *C*-terminus may therefore be to facilitate substrate binding in the JmjC active site.

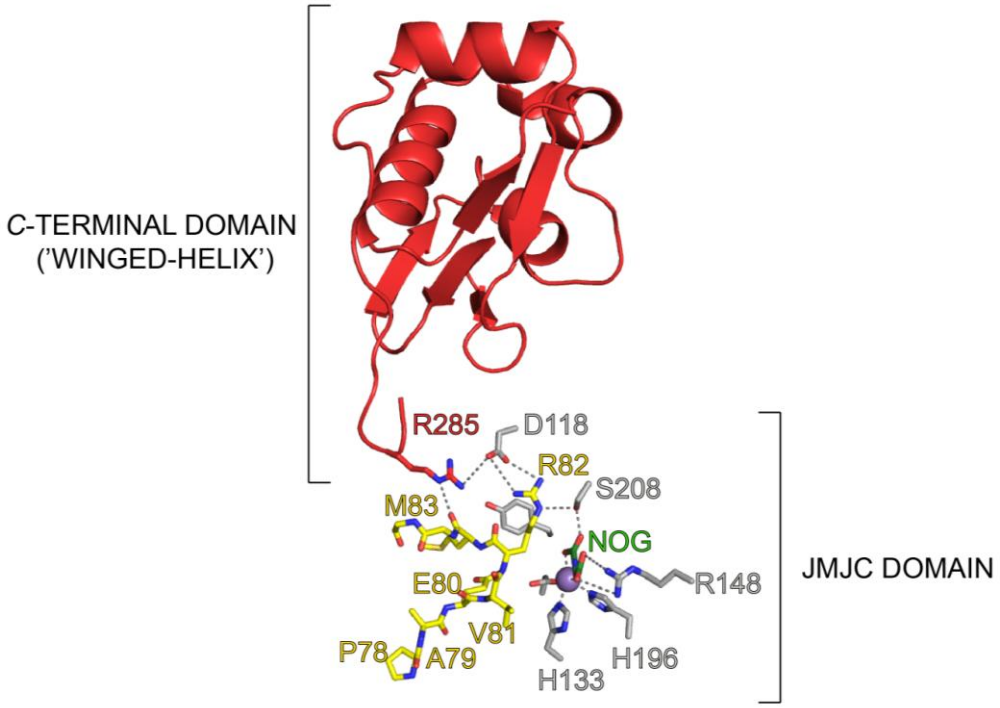


Figure 5.36: The *N*-terminal loop of the *C*-terminus protrudes into the JmjC domain of ycfD$_{RM}$; R285 from the *C*-terminal domain forms a hydrogen bonding interaction with M82 of L16$_{RM}$.

A Genetic Algorithm for Non-sequential, Gapped protein STructure Alignment (Gangsta+) (18) was then conducted to find structurally related proteins on ASTRAL40 (SCOP version 1.75) database; the search enables non-sequential protein structure alignment, which allows the detection of non-sequential structural analogues (proteins with overall structural similarity but different connectivity). The search defined the *C*-terminus as a variation of the winged helix (WH) fold. The WH domains, a subtype of the helix-turn-helix (HTH) family, are known as nucleic acid/protein interacting domains (19) and are found in different cellular pathways from transcriptional regulation to RNA processing (20). They are positively charged, enabling them to bind negatively charged DNA. In ycfD$_{RM}$, and other ROX, the overall charge of the fold is positive (calculated pI is 9.5-10.0), thus the possibility of binding rRNA cannot be excluded.

The closest structural homologue of the *C*-terminal domain identified by the Gangsta+ search engine was the 'peptide clamp domain' found in MccB (RMSD 3.3 Å) (Figure 5.37), an enzyme found in the biosynthetic pathway of microcin C7 antibiotic (21). The peptide clamp is important for the catalytic activity of MccB. The MccB was then superimposed onto the structure of ycfD$_{RM}$ to investigate the possibility that the *C*-terminus of ROX serves a similar purpose, i.e. peptide-binding (Figure 5.37). The position of the substrate fragment (MccA) in the MccB active site overlaps with the dimerisation helix in ycfD$_{RM}$ (Figure 5.37A); it is thus unlikely that the role of the peptide clamp in binding the substrate of MccB is analogous to the ROX *C*-terminal domain.

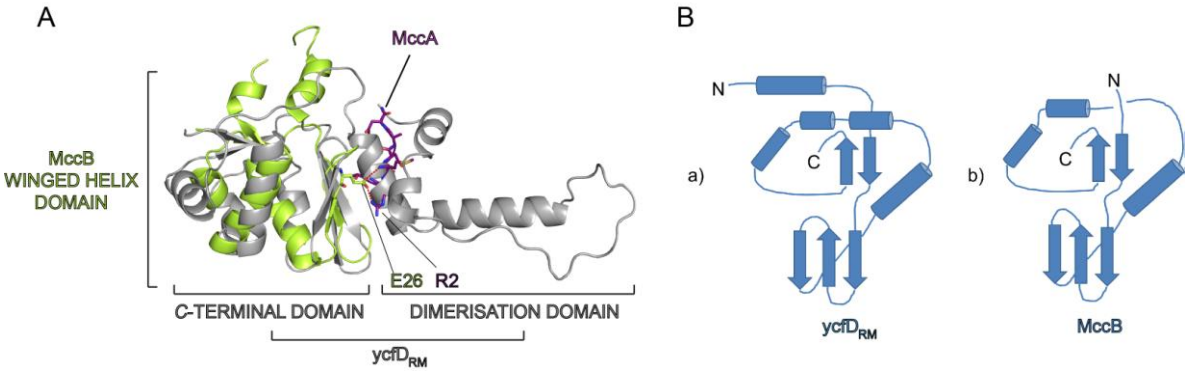


Figure 5.37: Superposition of the *C*-terminal domain of ycfD$_{RM}$ (residues 296-392) and the winged-helix domain of MccB (PDB: 3H9G). Structurally homologous fragments suggest a similar folding pattern, which may indicate analogous functions. B: Topology diagrams of a) ycfD$_{RM}$ and b) MccB (PDB: 3H9G) show a similar folding pattern, reminiscent of the 'winged-helix' fold. The two topologies are identical apart from the *N*-terminal helix present in ycfD$_{RM}$, which is missing in MccB. Arrows symbolise β-strands and the cylinders symbolise α-helices. *N*-/*C*-termini are indicated.

According to Gangsta+ search, the next closest structural homologue of ycfD$_{RM}$ *C*-terminus was the *C*-terminal domain of ribosomal protein L6 (RMSD 3.0 Å). In the structures of the *T. thermophilus* ribosome, L6 is found in the proximity of L16 but separated from L16 by 23S rRNA (22). It is possible, therefore that the role of the ROX *C*-terminus is to mimic the microenvironment of the ribosome structure where L16 binds. It is notable that the overall charge of the *C*-terminus oppositely matches that of L16 (Figure 5.38). In particular, an area of negatively charged residues spans the inside of the JmjC domain and the dimerisation domain; the body of L16 protein is positively charged, potentially providing electrostatically matched surfaces for the binding interaction of ycfD$_{RM}$ and L16$_{RM}$.

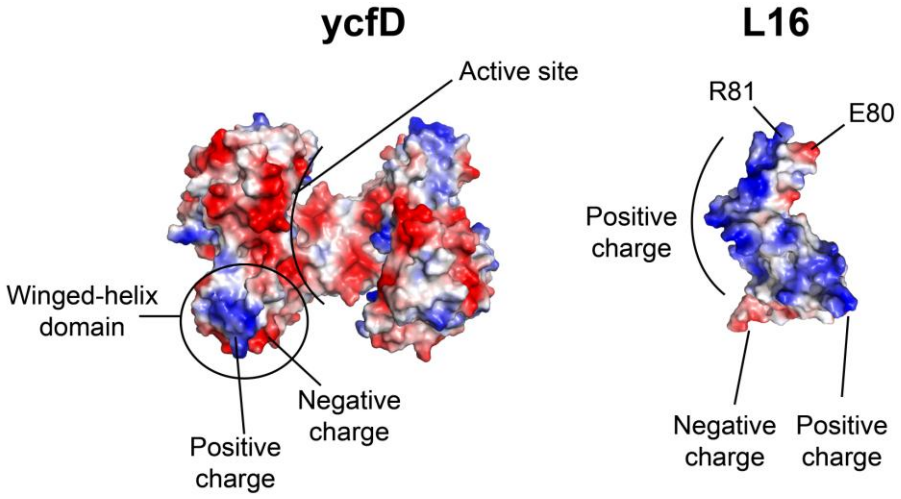


Figure 5.38: Charge distribution on the surface of ycfD and L16 reveals areas of complementary charges, which could indicate areas of interaction between ycfD and L16. Surface charge was calculated using the 'Vaccum electrostatics function in PyMol. Red colour indicates negative electrostatic potential; blue indicates positive electrostatic potential and white indicates neutral areas.

In order to elucidate the structural role of the *C*-terminus in substrate binding, a ycfD$_{RM}$.L16 protein-protein complex structure would be required. The remainder of the chapter presents the initial studies on obtaining a soluble ycfD.L16 protein-protein complex, which will inform further structural studies.

5.12. Towards the structure of a ycfD.L16 protein-protein complex

The ycfD$_{RM}$.Mn(II).NOG.L16$_{RM}$ structure allowed a more accurate prediction of points of close contact between ycfD$_{RM}$ and L16 fragment peptide. It was envisaged that a recombinantly expressed and purified protein-protein complex would provide useful information about the interactions between

ROX and ribosomal proteins; no ROX has been studied in complex with a whole ribosomal protein to date, therefore a number of questions remain unanswered, for example the stoichiometry of substrate binding and the role of the *C*-terminal domain.

In order to investigate the stoichiometry of L16 binding, a set of ycfD$_{EC}$/L16$_{RM}$ cysteine-containing variants were designed at points of close contact but sufficiently remote from the active site so as to prevent interference with accommodation of substrate residues in the active site.

Initially, the *L16$_{RM}$* gene was cloned from *R. marinus* gDNA (Figure 5.39). It was chosen over the *E. coli* homologue due to a higher affinity shown by ycfD$_{EC}$ for L16$_{RM}$ over L16$_{EC}$ sequences, as measured by kinetic studies discussed in Chapter 4.

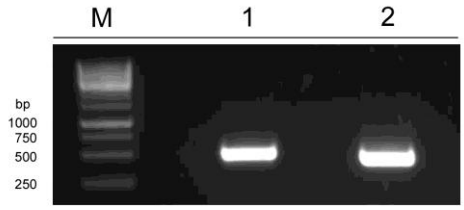


Figure 5.39: Agarose gel showing PCR amplification of gene coding for ribosomal protein L16$_{RM}$ from *R. marinus* gDNA. Two clones were generated: 1) region ± 20 nucleotides was amplified around L16, 2) L16$_{RM}$ was amplified from start to stop codon with 5' NdeI and 3' XhoI restriction sites.

The gene was sub-cloned into the pET28a vector using restriction sites NdeI and XhoI. Expression trials were conducted in order to establish optimal expression conditions for recombinant production of His$_6$-L16$_{RM}$. Concentration of IPTG used was 0.5 mM. Expression at 28^{o}C for 18 h led to the highest amount of recombinantly produced His$_6$-L16$_{RM,}$ as observed after small-scale purification through a HisTrap Ni(II)-affinity column (1 mL, GE Healthcare, Figure 5.40).

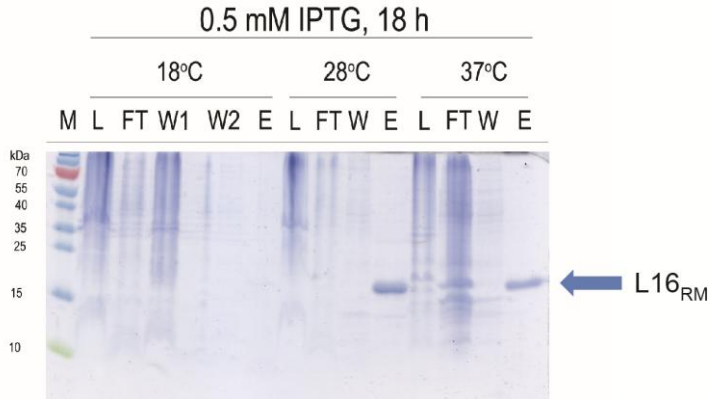


Figure 5.40: An SDS-PAGE gel (15%) of recombinantly produced His$_6$-L16$_{RM}$ purified by HisTrap Ni(II)-affinity column. M: molecular weight markers; L: lysate; FT: flow-through; W: wash; E: elution.

Having established that L16$_{RM}$ protein can be expressed in high amounts, cysteine-containing variants of L16$_{RM}$ and ycfD$_{EC}$ were designed based on the ycfD$_{RM}$ substrate structure. L16 substrate fragment was modelled into the ycfD$_{EC}$ active site. The model suggested the proximity of Cα of H277 ycfD$_{EC}$ to the Cα of L16 G84 (~2 Å distance Figure 5.41). Cysteine residues were then modelled at these positions and the distance between cysteinyl sulfur atoms was calculated at 2.5 Å (Figure 5.41).

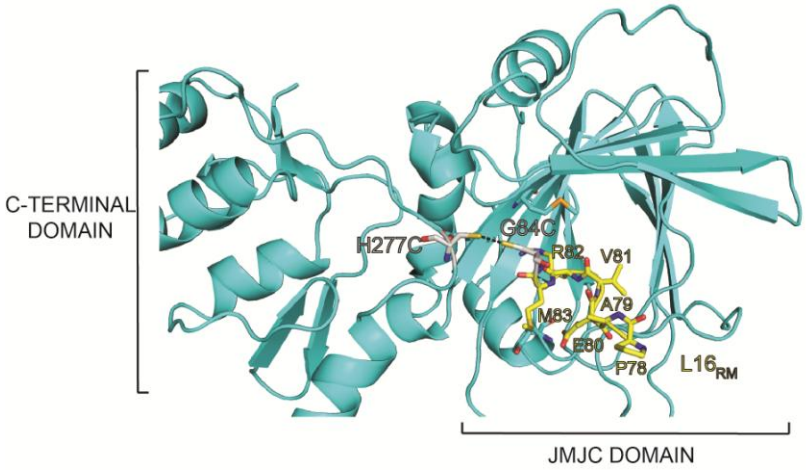


Figure 5.41: The L16$_{RM}$ substrate fragment chain was modelled into the active site of ycfD$_{EC}$ by superimposing the crystal structures of ycfD$_{RM}$ and ycfD$_{EC}$ and using the 'Pair-fitting' function in PyMol; (RMS 0.50). A model of ycfD$_{EC}$ H277C.L16$_{RM}$ G84C interaction. 'Mutagenesis' function (PyMol) was used to design the double variant and measure the predicted distance between cysteine side chains. The distance measured between ycfD$_{EC}$ H277C and L16$_{RM}$ G84C was 2.5 Å, which was considered optimal for the formation of a disulfide chemical cross-link.

Point mutations of ycfD$_{EC}$ (H277C) and L16$_{RM}$ (G84C, both as pET28a constructs) were created by site-directed mutagenesis. Protein production in BL21(DE3) was followed by co-lysis of H277C ycfD$_{EC}$ and G84C L16$_{RM}$-containing cell pellets. Purification of combined cell lysates used a single HisTrap column, resulting in two major bands at ~15 kDa and ~40 kDa, corresponding to L16$_{RM}$ and ycfD$_{EC}$ cysteine mutants, respectively (Figure 5.42).

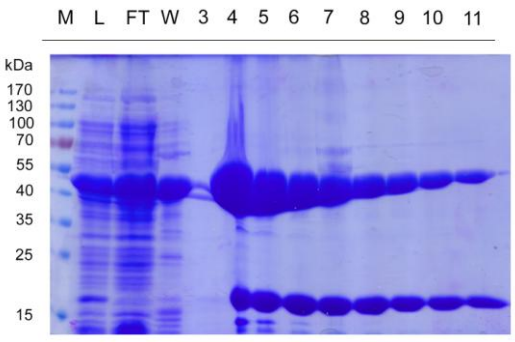


Figure 5.42: SDS-PAGE gel of ycfD$_{EC}$ H277C.L16$_{RM}$ G84C co-lysis. M: molecular weight markers; L: lysate; FT: flow-through; W: wash; numbers correspond to eluted fractions.

The combined eluate from HisTrap purification was further purified by size-exclusion chromatography. Two distinct peaks eluted at ~225 mL and ~260 mL (Figure 5.43). SDS-PAGE gel revealed that the first the peak (225 mL) contained a mixture of proteins of ~45 kDa and ~15 kDa, therefore corresponding to the ycfD$_{EC}$.L16$_{RM}$ complex (dissociated under the denaturing conditions of the SDS-PAGE gel). The eluate at 260 mL migrated at ~15 kDa, corresponding to L16$_{RM}$ G84C alone (Figure 5.43).

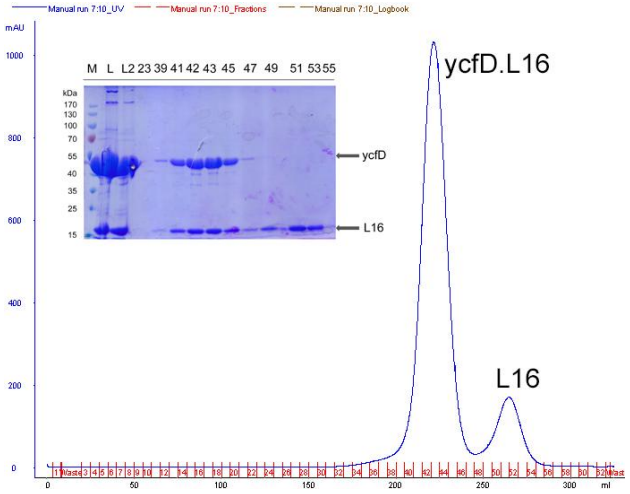


Figure 5.43: Chromatogram of size-exclusion purification of ycfD$_{EC}$ H277C.L16$_{RM}$ G84C. SDS-PAGE gel is shown as an inset with corresponding elution fractions indicated.

A non-denaturing mass spectrometry experiment was then conducted with ycfD$_{WT}$, and the ycfD$_{EC}$ H277C.L16$_{RM}$ G84C complex purified in the presence and absence of DTT (Figure 5.44). The non-denaturing mass spectrometry experiment was conducted by Eamonn Reading.

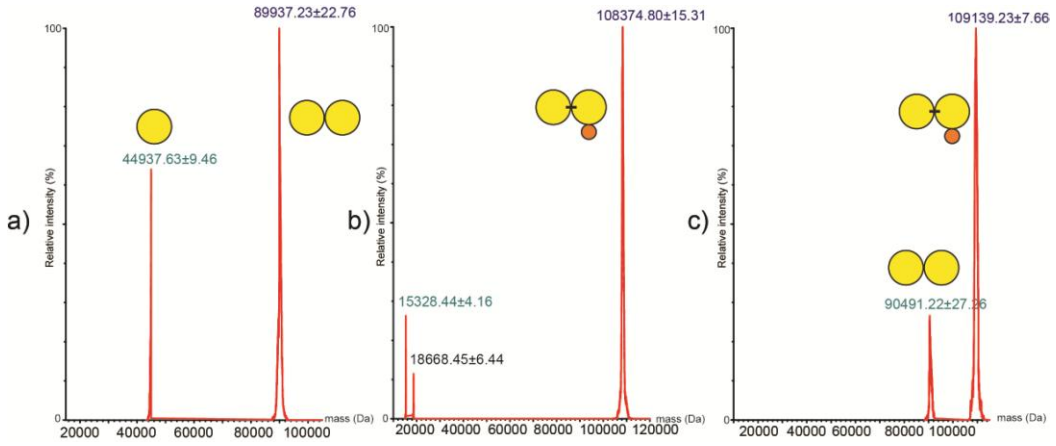


Figure 5.44: Deconvoluted non-denaturing mass spectrometry spectra of a) ycfD$_{EC}$ WT; b) ycfD$_{EC}$_H277C.L16$_{RM}$_G84C after gel filtration in the absence of DTT, c) ycfD$_{EC}$ H277C.L16$_{RM}$ G84C after gel filtration in presence of 100 μM DTT. Calculated mass: ycfD WT: 44605 Da, ycfD$_{EC}$ H277C: 44571 Da, L16$_{RM}$ G84C: 18347 Da.

The data demonstrated that two species of ycfD$_{WT}$ exist in the gas phase; the peak with lower abundance (44.9 kDa) corresponds to the monomeric form of ycfD$_{EC}$, and the peak with higher abundance (89.9 kDa) corresponds to the dimeric form of ycfD$_{WT}$ (Figure 5.44a). Interestingly, the ycfD$_{EC}$ H277C.L16G84C complex purified in the absence of DTT showed a single peak at 108.4 kDa (Figure 5.44b), corresponding to (ycfD$_{EC}$)$_2$L16 complex (calculated 108 kDa). This result suggests that, at least under the conditions tested, a ycfD dimer binds one molecule of L16. The ycfD$_{EC}$ H277C.L16G84C complex purified in the presence of 0.1 mM DTT contained two MS peaks, a low-abundance peak at 90.5 kDa, corresponding to the (ycfD)$_2$ dimer and a highly abundant peak at 109 kDa, corresponding to the (ycfD)$_2$L16 complex. The treatment with reducing agent therefore did not lead to dissociation of the complex, suggesting that the complex formation is highly specific.

Crystallographic characterisation of the complex was then attempted. The ycfD$_{EC}$ H277C.L16$_{RM}$ G84C complex was subjected to a broad crystallisation screen in the presence and absence of succinate. However, no crystals were observed after >1 month of equilibration at 20^{o}C. It is likely that the high concentration of NaCl required to stabilise the complex may interfere with formation of crystals and obtaining diffracting crystals was beyond the time-limits of the project.

This is the first report of a protein-protein complex of a ROX and an intact ribosomal protein. Non-denaturing MS data provides evidence that a ROX dimer binds one molecule of a ribosomal protein. In the future, molecular modelling coupled with collision cross-section experiments obtained by non-denaturing mass spectrometry will be used to predict the binding mode of L16 to ycfD dimer. The knowledge of stoichiometry may assist in correct modelling, which may also help to elucidate the role of the *C*-terminal domain, the function of which remains unclear.

5.13. Conclusions

This chapter describes the work leading to the first substrate complex crystal structure for a ROX. Initially, protein stabilisation studies identified FG2216 as a potent stabiliser of ycfD$_{RM}$ (ΔT_m of 6^{o}C), which was used to obtain a crystal structure of ycfD$_{RM}$ in the absence of the substrate. A crystal structure of ycfD$_{RM}$ in complex with L16$_{RM}$ substrate fragment was determined to 3.0 Å and revealed

the key interactions required for binding of the L16$_{RM}$ fragment. The substrate R82 of L16$_{RM}$ protrudes into a substrate-binding pocket lined by residues D118, M120, Y137 and S208. The side chain of D118 forms a salt-bridge interaction with the guanidinium side chain of substrate R82; the side chain of M119 lines one side of hydrophobic pocket to accommodate R82 and the hydroxyl group of S208 is positioned within 2.7 Å from the NOG oxalyl carboxylate and 3.5 Å from the substrate arginine Nε, indicating a potential hydrogen bonding interaction with both the side chain of substrate R82 and NOG. The backbone carbonyl oxygen of M83 L16$_{RM}$ forms a hydrogen bonding interaction with the guanidinium group of R285 ycfD$_{RM}$, which protrudes into the active site of the JmjC domain from the *C*-terminal domain. A substantial conformational rearrangement of the substrate-binding residues side chains was observed upon substrate binding, suggesting an induced-fit mechanism. In particular, the side chain of R212 forms a salt-bridge interaction with the L16$_{RM}$ residue E80 (2.6 Å), ~8.2 Å away from its position observed in the *apo* structure. This R212$_{ycfDRM}$-E80$_{L16}$ salt bridge appears to be crucial for the activity of ycfD$_{RM}$ against L16$_{RM}$ as demonstrated in Chapter 4.

The analysis of electrostatic interactions revealed that ycfD$_{RM}$ contains 30% more hydrophobic interactions and 100% more salt-bridge interactions than ycfD$_{EC}$, suggesting that hydrophobic interactions and salt bridges are important for thermal stabilisation of ycfD$_{RM}$. The binding mode of R82 in the active site of ycfD$_{RM}$ was compared to arginine hydroxylases VioC; comparison of the orientation of the arginyl/arginine side chains in the DSBH of the two oxygenases revealed structural explanations for the opposite stereochemistry of hydroxylation observed biochemically. Non-denaturing mass spectrometry on a chemically-crosslinked ycfD.L16 protein-protein complex provided evidence that one molecule of L16 binds to a dimer of variant ycfD.

The work has informed structural studies on evolutionarily related ROX, NO66 and MINA53, for which substrate structures have been obtained since the completion of the work described in this chapter.

5.14. References

1. Rose, N. R., McDonough, M. A., King, O. N., Kawamura, A., and Schofield, C. J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev* **40**, 4364-4397
2. Niesen, F., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **9**, 2212-2221
3. Hopkinson, R. J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K. H., Li, X. S., Kristensen, J. B. L., King, O. N. F., Chan, M. C., Yeoh, K. K., Choi, H., Walport, L. J., Thinnes, C. C., Bush, J. T., Lejeune, C., Rydzik, A. M., Rose, N. R., Bagg, E. A., McDonough, M. A., Krojer, T. J., Yue, W. W., Ng, S. S., Olsen, L., Brennan, P. E., Oppermann, U., Muller, S., Klose, R. J., Ratcliffe, P. J., Schofield, C. J., and Kawamura, A. (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci* **4**, 3110-3117
4. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221
5. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501
6. Tao, Y., Wu, M. H., Zhou, X., Yin, W., Hu, B., de Crombrugghe, B., Sinha, K. M., and Zang, J. Y. (2013) Structural Insights into Histone Demethylase NO66 in Interaction with Osteoblast-specific Transcription Factor Osterix and Gene Repression. *J Biol Chem* **288**, 16430-16437
7. Lam, S. Y., Yeung, R. C., Yu, T. H., Sze, K. H., and Wong, K. B. (2011) A rigidifying salt-bridge favors the activity of thermophilic enzyme at high temperatures at the expense of low-temperature activity. *PLoS Biol* **9**, e1001027
8. Chan, C. H., Yu, T. H., and Wong, K. B. (2011) Stabilizing salt-bridge enhances protein thermostability by reducing the heat capacity change of unfolding. *PLoS One* **6**, e21624
9. Hangasky, J. A., Taabazuing, C. Y., Valliere, M. A., and Knapp, M. J. (2013) Imposing function down a (cupin)-barrel: secondary structure and metal stereochemistry in the alpha KG-dependent oxygenases. *Metallomics* **5**, 287-301
10. Saban, E., Chen, Y. H., Hangasky, J. A., Taabazuing, C. Y., Holmes, B. E., and Knapp, M. J. (2011) The second coordination sphere of FIH controls hydroxylation. *Biochemistry* **50**, 4733-4740
11. Yin, X. H., McPhail, K. L., Kim, K. J., and Zabriskie, T. M. (2004) Formation of the nonproteinogenic amino acid 2S,3R-capreomycidine by VioD from the viomycin biosynthesis pathway. *Chembiochem* **5**, 1278-1281
12. Yin, X. H., and Zabriskie, T. M. (2004) VioC is a non-heme iron, alpha-ketoglutarate-dependent oxygenase that catalyzes the formation of 3S-hydroxy-L-arginine during viomycin biosynthesis. *Chembiochem* **5**, 1274-1277
13. Helmetag, V., Samel, S. A., Thomas, M. G., Marahiel, M. A., and Essen, L. O. (2009) Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. *FEBS J* **276**, 3669-3682
14. Aik, W., McDonough, M. A., Thalhammer, A., Chowdhury, R., and Schofield, C. J. (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struct Biol* **22**, 691-700
15. Chowdhury, R., McDonough, M. A., Mecinovic, J., Loenarz, C., Flashman, E., Hewitson, K. S., Domene, C., and Schofield, C. J. (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* **17**, 981-989
16. Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806

17. Lancaster, D. E., McNeill, L. A., McDonough, M. A., Aplin, R. T., Hewitson, K. S., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2004) Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J* **383**, 429-437
18. Guerler, A., and Knapp, E. W. (2008) Novel protein folds and their nonsequential structural analogs. *Protein Sci* **17**, 1374-1382
19. Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549
20. Teichmann, M., Dumay-Odelot, H., and Fribourg, S. (2012) Structural and functional aspects of winged-helix domains at the core of transcription initiation complexes. *Transcription* **3**, 2-7
21. Regni, C. A., Roush, R. F., Miller, D. J., Nourse, A., Walsh, C. T., and Schulman, B. A. (2009) How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. *Embo J* **28**, 1953-1964
22. Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C., and Ramakrishnan, V. (2009) Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat Struct Mol Biol* **16**, 528-533

6. Biochemical characterisation of MINA53 activity in cells

Contents

6.1. MINA53 and cancer 232

6.1.1. Localisation 233

6.1.2. MINA53 levels in cancer indicate a potential tumour suppressor role 233

6.1.3. A loss-of-function cellular model of MINA53 234

6.1.4. MINA53 and autophagy 235

6.1.5. Biochemical studies on MINA53 reveal rpL27a as a potential substrate 236

6.1.6. Aims 239

6.2. Isolation of ribosomal proteins from mouse livers 239

6.3. Identification of rpL27a in the UPLC chromatogram 240

6.4. MINA53 knock-down reduces rpL27A hydroxylation 243

6.5. Oxygen-dependence of rpL27a hydroxylation 246

6.6. Analysis of rpL27a hydroxylation in tumour samples 249

6.7. Effect of somatic mutations of *MINA53* on hydroxylation of rpL27a 250

6.8. *In vitro* testing of potential non-ribosomal MINA53 substrates 263

6.9. Conclusions 265

6.10. References 267

6.1. MINA53 and cancer

As was discussed in Chapters 1 and 2, the *E. coli* ycfD is a prokaryotic homologue of two human 2OG oxygenases, nucleolar protein 66 (NO66) and myc-induced nuclear antigen 53 (MINA53). MINA53 and NO66 are JmjC-domain containing enzymes that share homology with FIH and the JmjC histone demethylases (Figure 6.1).

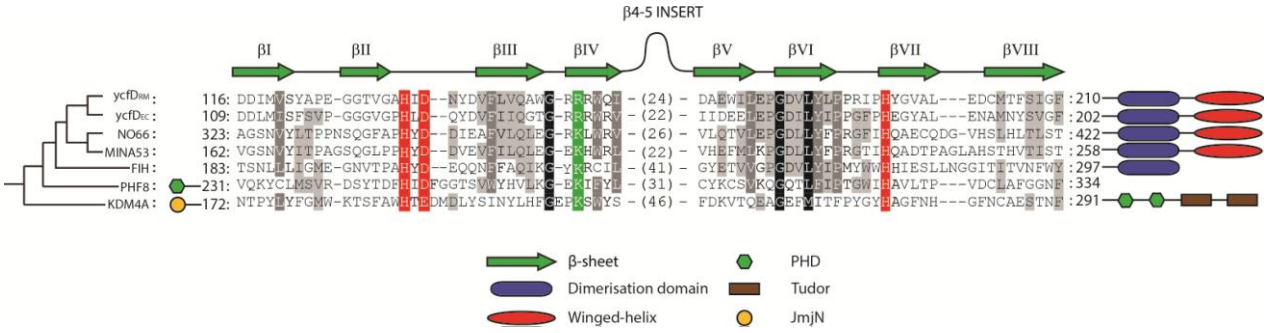


Figure 6.1: The protein alignment of the DSBH domain of human NO66 and MINA53 reveals homology with bacterial ycfD enzymes and human HIF hydroxylase FIH and histone demethylases PHF8 and KDM4A (JMJD2A). A structure based alignment was generated using STRAP alignment software (http://www.bioinformatics.org/strap) and edited in GeneDoc. Secondary structure assignment of the DSBH is indicated with green arrows above the alignment, Roman numerals indicate the identity of individual β-strands of the DSBH. Iron-binding residues are highlighted in red; the basic 2OG binding residue is highlighted in green. *N*- and *C*-terminal protein domains extending from the DSBH are indicated. The comparison of extra domains suggests that ycfD and NO66/MINA53 contain same domains (a DSBH is followed by a dimerisation domain, also present in FIH, and a winged-helix domain, which is absent in FIH), suggesting that they are the closest homologues within the JmjC-domain containing 2OG oxygenases.

The initial identification of MINA53 as a Myc target gene in glioblastomas (1) paved the way for subsequent studies of MINA53 expression in a multitude of different types of cancer. Although MINA53 over-expression was found to correlate with poor prognosis in some carcinomas, it was associated with good prognosis in others, suggesting that its role(s) in cancer are likely highly context-dependent (see below). These findings motivated a search for the biochemical function of MINA53, including its biological substrate, which would provide a rationale for the role of MINA53 in cancer. Importantly, the connection between the role of MINA53 as an oncogene or a tumour suppressor (evidence for both has been presented, see below) and its biochemical activity against the biological substrate(s) could provide potential means for therapeutic intervention in cancer.

6.1.1. MINA53 localisation

Immunofluorescence studies identified MINA53 localisation mainly in the nucleolus, in particular in the granular component, the cellular compartment crucial for the ribosome biogenesis process (1,2). RNAse treatment, RNA pol I inhibition (Actinomycin-D) or serum starvation abolished the nucleolar localisation of MINA53 (1,3), consistent with a role for it in ribosome biogenesis. Immunoprecipitation studies independently identified nucleophosmin and a number of ribosomal proteins as interaction partners of MINA53 (3), and MINA53 was enriched in the pre-ribosomal fraction of *Xenopus* oocytes (3). Together, these data indicated that MINA53 may be involved in a cellular process which involves rRNA transcription, likely ribosome biogenesis.

6.1.2. MINA53 levels in cancer indicate a potential tumour suppressor role

A number of clinical studies have reported elevated levels of MINA53 in different types of carcinoma. In renal cell carcinoma, MINA53 over-expression showed correlation with poor treatment prognosis (4). Similarly, MINA53 was over-expressed in 83% of studied oesophageal squamous cell carcinoma samples, which correlated with shorter survival times (5). Increased levels of MINA53 correlated with poor prognosis in neuroblastoma (6), tumour proliferation of gingival squamous cell carcinoma (7), and increased tumour size in hepatocellular carcinoma (8). A study of gastric carcinoma revealed increased levels of MINA53 protein in 91% of all studied tumours, whereas normal tissues had low levels of MINA53 expression (9). MINA53 was frequently over-expressed in aggressive types of B cell lymphoma (10). Surprisingly, in lung cancer, MINA53 expression levels had the opposite effect on survival – over-expression correlated with good prognosis, whereas down-regulated levels of MINA53 correlated with poorer prognosis (11).

With respect to the positive role of MINA53 in some tumours, targeting MINA53 by si/shRNA resulted in growth suppression in HeLa (cervical carcinoma, (12)), A549 (lung adenocarcinoma, Zhang 2005) and SW620 (colon carcinoma (1)). Furthermore, over-expression of MINA53 caused transformation of NIH/3T3 cells and enhanced tumour growth rate in an allograft model using nude mice (13). In contrast, targeting MINA53 by shRNA in H226B cells (squamous lung cancer, (11)) increased their invasiveness.

6.1.3. A loss-of-function cellular model of MINA53

In order to address functions of *MINA53* in cancer, a collaboration was established with Adam Zayer and Dr. Mat Coleman (Wellcome Trust for Human Genomics, Oxford), who developed an isogenic MINA53 KO model in immortalised K-Ras-transformed mouse embryonic fibroblasts (MEFs).

Adam Zayer and Dr. Mathew Coleman prepared a MEF cell model of *MINA53* KO derived from mouse embryos with homozygous MINA53 deletion. Primary MEFs were obtained from WT and KO mice, immortalised using SV40 Large T antigen and transformed with K-Ras G12V (a common human cancer oncogene (14)). Four genetic *MINA53* variants were prepared: WT mouse MEFs and *MINA53* KO MEFs reconstituted with either i) empty pBabe vector (15) (EV), ii) pBabe expressing WT *MINA53* or iii) pBabe expressing a catalytically inactive H179A variant of MINA53 (Figure 6.2).

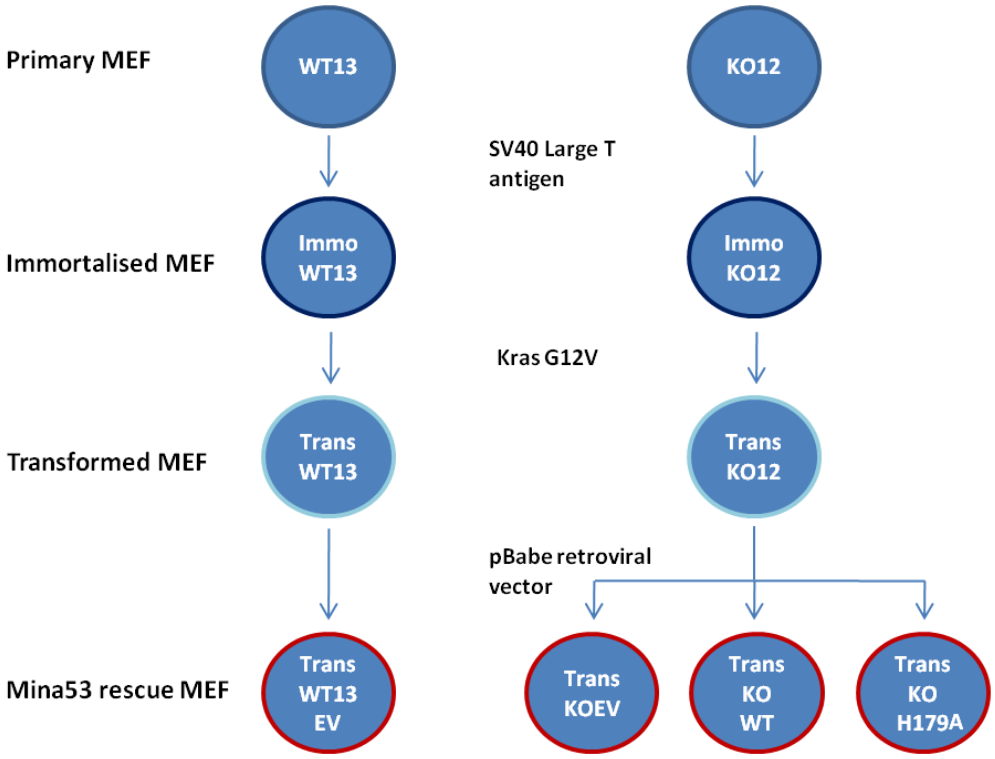


Figure 6.2: Schematic representation of the process used to prepare the WT and MINA53 KO MEF model (figure kindly provided by Adam Zayer, Wellcome Trust for Human Genomics, Oxford).

Using this model, Adam Zayer and Dr. Mathew Coleman demonstrated that, in contrast to some reports in other cell types, loss of MINA53 catalytic activity does not affect two-dimensional growth of cells in this model (data not shown). Interestingly, loss of MINA53 catalytic activity actually

increased colony formation on soft agar in an anchorage-independent growth (three-dimensional) assay, a surrogate for tumorigenic potential of a cell. Furthermore, MINA53 KO MEFs reconstituted with EV or WT MINA53 were used to inject BALC/c nude mice to create an allograft model, which revealed that tumours grown from transformed MEFs lacking MINA53 grew significantly faster than WT MEFs (Figure 6.3). Collectively, these results suggest that in the studied biochemical context, MINA53 catalytic activity appears to suppress tumorigenesis, consistent with a potential role of MINA53 as a tumour suppressor gene.

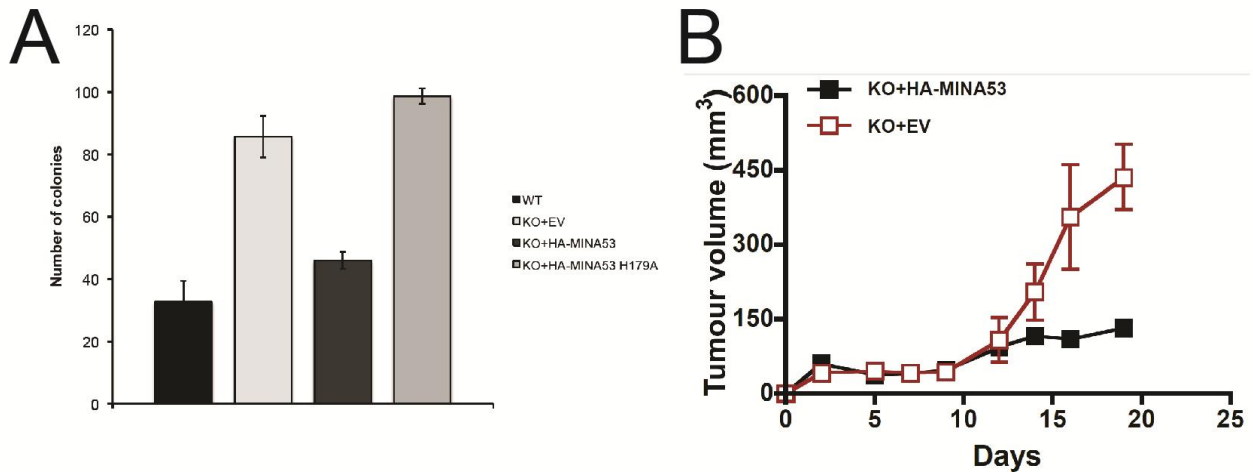


Figure 6.3: MINA53 regulates transformation and tumorigenesis. A) Transformed WT and MINA53 KO MEFs expressing i) empty vector (EV), ii) WT MINA53 or iii) catalytically inactive variant H179A were seeded on soft agar; after 4 weeks the number of colonies was counted by crystal violet staining. B) MINA53 catalytic activity affects the rate of tumour growth in mouse allograft models. BALC/c nude mice were injected with K-Ras transformed MEFs (WT+EV), MINA53 KO expressing empty vector (KO+EV), WT MINA53 (KO+HA-MINA53) or catalytically inactive variant H179A (KO+HA-MINA53 H179A). 7 mice were used for each type of tumour. Tumour growth was monitored over 55 days, 3 times a week. Modified from Adam Zayer's DPhil thesis.

6.1.4. *MINA53* and autophagy

The biochemical MEF model of MINA53 activity depends on a K-Ras G12V cellular transformation step (Figure 6.2). Since cellular transformation appeared to decrease in the clones with catalytically active MINA53 (as measured by anchorage-independent growth assay, Figure 6.3a), and since K-Ras driven transformation and tumorigenesis is dependent on autophagy (16,17), Adam Zayer and Dr Mathew Coleman investigated whether the role of MINA53 activity in K-Ras transformation was associated with altered autophagy. Autophagy is a catabolic process that recycles cellular organelles via lysosomal degradation during nutrient starvation or stress (18). Consistent with the hypothesis that, at least under the conditions tested, MINA53 activity regulates autophagy, they observed that

loss of MINA53 catalytic activity in the MINA53 MEF model upregulates basal autophagy markers (including LC3A/B-II and Lysotracker, Figure 6.4B/C), which were further induced by Rapamycin treatment; Rapamycin inhibits a key cellular nutrient sensing signalling pathway via mTOR kinase inhibition. The apparent regulation of Rapamycin-induced autophagy by MINA53 activity was confirmed in an independent assay that relied on a mitochondrial activity assay (MTS). Cells lacking MINA53 activity showed significant loss of mitochondrial activity (Figure 6.4) due to increased mitochondrial clearance (mitophagy; data not shown) associated with enhanced autophagy. The MTS assay represents an affordable and rapid means of screening for the effects of MINA53 variants on tumour-associated cell processes such as autophagy, and is employed later in this chapter.

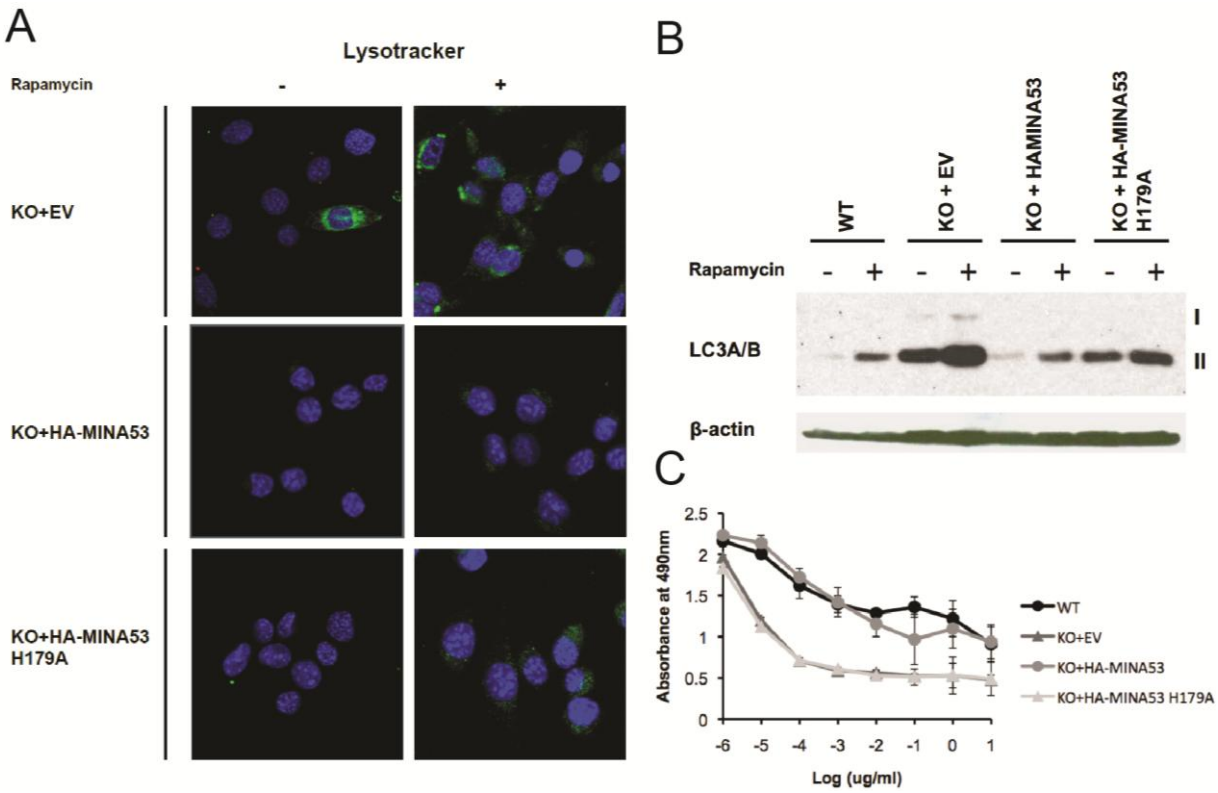


Figure 6.4: Loss of MINA53 catalytic activity upregulates autophagy. A) Confocal images demonstrate increased Lysotracker staining of clones that lack MINA53 catalytic activity, which suggests increased autophagy in response to Rapamycin. B) Treatment of MEF clones with Rapamycin leads to increased autophagy marker LC3 A/B II in cells that lack MINA53 activity. C) Loss of MINA53 catalytic activity induces autophagy, associated with mitochondrial loss in MINA53 KO MEF clones expressing empty vector (EV) or a catalytically inactive MINA53 variant treated with Rapamycin, measured indirectly by mitochondrial MTS assay. Modified from Adam Zayer, DPhil Thesis.

6.1.5. Biochemical studies on MINA53 reveal rpL27a as a potential substrate

The data described above implicates a role for the catalytic activity of MINA53 in the regulation of a cellular cancer-associated processes, such as autophagy. It is unclear at the present whether the role of

MINA53 in this process is direct or indirect; therefore, biochemical identification and characterisation of MINA53 substrates would be invaluable in order to understand the role of *MINA53* in cancer.

MINA53 has been confirmed as a 2OG- and Fe(II)-dependent oxygenase by ^{14}C-radioactive turnover studies which demonstrated that it requires Fe(II), 2OG as well as reducing agent ascorbate for maximum catalytic activity (Nicholas Granatino, DPhil thesis, University of Oxford). It has been suggested that MINA53 may function as a histone demethylase that is apparently specific for H3K9me3 at promoters of rRNA transcription (19). Given the relatively high sequence similarity with canonical JmjC-domain containing histone demethylases, this proposal appeared reasonable. However, several attempts by Schofield and co-workers failed to reproduce the reported results (Dr. Wei Ge and Sophie Williams, personal communications).

Immunoprecipitation studies using doxycycline-inducible FLAG-tagged MINA53 in U2OS osteosarcoma tumour cells identified nuclear protein nucleophosmin, as well as a large number of ribosomal proteins, including rpL27a, rpL3, P0, rpL21 and rpL4 as interaction partners of MINA53 (Nicholas Granatino, DPhil thesis). This interaction profile, together with the nucleolar localisation of MINA53 and its proposed role in ribosomal biogenesis could be consistent with its hydroxylation of a ribosomal protein. In the light of the characterization of ycfD as a ribosomal oxygenase (ROX), and its homology with MINA53, a ribosomal protein substrate appeared plausible.

Using a similar tiled-array approach that had been used for ycfD, a library of overlapping peptide sequences corresponding to fragments of ribosomal proteins identified in the co-immunoprecipitation experiment were prepared (experiment performed by Dr. Wei Ge). A peptide corresponding to rpL27a (residues G33-R42) resulted in +16 Da and -2 Da shifts, when incubated with 2OG, Fe(II), ascorbate and MINA53 (Figure 6.5), suggesting that rpL27a may be a hydroxylation substrate for MINA53. The -2 Da peak was assigned as hydroxylation followed by dehydration. Hydroxylation/dehydration occurred on H39, which was validated using NMR, as well as alanine-scanning experiments (20).

Identification of a candidate substrate provides a potential starting point for elucidation of the biological functions of *MINA53*. However, peptide screening against recombinant enzymes frequently

produces false positive results resulting from favourable interactions between a peptide fragment and the enzyme. Recent findings in from the Schofield lab suggest that many positively-charged peptides derived from ribosomal protein sequences undergo hydroxylation on the *N*-terminus, although there is no evidence that the corresponding protein is modified in cells. The identification of candidate substrates identified *in vitro* therefore requires validation in cells. LC-MS/MS of protease-fragmented proteins purified from cells has been used with success on related projects (e.g. NO66 hydroxylation of rpL8, OGFOD1 hydroxylation of rpS23). However, the sequence context of the rpL27a hydroxylation site renders the proteolytic fragment inappropriate for detailed LC-MS/MS studies, presumably due to the high G/H content of the fragment (Adam Zayer, personal communication). An alternative method for robustly and efficiently studying the activity of MINA53 in cells was therefore required; the mass of rpL27a (~16 kDa) renders it suitable for detection by ESI-TOF mass spectrometers.

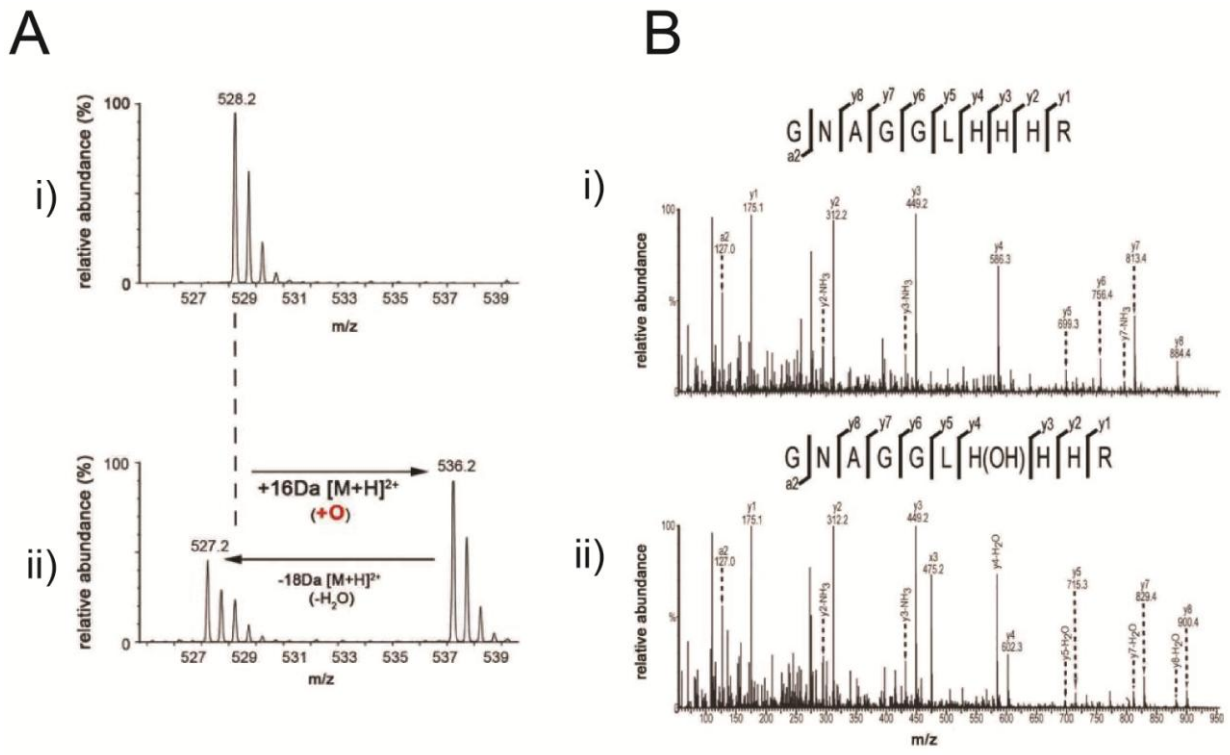


Figure 6.5: MINA53 hydroxylates rpL27a fragment (G33-R42) on H39. A) MALDI spectrum of rpL27a fragment peptide (G33-R42) incubated with Fe(II), 2OG and ascorbate in i) absence and ii) presence of MINA53. In ionization state $[M+H]^{2+}$ the +8 shift indicates a hydroxylation event, whereas the -1 shift indicates a hydroxylation followed by dehydration. B) MS/MS analysis of rpL27a fragment peptide identifies hydroxylation on H39. Peptide fragment shown in A) was analyzed by MS-MS. Reproduced, with permission, from (20).

6.1.6. Aims

Given the role of MINA53 activity in tumorigenesis it was considered important to characterise the biological substrates proposed for MINA53 in order to understand the mechanism involved. In order to establish whether rpL27a is a biologically-relevant substrate of MINA53, it was proposed that intact protein LC-MS methodology could be applied in conjunction with genetic manipulations (performed by Adam Zayer, Wellcome Trust for Human Genomics, Oxford) in human and mouse tissues. The effect of hypoxia on MINA53 activity and rpL27a hydroxylation would be examined. Furthermore, the effect of mutations that occur in tumours on the activity of MINA53 would be studied by measuring rpL27a hydroxylation. The ability of MINA53 cancer mutants to hydroxylate rpL27a would be correlated with biological function in transformation and autophagy assays in order to consider ribosomal protein hydroxylation as a potential driver of MINA53-related pathologies.

6.2. Isolation of ribosomal proteins from mouse livers

Intact protein MS requires considerable amounts of biological material for good signal/noise ratios (the minimum concentration of total ribosomal protein sample required for interpretable MS results was empirically determined at 1 mg/mL). Although actively translating ribosomes represent 30% of the mass of a dividing (bacterial) cell, over 500 mg of cell tissue is required for successful detection of individual ribosomal proteins. Liquid cultures are the preferred approach of culturing biological material for ribosomal analyses due to higher cell growth densities. Bacterial systems, such as *E. coli* and *R. marinus* discussed in Chapters 2-4 are therefore ideally suited for ribosomal studies, because they yield high amounts of material. Since the culture of adherent mammalian cells is usually undertaken in two-dimensions (due to a requirement for matrix-adhesion to support growth), contact inhibition upon cell confluence limits the yield of the resulting biological material relative to investment of both time and material. Therefore, sources of biological material other than tissue culture are often favoured.

The purification and mass spectrometric study of mammalian ribosomal proteins was initially optimized using mouse liver tissue. Livers are the largest mouse organ, often weighing 500-1000 mg

(equivalent to a cultured cell volume produced in >10 confluent 150 cm^2 culture dishes), are easily dissected and can be flash-frozen within 3 min of sacrificing the animal. Samples of WT mouse livers were initially kindly provided by Dr. Tammie Bishop. Subsequent samples were provided by Drs. Matthew Cockman, Patrick Pollard (OGFOD1 project) and Mathew Coleman (MINA53 project).

Ribosomal proteins were extracted from mouse liver extracts using the same procedure as described in Chapter 2. SDS-PAGE gel analysis of the resulting protein pellet demonstrated significant enrichment of proteins in 10-40 kDa range, consistent with ribosomal protein/ribosome associated factor masses (Figure 6.6).

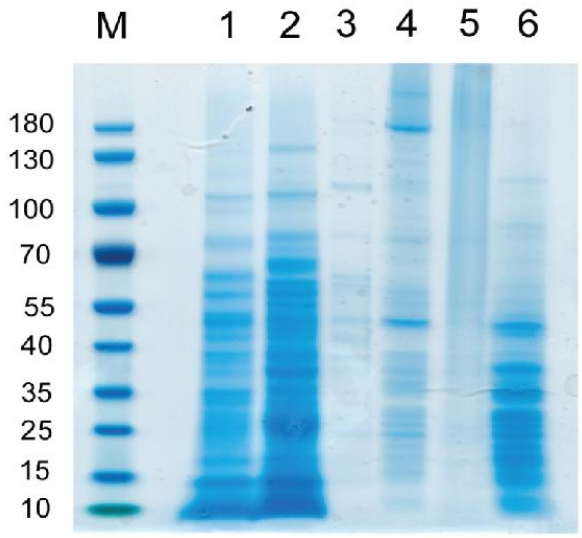


Figure 6.6: SDS-PAGE gel of the ribosomal protein extraction process. Tissue was homogenised in buffer containing Tritonx100 and protease inhibitor cocktail; the lysate was ultra-centrifuged (3 h, 100,000 *g*), ribosomes are isolated from the resulting pellet by purification through sucrose cushion (1.1 M sucrose); ribosomes dissociated by high Mg(II) concentration and rRNA is removed by acetic acid precipitation. Resulting ribosomal proteins are precipitated by acetone and dried under air. M: molecular weight marker; 1: whole-cell extract; 2: supernatant after first 100 k*g* centrifugation; 3: supernatant from centrifugation through 1.1M sucrose cushion; 4: pellet from 100 k*g* centrifugation through 1.1M sucrose cushion; 5: RNA fraction of pelleted ribosomes; 6: protein fraction of pelleted ribosomes showing enrichment of proteins in 10-40 kDa range.

6.3. Identification of rpL27a in the UPLC chromatogram

For LC-MS studies, chromatographic properties of rpL27a were established. RpL27a was first isolated using a HPLC protocol originally established by Dr. Matthew Cockman (Wellcome Trust for Human Genomics, Oxford). A HPLC eluate of a crude ribosomal fraction was analysed by dot-blotting; the fraction that showed a high degree of cross-reactivity with a α-rpL27a antibody was freeze-dried and analysed by UPLC-ESI-MS alongside the total isolated ribosomal proteins from mouse livers using identical chromatographic conditions (Figure 6.7).

UPLC analysis of the HPLC purified rpL27a fraction (Figure 6.7a) suggested the presence of 4 major components in the sample: within experimental error, deconvoluted masses were consistent with predicted masses of rpL35 (14422 Da at 6.8 min; predicted mass 14421 Da), rpL17 (21267 Da at 10.5 min; predicted mass 21266 Da), modified rpL27a (16490 Da at 15 min; predicted mass 16474 Da) and rpS19 (15955 Da at 18.5 min; predicted mass 15954 Da). This indicates that the existing HPLC methodology of purifying ribosomal proteins may suffer from cross-contamination with other ribosomal proteins. The same elution method was then applied to a sample of total mouse liver ribosomal proteins (in 0.1% $HCOOH_{aq}$). A peak at 15 min deconvoluted to a mass of 16490 Da (Figure 6.7), suggesting that this was the elution time corresponding to rpL27a. This was used to identify the mass of rpL27a in all subsequent experiments. RpL27a purified from mouse liver was also subjected to MS/MS analysis which localised the hydroxylation to residue H39 of the tryptic fragment (Figure 6.8, kindly provided by Adam Zayer).

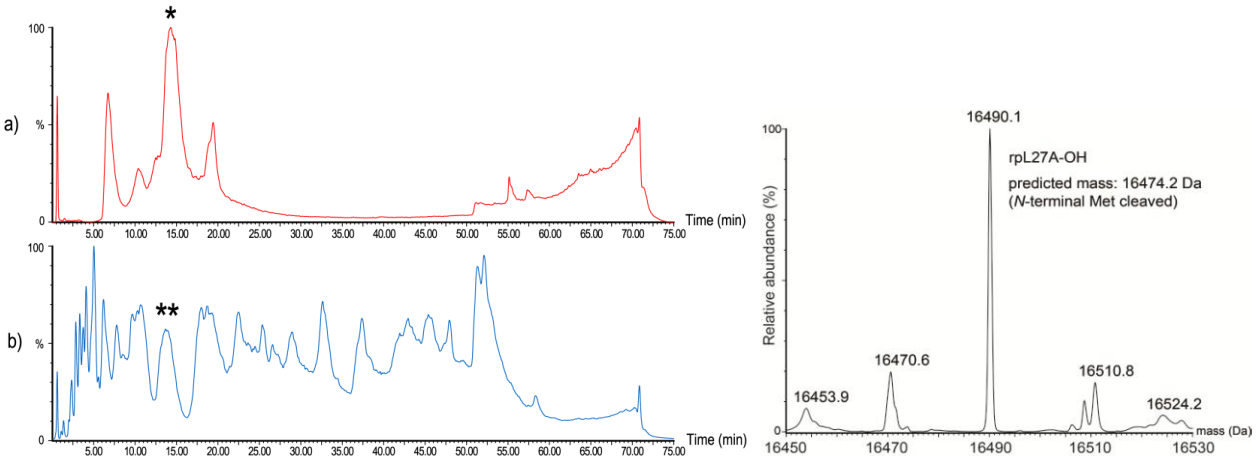


Figure 6.7: UPLC chromatogram of rpL27a-containing samples from *M. musculus* (left). a) Ribosomal protein rpl27A (marked *) from *M. musculus* purified by semi-preparative HPLC and identified by antibody staining (the sample kindly provided by Dr. M. Coleman, Wellcome Trust Centre for Molecular Physiology); deconvolution of four major peaks resulted in mass signals of 14422 Da (6.8 min), 21267 Da (10.5 min), 16490 Da (15 min) and 15955 Da (18.5 min), consistent with predicted masses of rpL35 (predicted 14421 Da), rpL17 (21266 Da), modified rpl27a (predicted mass 16474 Da) and rpS19 (predicted 15954 Da) b) UPLC chromatogram of a total *M. musculus* ribosomal protein fraction. Elution time of rpL27A, identified by intact mass, is marked (**). An identical chromatographic method was employed to obtain the two chromatograms, which demonstrates that mass-based protein identification is comparable to antibody-based identification. **Deconvoluted ESI-MS spectrum of rpL27a from WT mouse liver ribosomes (right).** Predicted mass of rpL27a (*N*-terminal methionine removed) is 16474.2 Da. The major peak at 16490.1 Da is consistent with hydroxylation of rpL27a (+16 Da). Peaks at 16454, 16471, 16511 and 16524 likely correspond to deconvolution artefacts. Mass spectrum acquired using a Synapt ESI-TOF-MS (Waters).

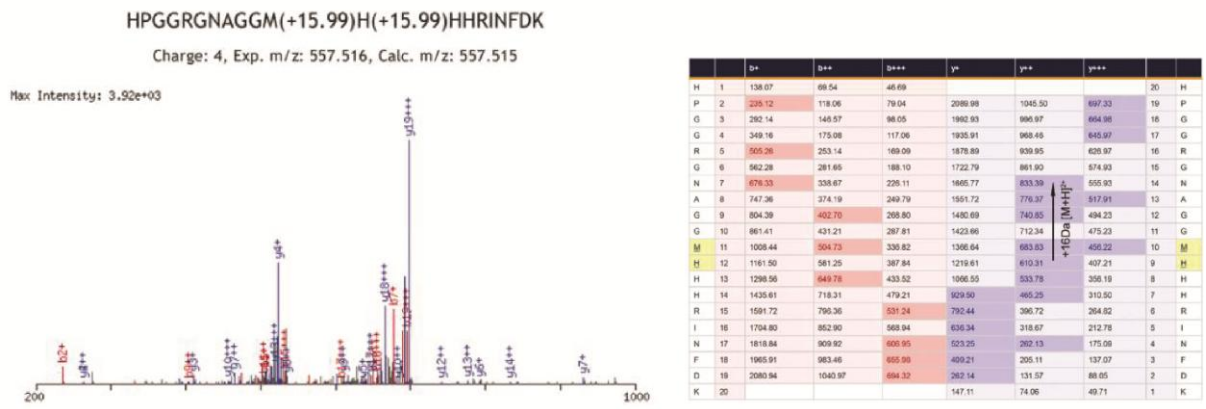


Figure 6.8: Mouse liver rpL27a is hydroxylated on H39. Endogenous rpL27a was purified from mouse liver as described in methods prior to trypsinolysis and MS/MS. The tables summarise the ions detected from fragmented H39-hydroxylated rpL27a tryptic peptide. Note the mass shift of 8 m/z in the doubly-charged y-ion (y++) series from y9 onwards, indicating hydroxylation of H39. Figure kindly provided by Adam Zayer.

To investigate whether rpL27a hydroxylation is widespread, the analysis was then extended to other mouse tissues, including kidney, brain, lungs and testes. Due to small amount of material (100-200 mg tissue), the MS signal/noise ratio was markedly lower compared to liver for most of the organs analysed; within experimental error (±3Da) the mass corresponding to hydroxy-rpL27a (16493 Da) was detected in all samples (liver, kidney, brain, lungs and testes) at elution time 15±1 min, suggesting that rpL27a hydroxylation is widespread (Figure 6.1).

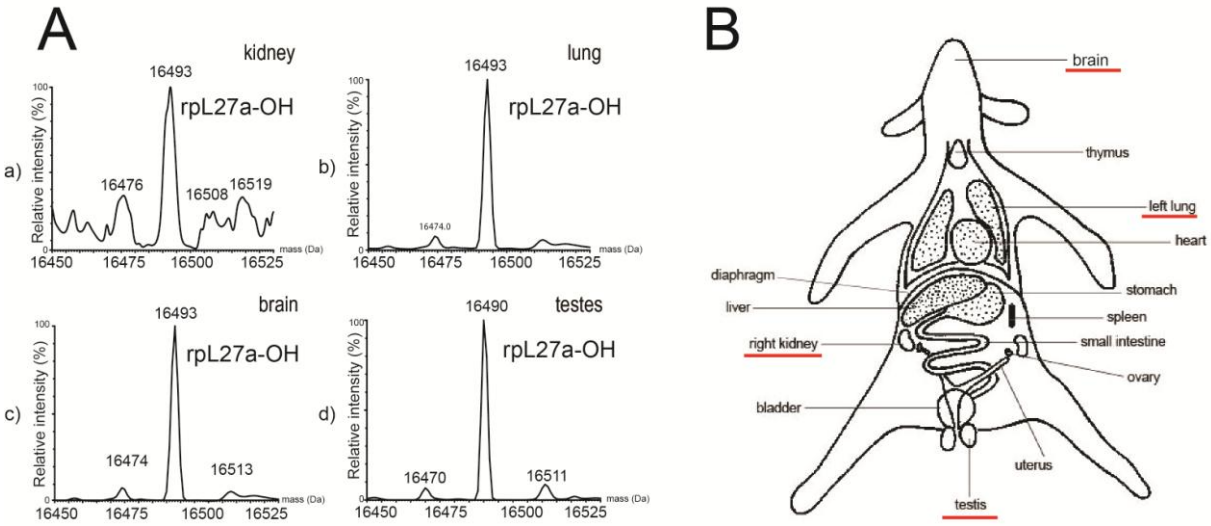


Figure 6.9: Deconvoluted ESI-MS spectra of rpL27a from mouse organs reveal complete hydroxylation in all organs analysed: a) kidney, b) lung, c) brain and d) testes. Predicted mass of rpL27a is 16474 Da. Due to calibration error all masses are shifted +2 Da; sample of testes was analysed separately under calibration which resulted in predicted mass 16490 Da. The peaks corresponding to rpL27a were identified based on the known elution time of the protein (15 min) and compared to masses of other ribosomal proteins, which revealed the mass shift. Due to the limited amount of material available, the samples could not be re-analysed. B) A schematic representation of mouse anatomy (adapted from http://en.wikibooks.org/wiki/).

RpL27a hydroxylation was also detected in human cell lines. Initially, human embryonic kidney (HEK) 293T cells were used to isolate ribosomes and determine the mass of rpL27a. Within experimental error (±2 Da), the mass of 16448 Da corresponded to rpL27a with *N*-terminal methionine removed (predicted 16430.1 Da). RpL27a was independently subjected to MS/MS analysis which identified hydroxylation on H39 (Figure 6.10b, kindly provided by Adam Zayer). RpL27a was also found in hydroxylated form in rat kidney samples (kindly provided by Dr. Mark Cole). Together, the results suggested that rpL27a hydroxylation is conserved in mammals.

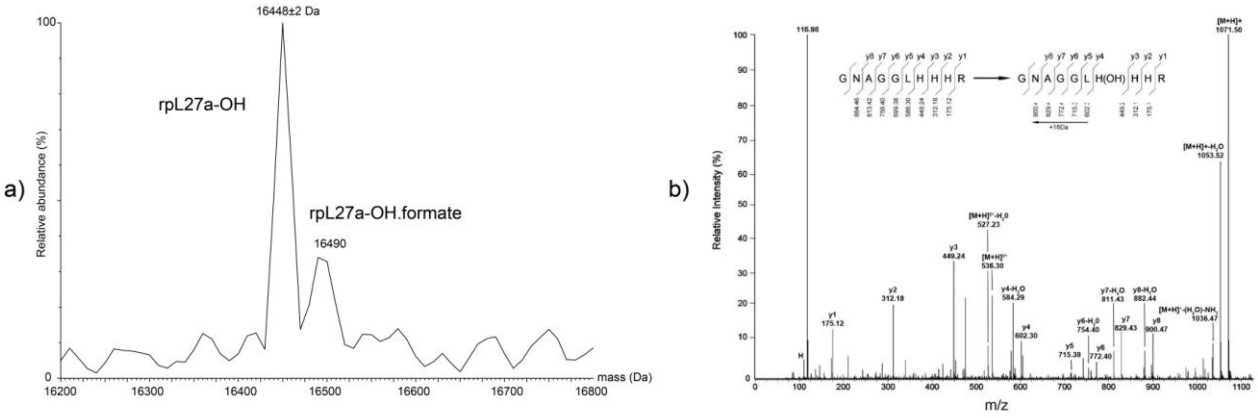


Figure 6.10: Endogenous rpL27a from HEK293T cells is hydroxylated on H39. a) Deconvoluted ESI-MS spectrum of rpL27a demonstrates that rpL27a from HEK293T cells is hydroxylated. Predicted mass (*N*-terminal methionine removed) is 16430.1 Da. Within experimental error (± 2Da) the mass of rpL27a corresponds to hydroxylation. b) MS/MS spectrum of endogenous rpL27a purified from HEK293T cells prior to trypsinolysis and MS/MS sequencing (kindly provided by Adam Zayer, Wellcome Trust for Human Genomics, Oxford). The +16 Da shift in the y-ion series from y4 onwards indicates hydroxylation of H39. The unhydroxylated peptide was not detected (consistent with the whole protein MS data).

6.4. *MINA53* knock-down reduces rpL27A hydroxylation

In order to test the dependence of rpL27a hydroxylation on MINA53, HEK293T cells were subject to shRNA knock-down of *MINA53* by Adam Zayer (Wellcome Trust for Human Genomics, Oxford). Stable shRNA knockdown achieved >95% suppression of MINA53 protein expression, as measured by immunoblotting using a α-MINA53 antibody. ESI-MS analysis suggested that MINA53 shRNA reduced rpL27a hydroxylation from >95% to 55% (Figure 6.11); rpL27a protein levels were not affected by MINA53 knock-down, as demonstrated by a Western blot using an α-rpL27a antibody. However, the deconvoluted MS spectra were noisy and difficult to analyse due to a low amount of biological material available for analysis, which led to relatively low signal/noise ratio during the LC/MS analysis.

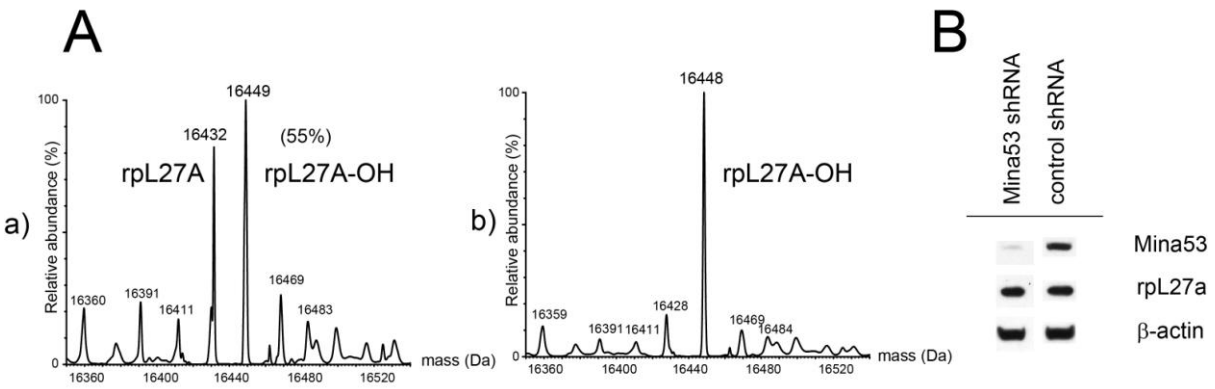


Figure 6.11: A) Deconvoluted ESI-MS spectra of rpL27a from HeLa cells transfected with a) MINA53 shRNA, b) control shRNA. B) Western blots (performed by Adam Zayer) showing that MINA53 protein levels have been depleted >95%, whereas the levels of rpL27a and β-actin loading control remained constant. Predicted mass of rpL27a (*N*-terminal methionine cleaved) is 16430 Da. The peaks at 16448/9 Da were assigned as hydroxylated rpL27a; peaks at 16432 Da were assigned as unhydroxylated rpL27a. Approximately 45% of rpL27a in shMINA53 sample was unhydroxylated. Mass spectra acquired on Synapt ESI-TOF-MS (Waters) without internal calibration.

Therefore, shRNA-mediated MINA53 knock-down was repeated in an independent cellular background in order to independently validate the HEK293T results. ShRNA knockdown of MINA53 in lung adenocarcinoma cells resulted in undetectable levels of MINA53 protein expression. RpL27a from the control shRNA sample revealed a mass of 16446.1 Da, consistent with hydroxylated rpL27a (predicted mass of NTM cleaved rpL27a from *H. sapiens* is 16430.1 Da), whereas deconvoluted MS spectrum of the knock-down sample only showed a strong signal at 16431.5 Da, consistent with predicted mass of unmodified rpL27a (NTM removed); two peaks were observed at 16450 and 16456 Da at 10% intensity relative to the dominant peak at 16431.5 Da, which likely corresponded to m/z deconvolution artefacts (Figure 6.12). Again, MINA53 knock-down did not affect rpL27a protein levels. Together, the data suggested that rpL27a hydroxylation in human cells requires MINA53.

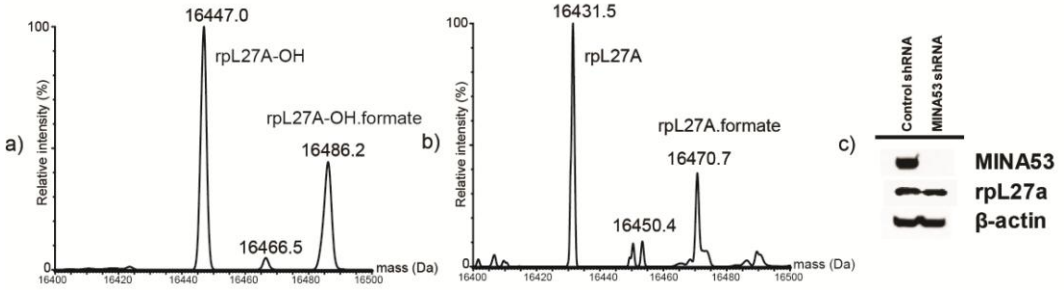


Figure 6.12: Deconvoluted ESI-MS spectrum of rpL27a isolated and characterized by UPLC-MS from A549 cells transfected with a) control shRNA, b) MINA53 shRNA. c) Western blots (performed by Adam Zayer) showing that MINA53 protein levels have been depleted >95%, whereas the levels of rpL27a and β-actin loading control remained constant. The peak at 16447 Da corresponds to hydroxylated rpL27a (predicted mass of *N*-terminal methionine-cleaved rpL27a is 16430 Da), the peak at 16486 Da corresponds to a formyl (+41 Da) adduct of hydroxylated rpL27a. The peak at 16431 Da corresponds to unhydroxylated rpL27a (NTM removed) and the peak at 16470 Da corresponds to a formyl adduct of unhydroxylated rpL27a. Mass spectra acquired on LCT-Premier ESI-TOF-MS (Waters) in Lock-Spray mode using leucine enkephalin for on-line mass calibration.

To determine whether the requirement for MINA53 is dependent on the catalytic activity of the enzyme, the hydroxylation of rpL27a in the *MINA53* MEF model was next analysed (samples prepared by Adam Zayer and Dr. Mathew Coleman). LC-MS analysis revealed that WT MEFs contained a major species at 16491 Da at an elution time consistent with rpL27a, confirming the presence of hydroxylated rpL27a (Figure 6.13a). Minor peaks observed at ~15% intensity relative to major peaks at 16453, 16471, 16511, and 16528 Da also occurred at deconvolution of a number of other proteins from the sample, strongly suggesting that they are deconvolution artefacts. KO MEFs transfected with pBabe EV contained a 16475 Da signal eluting at the same elution time as the 16491 Da species from the previous WT sample; within experimental error this mass corresponds to unhydroxylated rpL27a (NTM removed; predicted 16474 Da). Transfection of KO MEFs with human influenza hemagglutinin (HA)-tagged WT MINA53 resulted in recovery of the 16491 Da peak, and no signal at 16475 Da, consistent with reconstituted MINA53 activity. In contrast, transfection of KO MEFs with the H179A variant of MINA53, predicted to abrogate Fe(II)-binding without disrupting the tertiary fold or expression of MINA53, resulted in a dominant species of 16475 Da, the mass of unhydroxylated rpL27a. Therefore, the hydroxylation of endogenous rpL27a is dependent on both the expression and catalytic activity of MINA53.

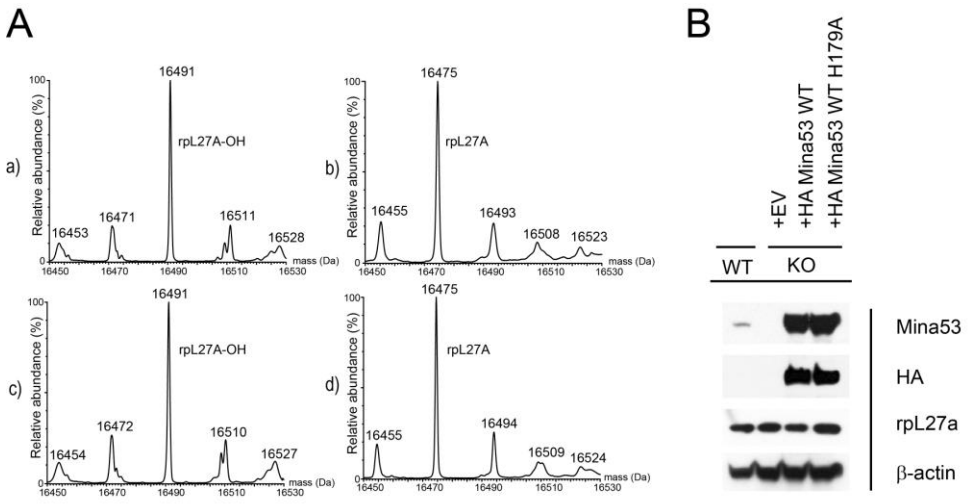


Figure 6.13: A) Deconvoluted ESI-MS spectra of rpL27a isolated from a) WT mouse embryonic fibroblasts (MEFs), MINA53 KO MEFs transfected with b) EV, c) WT MINA53, d) MINA53 variant H179A. The mass at 16491 Da corresponds to hydroxylated rpL27a, sub-peaks spaced at 20 Da intervals are likely harmonics of dominant mass, resulting from deconvolution affected by noise in the spectrum. Peak at 16475 Da corresponds to unhydroxylated rpL27a (predicted mass 16474 Da, NTM cleaved). Same sub-peak pattern is observed in samples b) and d), shifted relative to samples a/c) by 16 Da. Mass spectra acquired on Synapt ESI-TOF-MS (Waters) without internal calibration. **B) Western blot showing over-expression of HA-tagged MINA53 in samples c) and d).**

Considering the significant protein sequence homology between MINA53 and NO66 the potential redundancy of these two enzymes with respect to rpL27a hydroxylation was explored next. ESI-MS spectra of ribosomal proteins purified from HeLa cells stably-transfected with either control or NO66 shRNA demonstrated no effect on rpL27a hydroxylation, despite efficient NO66 knockdown (Figure 6.14).

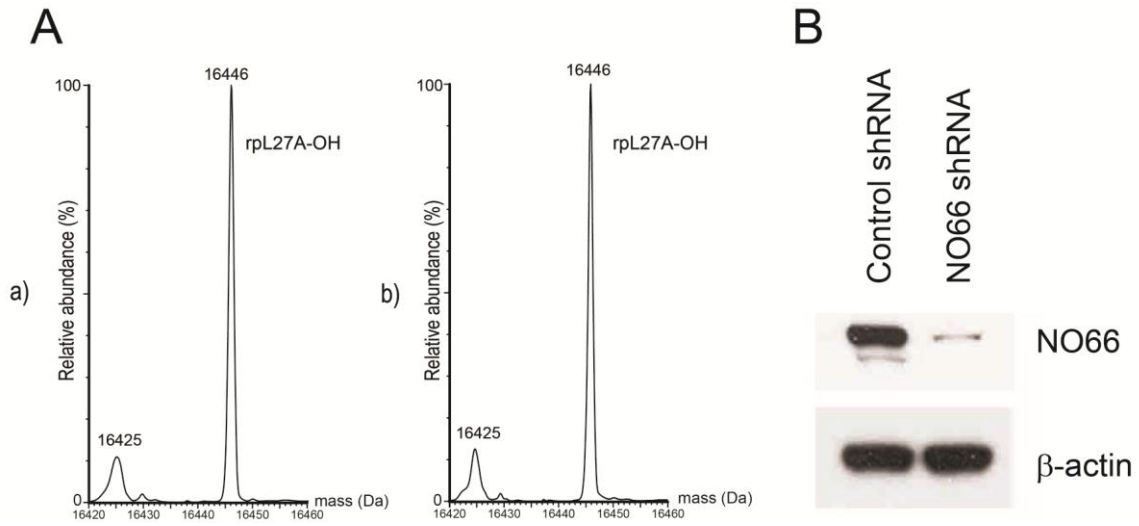


Figure 6.14: Deconvoluted ESI-MS spectra of rpL27a from cells treated with NO66 shRNA HeLa (NO66) cells (transfection and cell culture performed by Penny Feng). HeLa cells were treated with a) control shRNA or b) NO66 shRNA. Predicted mass of rpL27a (*N*-terminal methionine cleaved) is 16430.1 Da. The major peak at 16446 Da in sample b) demonstrates that treatment of HeLa cells with NO66 shRNA has no effect on the mass of rpL27a suggesting that rpL27a is not a hydroxylation substrate of NO66. Mass spectra acquired on LCT-Premier ESI-TOF-MS (Waters) in Lock-Spray mode using leucine enkephalin for on-line correction of mass. **B) Western blot demonstrating efficient NO66 knockdown by shRNA**

6.5. Oxygen-dependence of rpL27a hydroxylation

In eukaryotes, two sub-families of 2OG- and Fe(II)-dependent oxygenases were proposed to serve as oxygen sensors. PHD2 from the prolyl hydroxylase subfamily serves as a sensor of subtle changes of oxygen levels due to its relatively high K_m^{app} for O_2 (21); on the other hand, the activity of FIH (a JmjC-domain containing oxygenase and an evolutionary relative of MINA53), is only inhibited by severe hypoxia due to its relatively low K_m^{app} for O_2 (22). In Chapter 4, a potential role of ycfD$_{RM}$, a prokaryotic homologue of MINA53 as an oxygen sensor was described. Its sensor role was apparent through decreasing levels of L16$_{RM}$ hydroxylation with decreasing availability of dissolved O_2 at higher temperatures. In Chapter 2, the dependence of L16$_{EC}$ hydroxylation by ycfD$_{EC}$ on atmospheric O_2 concentration was presented; L16-OH increased linearly with increasing atmospheric O_2 in which

E. coli cultures were grown. The experiment was feasible because *E. coli* is a facultative anaerobe. In human cells, analogous experiments are technically challenging due to the scale of material required for MS detection of individual ribosomal proteins and the severe effects of hypoxia on protein synthesis, cell growth, and survival. Due to the relatively long half-life of mammalian ribosomes (3-5 days) together with reduced ribosome biogenesis and translation upon hypoxia (23), the level of incorporation of unhydroxylated ribosomal proteins is expected to be low.

The Ras-transformed MEFs used in this study are more resistant to severe hypoxia compared to other immortalised cell lines (Adam Zayer, personal communication), and were therefore selected as a suitable model in which to study the effects of O_2 on rpL27a hydroxylation. The MEFs were grown at 0.1% O_2 for 3 days before harvesting ribosomes as previously described. Although a small amount of ribosomal material led to low signal intensity during the MS analysis, analysis of the rpL27a LC-MS data revealed the presence of a minor peak at 16473 Da, 17 Da less than the major peak at 16490 Da (Figure 6.15). The 16473 Da peak likely corresponds to unhydroxylated rpL27a, indicating that severe chronic hypoxia reduced rpL27a hydroxylation by 20%. This effect is relatively small compared to analogous ROX systems in bacteria (see Chapters 2 and 4). Nonetheless, the increase in the unhydroxylated rpL27a peak was larger than for rpS23 (the substrate of an unrelated 2OG oxygenase OGFOD1; Dr. Matthew Cockman, personal communication), which showed a ~5% increase in unhydroxylated protein measured in the same biological sample, data not shown). It is possible that MINA53 activity is less substantially inhibited under the hypoxic conditions than in that of the rpS23 hydroxylase OGFOD1. Interestingly, OGFOD1 is evolutionarily closely related to hypoxia sensor PHD2, which is substantially more sensitive to decreasing oxygen levels compared to FIH, which is evolutionarily related to MINA53. The observation that rpL27a hydroxylation is less inhibited by hypoxia than rpS23 hydroxylation is therefore in keeping with their evolutionary history. Furthermore, it cannot be excluded that more than 20% of unhydroxylated rpL27a could be produced during exposure to anoxia, without incorporation into ribosomes. Alternatively, MINA53 may have a very high affinity for O_2 and not function as a classical oxygen sensor. Hypoxia (1% O_2 for 24 h) also reduced the hydroxylation of the NO66 substrate rpL8, resulting in an approximately 30% increase in

unhydroxylated rpL8 (Penny Feng, personal communication). This greater reduction in substrate hydroxylation from a more moderate hypoxia treatment, suggests that NO66 activity may be slightly more sensitive to hypoxia than MINA53.

Next, the effect chronic hypoxia on rpL27a hydroxylation was studied in an *in vivo* model. In collaboration with Dr. Tammie Bishop, rpL27a hydroxylation was investigated in mice exposed to prolonged mild hypoxia (7 days, 10% O_2) or normoxia. Experimental procedures used for mouse tissue analysis passed ethical review by the Medical Sciences divisional Local Ethical Review panel (Oxford University) and were performed under UK Home Office regulations (Animal Scientific Procedures Act 1986). Ribosomes were isolated from mouse livers and analysed by LC-MS. Comparison of rpL27a from mice exposed to prolonged mild hypoxia and rpL27a from the control sample revealed that there was no difference in hydroxylation levels (Figure 6.15). Due to ethical considerations, 10% O_2 is the lowest permitted concentration of O_2 in which mice can be conditioned and it was therefore not possible to analyse the effect of more severe hypoxia (0.1% O_2).

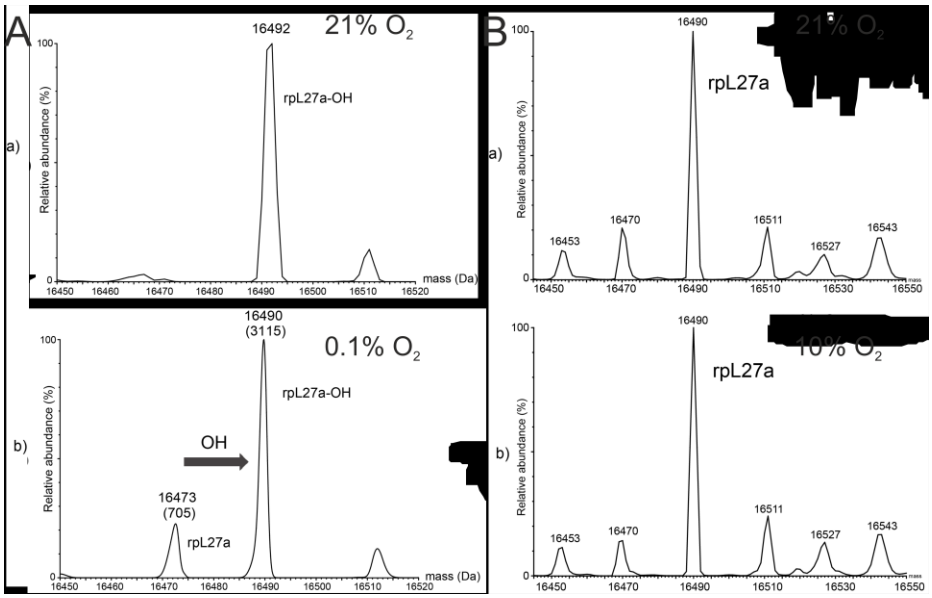


Figure 6.15: A) Deconvoluted ESI-MS spectra of rpL27a from mouse embryonic fibroblast cells. Incubation at a) 21% O_2 or b) 0.1% O_2 for 3 days. Cells were harvested by trypsinolysis and lysed by Dounce homogenisation before performing an ultracentrifugation step to isolate ribosomal proteins. The peaks at 16490/2 Da correspond to hydroxylated rpL27a; 16473 Da is consistent with unhydroxylated rpL27a. Reconstituted intensities of MS peaks are indicated in brackets. The intensity of unhydroxylated rpL27a is 18% of total rpL27a present. Hydroxylation of rpL27a in normoxic MEF is > 95%. Mass spectra recorded on the LCT Premier ESI-TOF-MS spectrometer. **B) Deconvoluted ESI-MS spectra of rpL27a from mouse livers of animals exposed hypoxia.** a) 21% O_2 or b) 10% O_2 for 7 days. The peak at 16490 Da corresponds to hydroxylated rpL27a; subpeaks occurring at ±20 and ±37 Da are deconvolution artefacts. Mass spectra recorded on the Synapt ESI-TOF-MS spectrometer.

6.6. Analysis of rpL27a hydroxylation in tumour samples

The analysis to date suggests that rpL27a is fully hydroxylated in all normal cell lines and tissues, and only modestly reduced by severe chronic hypoxia *in vitro*. However, some pathological conditions are associated with chronic and severe hypoxia and anoxia, particularly the microenvironment of solid tumours. Therefore, to assess rpL27a hydroxylation in severely hypoxic human 'tissue', clinical samples of non-Hodgkin lymphoma (labelled 6E5 and 6E63) were obtained from the John Radcliffe hospital. Due to ethical restrictions on the use of human tissue, it was necessary to use placental tissue as a control 'normal tissue'. All tissue was flash-frozen in $N_{2(l)}$ and lysed in the presence of a pan-2OG oxygenase inhibitor NOG (1 mM in lysis buffer) to prevent hydroxylation post-lysis. Ribosomal protein masses were determined using LCT Premier ESI-MS to 0.3 Da accuracy, using leucine enkephalin as the internal standard for on-line calibration. Surprisingly, LC-MS analysis revealed a major peak at 16446±0.3 Da in all three samples; no signal was observed at 16430 Da, suggesting that rpL27a is >95% hydroxylated in placental tissue as well as non-Hodgkin lymphoma (Figure 6.16).

These data support the *in vitro* analysis indicating that although MINA53 activity and rpL27a hydroxylation are clearly oxygen-dependent; neither is substantially affected by short-term hypoxia, likely due to a high affinity of this enzyme for O_2, similar to FIH. Therefore, MINA53 is unlikely to be a meaningful 'oxygen sensor' akin to the HIF PHDs.

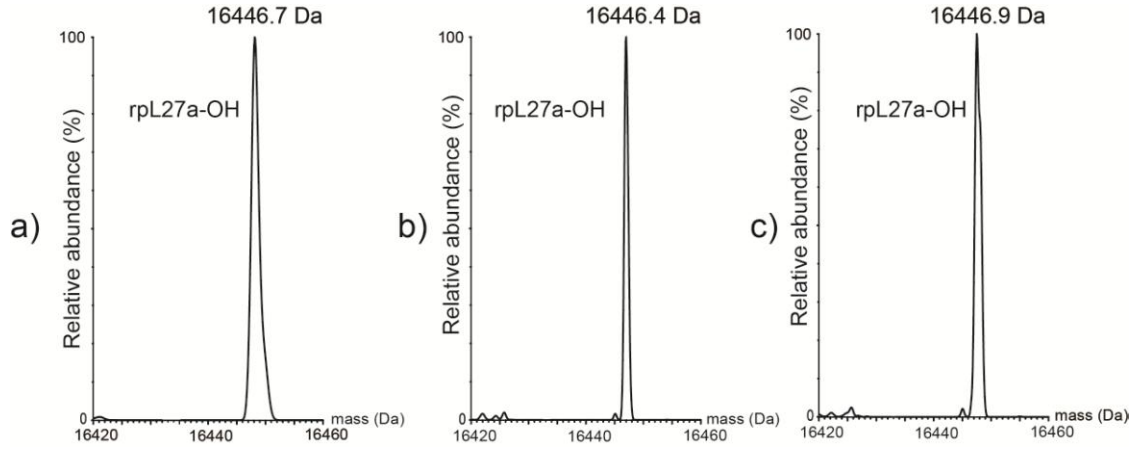


Figure 6.16: Deconvoluted ESI-MS spectra of rpL27a from a) human placenta tissue (provided by Dr. Wei Ge), b) non-Hodgkin lymphoma (sample 6E5) and c) non-Hodgkin lymphoma (sample 6E63). The mass of rpL27 Da was determined by manual component analysis of m/z data, which suggested a maximum margin of error at 0.3 Da. The mass of 16446.7 Da corresponds to hydroxylated rpL27a (NTM removed). No difference was observed between placenta and lymphoma, suggesting that modification of rpL27a is present in highly hypoxic tumour tissue. Mass spectra recorded on an LCT Premier ESI-TOF-MS spectrometer.

6.7. Effect of somatic mutations of *MINA53* on hydroxylation of rpL27a

The functional characterisation of the *MINA53* knockout MEF model suggested a tumour suppressor role for MINA53 in human tumours (see section 6.1.2). Tumour suppressor genes are often deleted, silenced or mutationally inactivated in cancer. Therefore, our collaborators considered that MINA53 may be subject to similar alterations in tumours, and that these could be associated with reduced rpL27a hydroxylation.

With the advent of affordable genome sequencing, genomes of cancer cells are becoming widely available. New cancer genomes are regularly sequenced and available for analysis on databases, such as the catalogue of somatic mutations in cancer (COSMIC), curated by the Wellcome Sanger Institute, Cambridge, UK. Our collaborators interrogated the COSMIC database and identified several point mutations in a variety of tumour types. Although the overall mutation frequency was just 0.5%, the mutations are enriched in specific tumours such as gastric cancer (4%), endometrial cancer (2%, data kindly provided by Jerry Zak and Dr. Mathew Coleman, Wellcome Trust for Human Genomics). Consistent with it being a tumour suppressor, 8% of the curated MINA53 mutations were 'nonsense' mutations (mutations that introduce a premature stop codon). Interestingly, although missense mutations are present throughout the MINA53 coding sequence, they tend to cluster in specific functional domains such as the catalytic and dimerisation domains. Together with the observation that specific residues (e.g. R92, R197 and G289, Figure 6.17) are mutated in multiple different tumour types, this strongly argues for a functional consequence of tumour-associated mutations of *MINA53*.

In order to test whether cancer-associated *MINA53* mutations lead to loss-of-function the *MINA53* KO MEF model (described above) was used to introduce MINA53 variants containing cancer mutations. Using the intact-protein MS methodology described above, the hydroxylation activity of MINA53 variants against rpL27a was measured. In the early phase of this project only a limited number of mutations were described in COSMIC, which were analysed first and are described here. A preliminary structural analysis suggested that mutations occur in functionally important regions of MINA53 (Figure 6.18).

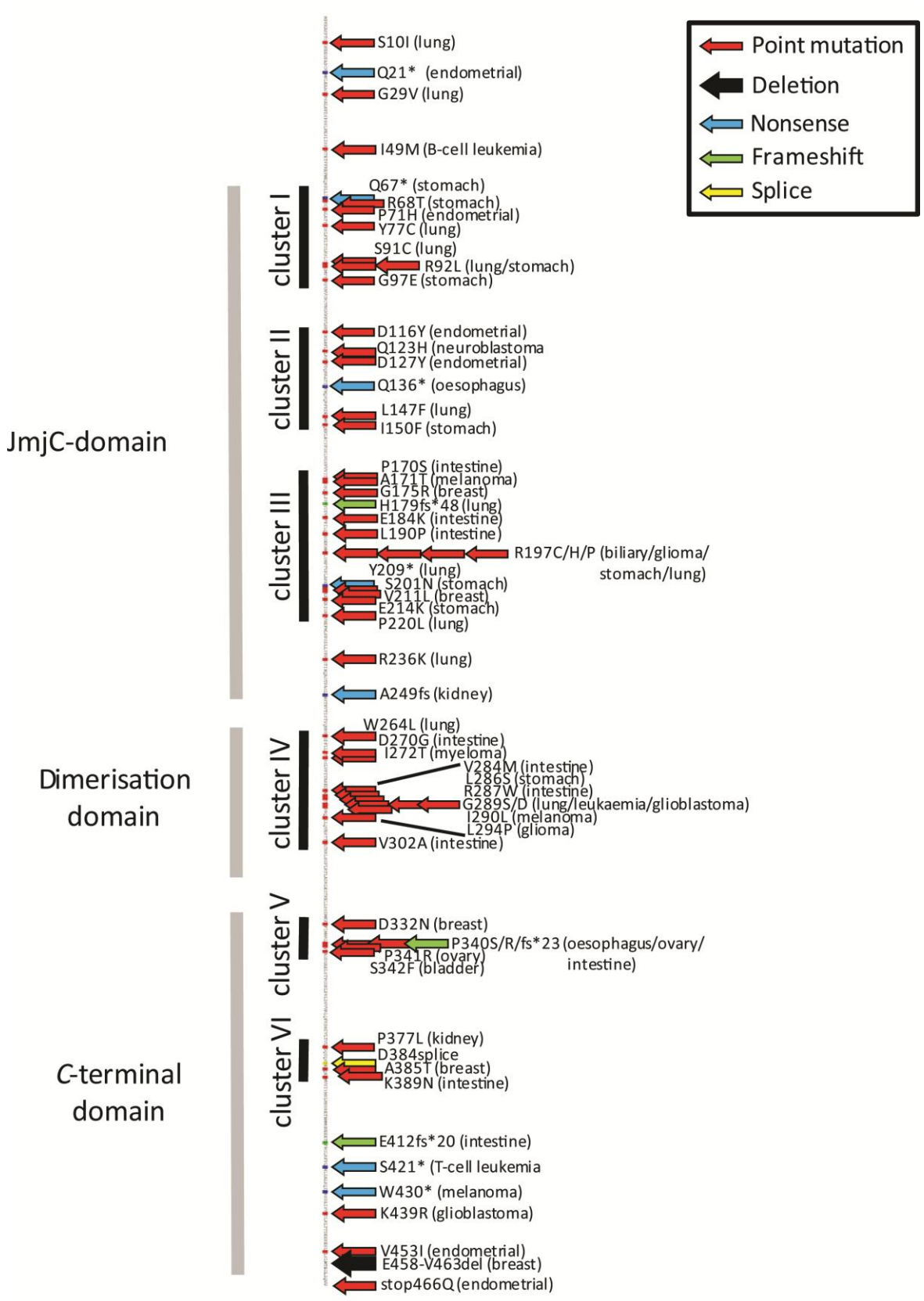


Figure 6.17: Complete list of MINA53 mutations deposited on COSMIC and TCGA databases by 12.1.2014. Identified mutations are superimposed onto the primary sequence of MINA53; component domains are indicated. 6 clusters with increased frequency of mutations are indicated by roman numerals. Each mutation identified is indicated with a single arrow; arrows are colour-coded according to the type of mutation. Original tissue in which the mutation was identified is indicated in brackets. Figure kindly provided by Dr Mathew Coleman.

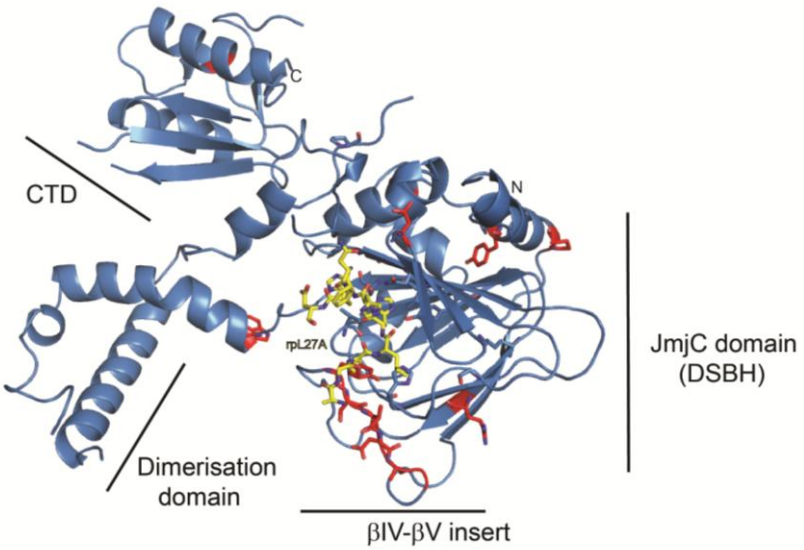


Figure 6.18: A model of crystal structure of MINA53 (blue) with substrate fragment of rpL27a (yellow) covalently attached via a chemical cross-link (courtesy of Dr. Rasheduzzaman Chowdhury). Locations of clinical mutations are indicated in red.

As further mutations were curated from newly sequenced tumours, these were also added to the analysis and are described below. The mutations analysed in the initial phase included: Y209*, H179fs48*, P341R and G175R, W264L. These mutants were reconstituted into *MINA53* KO MEFs. Unsurprisingly, variants that contained premature stop codons (Y209*) or frame-shifts (H179fs48*) resulted in loss of soluble MINA53, as measured by Western blotting, suggesting defects in gene transcription/translation, whereas mis-sense mutations resulted in soluble protein (Figure 6.19). As expected, levels of exogenous MINA53 were markedly higher than in WT MEFs and no MINA53 protein was detected in KO MEFs expressing empty vector pBabe (Figure 6.19).

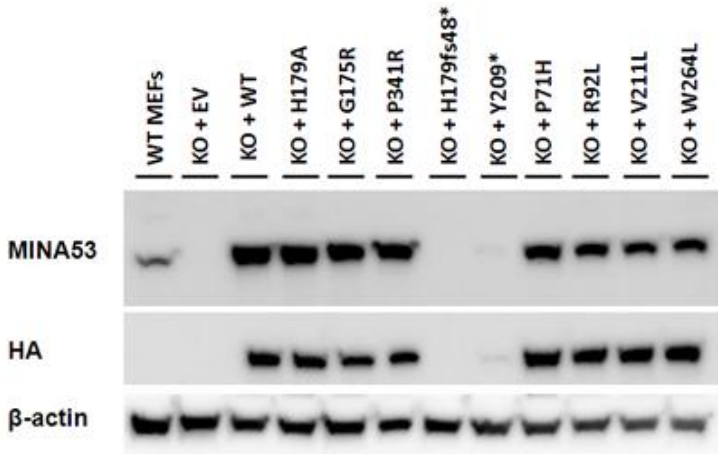


Figure 6.19: Western blot analysis of MINA53 variants containing mutations identified in the COSMIC database demonstrates that only nonsense mutations (premature stop codons, e.g. Y209*, or frame-shifts, e.g. H179fs48*) lead to insoluble protein. MINA53 protein levels in MINA53 KO MEFs over-expressing HA-tagged MINA53 variants are markedly increased relative to endogenous levels of MINA53 in WT MEFs. B-actin was used as a loading control. Figure kindly provided by Adam Zayer.

Initially, WT MEFs and *MINA53* KO MEFs re-expressing either WT MINA53 or EV were tested (Figure 6.20A). As observed before, WT MEFs and KO MEFs expressing WT MINA53 resulted in peaks corresponding to hydroxylated rpL27a (16490 Da, Figure 6.20A i, iii). KO MEFs expressing EV only contained a peak at 16474 Da, 16 Da lower compared to WT, consistent with unhydroxylated rpL27a (Figure 6.20 ii). As further controls, active site variants predicted to abrogate binding of Fe(II) or 2OG were tested. As expected, variants of the HxD Fe(II)-binding triad (H179A and D181A) resulted in complete ablation of rpL27a hydroxylation in cells (Figure 6.20B i, ii). Interestingly, cells expressing a variant of the basic 2OG-binding residue (K194A) contained 20% hydroxylated rpL27a (Figure 6.20B), suggesting that MINA53 can partly retain its catalytic activity in the absence of binding the C5-carboxylate of 2OG. Together, these results validate the MINA53 KO MEF model as a model for studying MINA53 mutation *in vivo*.

WPMS analysis of ribosomes purified from MEFs reconstituted with the cancer mutants G175R, H179fs, Y209stop and P341R mutations revealed that these mutations either reduced MINA53 activity or completely ablated it (Figure 6.20 C/D). For example, MINA53 G175R protein was soluble (Figure 6.19), but completely inactivated (Figure 6.20Ci), suggesting that the variant interferes with the catalytic mechanism of rpL27a hydroxylation. No soluble protein was detected for the H179fs variant (Figure 6.19), suggesting defects in gene transcription/translation and providing an explanation for the observed lack of rpL27a hydroxylation (Figure 6.20Cii). For Y209* (premature stop codon), an ambiguous mass signal at 16492 Da at ~10% intensity was observed likely corresponds to hydroxylated rpL27a (Figure 6.20Ciii). However, western blot analysis surprisingly revealed a faint band corresponding to full-length soluble MINA53 (Figure 6.19), which was attributed to possible contamination of the sample or cell line. Indeed, an independent Y209* cell line displayed no MINA53 expression and no observable rpL27a hydroxylation (data not shown). Interestingly, the P341R mutation, located in the *C*-terminal domain, also led to a reduction in catalytic activity (only 20% of hydroxylated rpL27a was observed, Figure 6.20Di). Given that residue P341 was independently mutated in three types of cancer (P341R, P341P (silent) and P341S, COSMIC database), mutations at that position may promote the tumorigenic potential of a cancer cell.

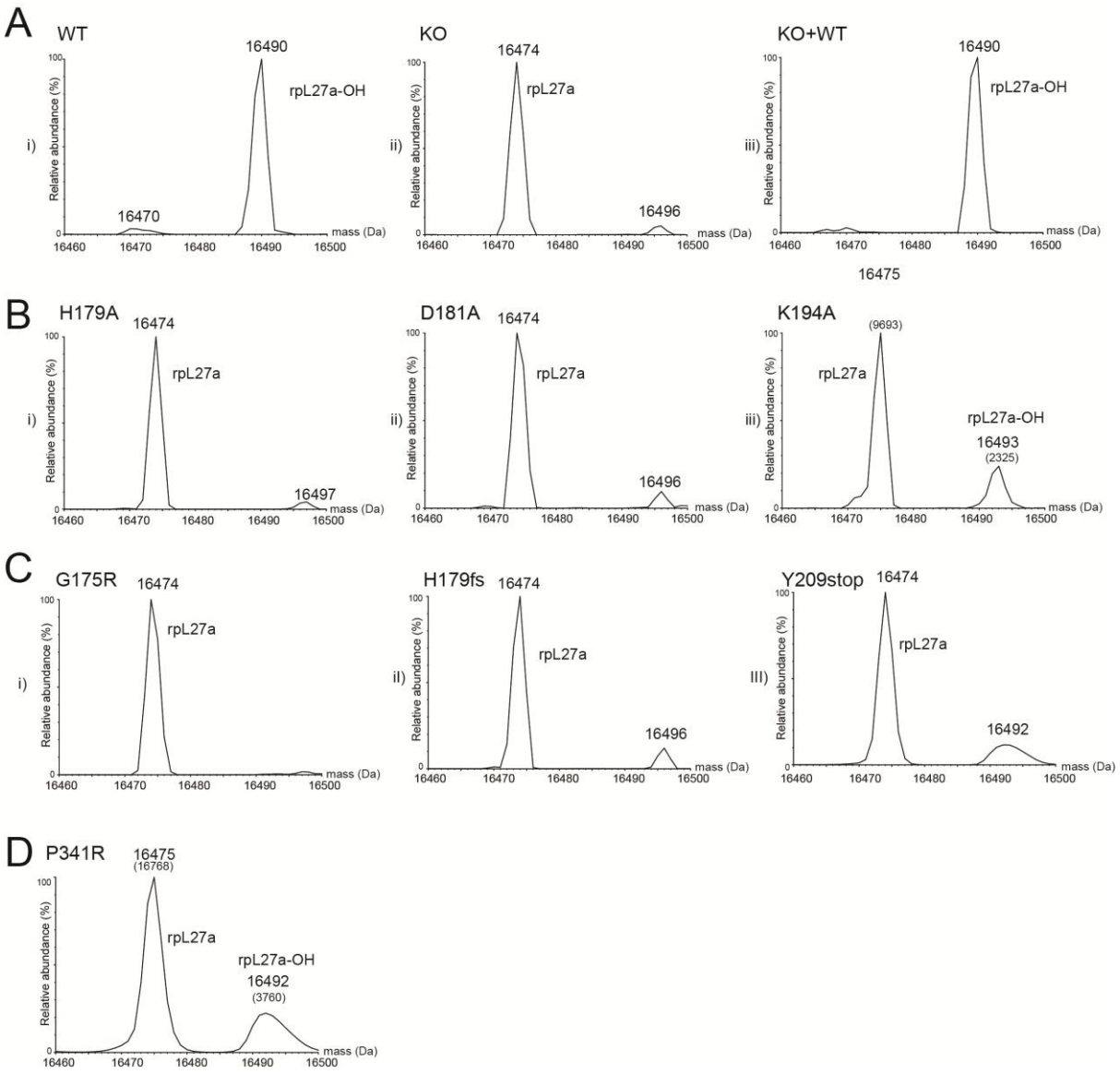


Figure 6.20: Deconvoluted ESI-MS spectra of rpL27a Mina43 knock-out MEFs reconstituted with MINA53 variants that were identified in clinical tumour samples by the Sanger Institute. A: control set: i) WT MEFs, ii) MINA53 KO MEFs transfected with empty vector (EV), c) MINA53 MEFs reconstituted with WT MINA53. B: replacement of Fe(II)- and 2OG-binding residues reduces the catalytic activity of MINA53; i) H179A, ii) D181A, iii) K194A (2OG-binding basic residue). Reconstituted intensities of MS signals are listed in brackets. K194A retained ~20% activity. C: MINA53 KO MEFs reconstituted with human MINA53 harbouring mutations identified by the Sanger COSMIC database, located in the JmjC domain: i) G175R, ii) H179fs, iii) Y209stop. D: MINA53 clinical mutation located in the C-terminal domain (P341R). Mass spectra recorded on the LCT Premier ESI-TOF-MS spectrometer using leucine enkephalin for on-line correction of mass.

In order to further characterise the biochemical consequences of these mutations on catalytic activity, the variants G175R and P341R were selected for *in vitro* kinetic profiling. Consistent with cellular Western blot data, both variants could be expressed in soluble form (Figure 6.21). Intriguingly, both variants showed catalytic activity (Figure 6.21B). However, G175R showed a decreased initial rate and a lower overall extent of hydroxylation of rpL27a peptide. Given the position of this mutation in

the core of the DSBH (Figure 6.18), it is likely that G175R interferes with 2OG binding; a kinetic assay was performed with excess 2OG. The K_m^{app} for 2OG of WT MINA53 was determined at 51±0.7 μM (Figure 6.22). Interestingly, the P341R mutant catalysed the hydroxylation of rpL27a fragment peptide at initial rate (0-10 min) that was equivalent to WT, although the overall extent of hydroxylation was also reduced, similarly to G175R.

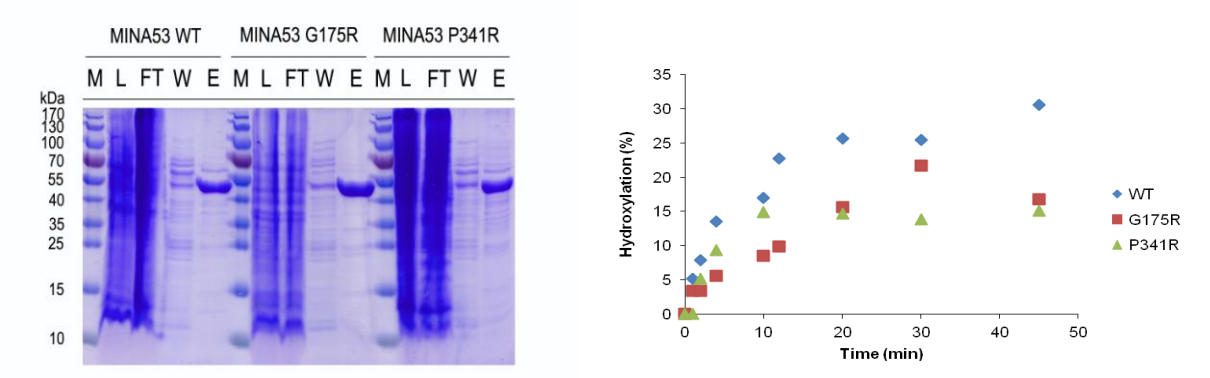


Figure 6.21: MINA53 variants (WT, G175R and P341R) produced recombinantly from a modified pET28a expression vector in *E. coli* BL21 (DE3) are soluble but show reduced levels of catalytic activity. A) SDS-PAGE gel of MINA53 demonstrates that all variants were purified to >85% purity. When *E. coli* cells reached OD_{600} 0.6, protein expression was induced with 0.5 mM IPTG and temperature reduced from 37°C to 18°C; after overnight growth, cells were harvested and lysed by sonication. MINA53 variants were purified by Ni(II)-affinity chromatography and buffer-exchanged into 50 mM HEPES, 500 mM NaCl, 5-10 % glycerol, 1 mM DTT, pH 7.5) using PD10 columns. M: molecular weight marker; L: lysate; FT: flow-through; W: wash; E: eluate. B) MALDI-MS time-course of MINA53 variant hydroxylation activity against rpL27a substrate fragment. Blue diamonds: WT MINA53, red squares: G175R; green triangles: P341R. Single measurements were performed at each time-point and the extent of hydroxylation analysed by MALDI-MS.

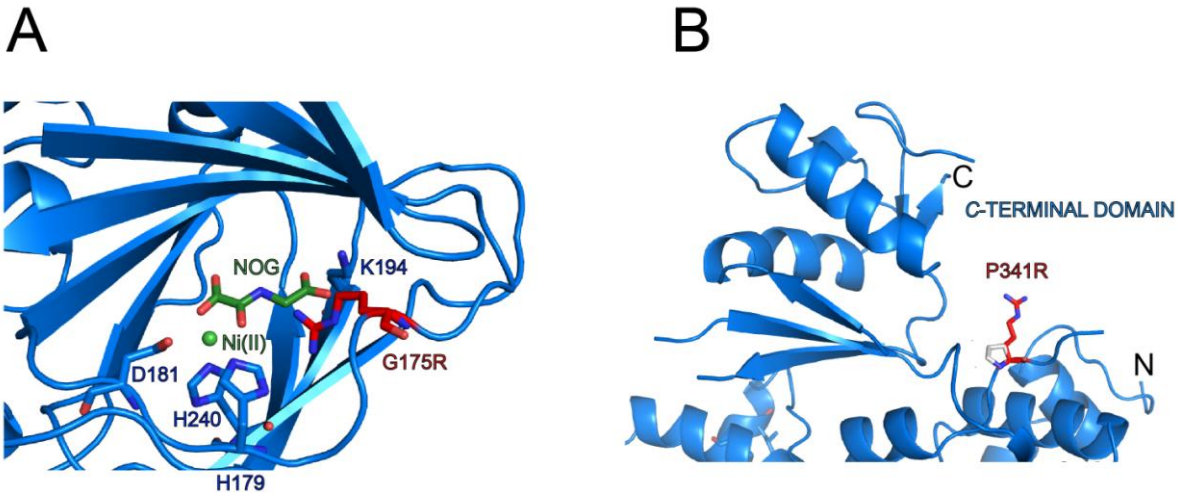


Figure 6.22: A view from a crystal structure of MINA53 with models of clinical mutations reveal potential catalytic and structural importance for residues G175 and P341. A) A model of G175R mutation in the core of MINA53 DSBH suggests that residue G175R would interfere with 2OG binding, and likely sterically clash with 2OG-binding residue K194; B) mutation P341R replaces a structurally important prolyl residue located on a loop at the interface between the jmjC and the *C*-terminal domains, placing a strongly positively charged residue at the interdomain interface, thereby explaining the observed decrease in catalytic activity of the P341R variant compared to WT. MINA53 variants were modelled using the 'Mutagenesis' tool of PyMol Molecular Graphics System (Schrodinger).

Next, it was investigated whether the G175R and P341R substitutions affect the thermal stability of MINA53. Differential scanning fluorimetry (DSF), discussed in Chapter 4, was used to determine the melting points (T_m) of the two variants compared to WT MINA53 (Figure 6.23). The T_m of the G175R variant was 45.7°C, within experimental error of the T_m for WT (46.5°C), suggesting that the substitution in the DSBH did not affect the tertiary structure. In contrast, the P341R substitution decreased the T_m from 46.5°C to 37.3°C, indicating that P341 has a structurally important role and its replacement with a basic residue interferes with the tertiary structure. Overall, the kinetic and stability measurements indicate that both kinetically deactivating as well as structurally destabilising mutations of MINA53 occur in cancer.

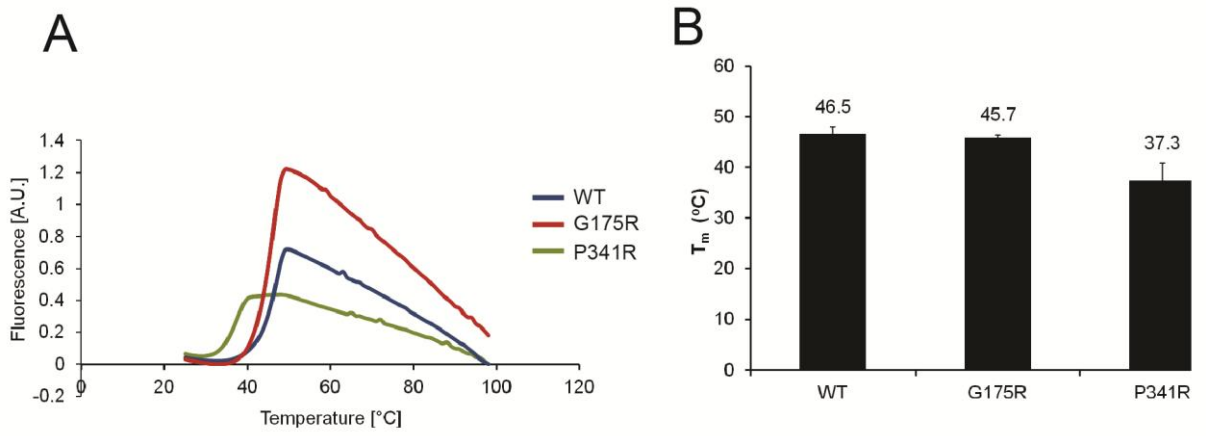


Figure 6.23: DSF measurement of thermal stability of MINA53 variants. MINA53 variants (WT, G175R and P341R, 2 μM) were incubated with SyproOrange™ dye and increase in fluorescence was measured between 25-98°C using OpticonMonitor software connected to MJ MINI thermal cycler (Bio-Rad). Results were analysed by fitting a Boltzmann sigmoidal function using GraphPad Prism (version 5.00). Tm values were measured in triplicate. Panel A shows representative thermal denaturation spectra. Panel B shows the average T_m value of three experiments, with error bars denoting standard deviation from the mean. The results demonstrated that variant P341R is destabilised by ~9oC relative to WT, whereas the T_m value for G175R is within experimental error of the WT value.

Initial biological characterisation was performed on mutants G175R and P341R by Adam Zayer, using the soft-agar assay described previously, in order to assess the tumorigenic potential of cell expressing MINA53 cancer mutants. Consistent with the biochemical analysis of rpL27a hydroxylation, the phenotypic behaviour of both mutants in this assay was indistinguishable from KO or H179A MINA53 (Figure 6.24). Therefore, the reduced biochemical activity of these mutants towards rpL27a correlates with inactivation of tumour suppressor activity.

KO MEFs were then reconstituted with a second cohort of variants that had been subsequently added to the COSMIC database. Due to material and temporal constraints, it was not possible to prepare and analyse them all. Therefore, a selected number were prepared and analysed so as to cover the major mutational 'hot-spots' that were becoming apparent (cluster I, spanning Q67-G97; cluster II, D116-I150; cluster III, P170-P220; cluster IV, W264-V302; D332-S342; cluster VI, P377-K389; Figure 6.17).

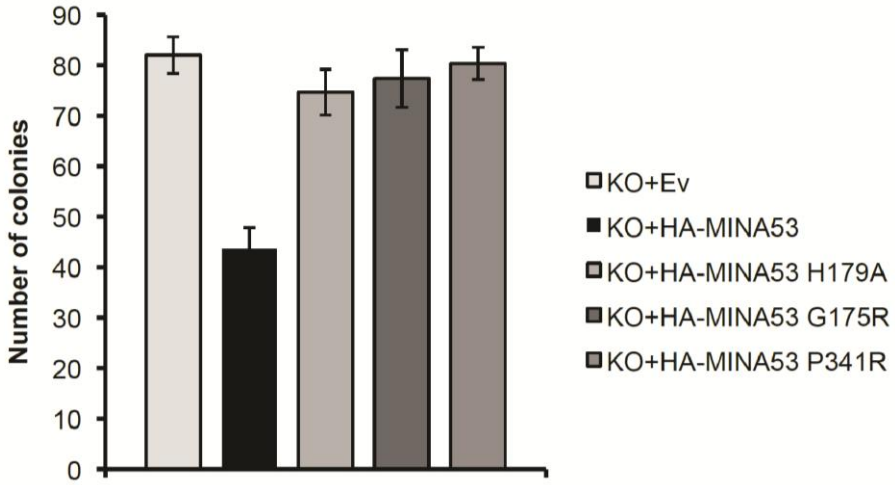


Figure 6.24: Catalytically deactivating mutations of MINA53 identified on COSMIC database increase anchorage-independent growth. K-Ras transformed MINA53 KO MEFs reconstituted with EV, HA-MINA53, H179A, G175R and P341R were grown on soft agar. Number of counted colonies (± s.d.) after 4 weeks is shown. Figure kindly provided by Adam Zayer.

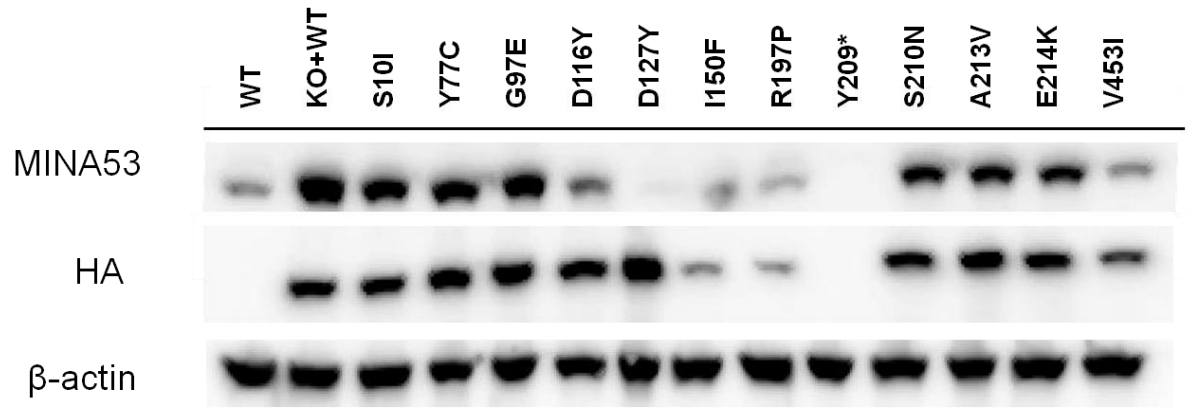


Figure 6.25: Western blot analysis of MINA53 variants containing mutations identified in the COSMIC and TCGA databases demonstrates that several mutations, including mis-sense and nonsense mutations, affect the solubility of the protein. MINA53 protein levels in MINA53 KO MEFs over-expressing HA-tagged MINA53 variants are markedly increased relative to endogenous levels of MINA53 in WT MEFs. Levels of D116Y and D127Y variants appear lower compared to WT control, whilst the level of their HA-tags is comparable to WT, suggesting that the mutations interfere with the antigen activity of the α-MINA53 antibody. Variants I150F and R197P are less soluble that WT, as demonstrated by lower intensity of the signals generated by α-MINA53 and α-HA antibodies. Y209* is insoluble (no signal was detected at the predicted size of the truncated protein). V453I is less soluble than WT. Other mutations produced soluble protein. β-actin was used as a loading control. Figure kindly provided by Adam Zayer.

MINA53 variant S10I has no effect on rpL27a hydroxylation (Figure 6.26). However, its location in the *N*-terminal region has not been crystallographically characterised so it was not possible to rationalise this further. A number of mutations occurred in 'cluster I', *N*-terminal to the DSBH of MINA53. P71H was identified in endometrial cancer, which has the highest frequency of MINA53 mutations found on COSMIC. WPMS revealed that the protein was 90% active. Similarly, variants Y77C and R92L (identified in two independent types of cancer (lung and stomach, Figure 6.17), as well as G97E, were 90% active. Similarly, mutations in cluster II did not affect hydroxylation of rpL27a: D116Y and D127Y (both identified in endometrial cancer) were >90% active; I150F (stomach cancer) showed 88% activity, consistent with reduced protein expression levels observed by Western blotting (Figure 6.25). It cannot be excluded that effects on MINA53 activity may be revealed by more detailed kinetic analyses.

The residue R197 was found to be independently mutated in four cancers (biliary, glioma, stomach and lung cancers) to generate variants R197C, R197H and R197P. It was hypothesised that the latter (R197P) would lead to the largest chemical and structural differences MINA53 and presented the highest chance of observing an effect on MINA53 activity. R197 is a highly conserved residue on strand βIV, also found in NO66 and bacterial ycfD homologues (see Figure 6.1). The variant R197P was markedly less soluble than the WT (Figure 6.25) and showed complete inactivation of MINA53 (Figure 6.26i). The R197P substitution introduced a conformationally-constrained residue into a structurally important, evolutionarily conserved strand βIV in the DSBH. R197 may interact with E214 *via* hydrogen bonding interaction (R197 NH2 - E214 OE2 4.9 Å) that could stabilise the conformation of βIV-βV loop, important in substrate recognition. Intriguingly, E214 is also mutated in cancer (stomach, Figure 6.17), strongly arguing for an important role of this intra-molecular interaction in tumour suppression.

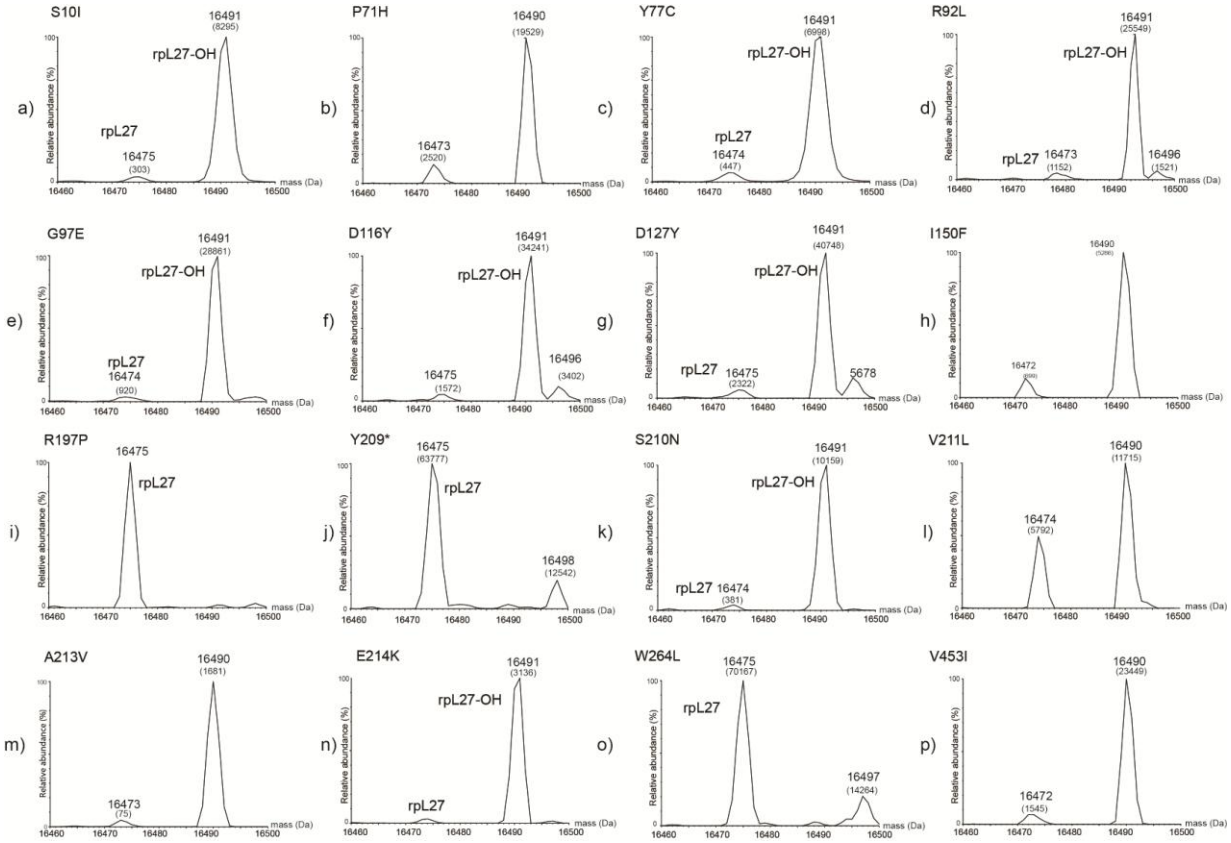


Figure 6.26: Deconvoluted ESI-MS spectra of rpL27a from MINA53 knock-out MEFs reconstituted with MINA53 variants that were sequenced from clinical tumour samples by the Sanger Institute. a) S10I, b) P71H, c) Y77C, d) R92L, e) G97E, f) D116Y, g) D127Y, h) I150F, i) R197P, j) Y209*, k) S210N, l) V211L, m) A213V, n) E214K, o) W264L, p) V453I. Peak at 16491 Da was assigned as hydroxylated rpL27a; peaks at 16475 Da were assigned as unhydroxylated rpL27a. Mass spectra recorded on the LCT Premier ESI-TOF-MS spectrometer using leucine enkephalin for on-line correction of mass.

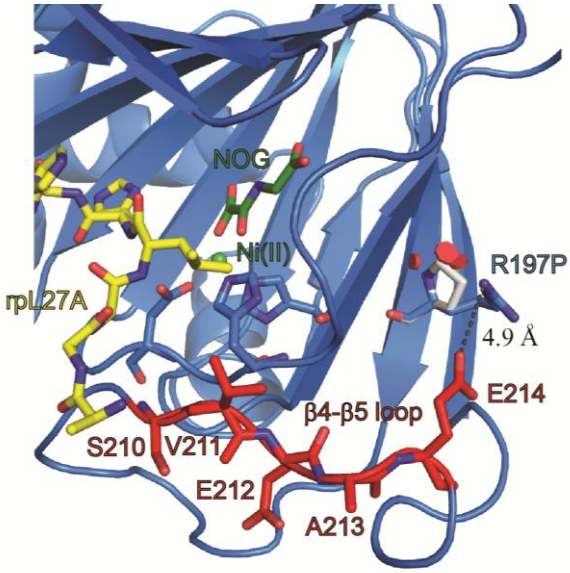


Figure 6.27: A model of R197P mutation in the crystal structure of MINA53.rpL27a (courtesy of Dr. Rasheduzzaman Chowdhury) reveals a potential interaction of R197 with E214 from the β4-β5 insert, involved in substrate recognition. Western blot data suggests that variant R197P leads to a decrease in protein solubility; the high evolutionary conservation of R197 suggests its structural importance; replacement with a prolyl residue likely leads to conformational changes which distort the DSBH, leading to a decreased solubility of the protein.

Interestingly E214 forms part of a mutational 'hot-spot' on the βIV-βV substrate-recognition loop, which strongly argues for a functional consequence of MINA53 mutation in cancer; each residue between positions 210-214 is mutated at least once (Figure 6.17). These mutations represent the extension of mutational Cluster III and were therefore expected to lead to deactivation of MINA53. However, substitutions S210N, A213V and E214K did not have an effect on MINA53 activity against rpL27a (Figure 6.26). This was especially surprising for the E214K substitution, which is chemically non-conservative. This particular example could be rationalised based on the fact that the side-chain of residue E214 points away from rpL27a chain; therefore only the main chain interacts with the substrate and the acidic to basic side-chain substitution does not affect enzymatic activity.

In contrast, the residue V211 was predicted to form a hydrophobic interaction with the substrate residue L38 (methylene distance 4.9 Å). The V211L variant resulted in a 30% decrease of rpL27a hydroxylation, consistent with less favourable interaction between residue 211 and L38 (methylene distance reduced to 3.6 Å, Figure 6.28)

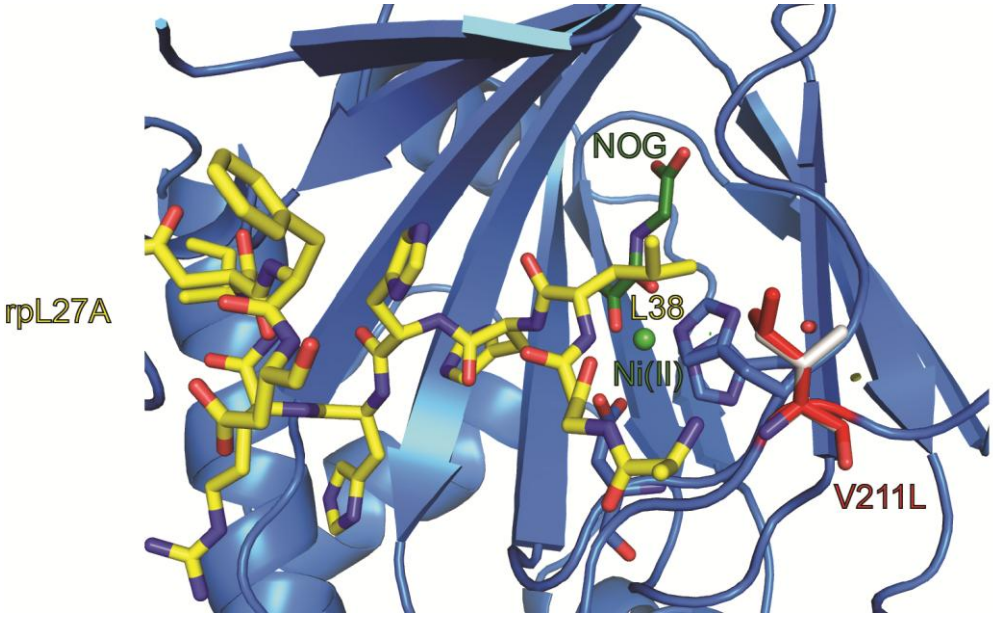


Figure 6.28: A model of V211L mutation in the crystal structure of MINA53.rpL27a (courtesy of Dr. Rasheduzzaman Chowdhury) reveals a potential hydrophobic interaction of V211 located on the β4-β5 insert with residue L38 from the rpL27a substrate fragment. A hydrophobic residue at position -1 relative to the hydroxylated residue is conserved in the ycfD.L16 complex (but not in NO66.rpL8), indicating potential functional importance. V211L variant introduces an extra methylene unit in the vicinity of L38, thereby decreasing the L38-V211 distance from 4.9 to 3.6Å and increasing steric crowding, which may account for the observed decrease in enzymatic activity of V211L variant against rpL27a.

Whilst the substitution V453I, which has been identified in endometrial cancer, did not lead to loss of catalytic activity against rpL27a (Figure 6.26p), the substitution W264L resulted in complete loss of activity against rpL27a (Figure 6.26o). W264 is the first residue of the MINA53 dimerisation domain. It is the first residue of the dimerisation helix α-1 and therefore likely important for the maintenance of the secondary structure; the protein product is soluble (Figure 6.19). W264 is located in the centre of a unique hydrophobic/aromatic pocket at the dimer interface, formed by residues T288, P291, L294, V303 from monomer A, and F311 F267, L289, L308, L325 from monomer B. The indole ring of W264 slots in the centre of a cage formed by hydrophobic residues from both monomers and likely participates in strong hydrophobic and π-stacking interactions required for formation of the dimer (Figure 6.29).

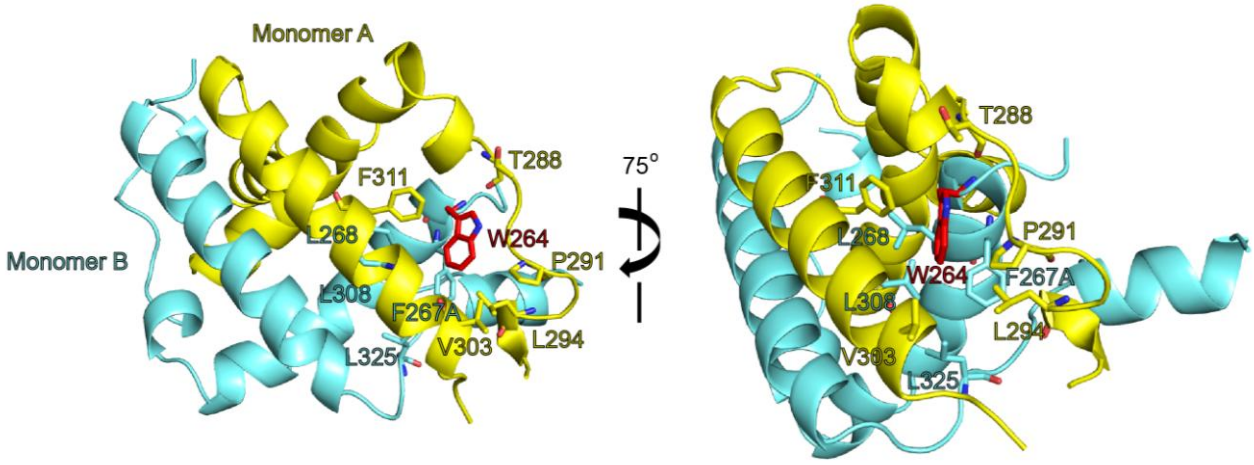


Figure 6.29: W264 is the first residue of the MINA53 dimerisation domain and the first residue of the dimerisation helix α-1. W264 is located in the centre of a hydrophobic/aromatic pocket at the dimer interface, formed by residues T288, P291, L294, V303 from monomer A, and F311 F267, L289, L308, L325 from monomer B. The indole ring of W264 slots in the centre of a cage formed by hydrophobic residues from both monomers and likely participates in strong hydrophobic and π-stacking interactions required for formation of the dimer.

It is unclear at present why several of the MINA53 mutations studied do not affect rpL27a hydroxylation but are apparently selected for in cancer cells. Overall, under the tested conditions, six of the studied mutations inactivated rpL27a hydroxylation, two modestly reduced it, and 12 had no effect. Of the latter, there are strong arguments for a functional affect of the mutation, including; i) occurrence in multiple tumours (e.g. R92L) or ii) presence in structurally relevant hot spots (e.g. S210N, A213V and E214K). If correct, this would imply that these mutations would affect the

tumour-suppressive activity of MINA53, and that the role of MINA53 in cancer is independent of rpL27a hydroxylation. To test this hypothesis, Adam Zayer assessed the activity of several cancer variants in the MEF autophagy assay (mitochondrial clearance in response to Rapamycin, as described above), including those that ablated rpL27a hydroxylation (W264L, P341R and G175R) and those that did not affect rpL27a hydroxylation levels relative to WT (P71H, R92L, V211L). Intriguingly, all of the cancer variants failed to restore normal autophagy in MINA53 KO MEFs, suggesting that despite differences in their biochemical activity towards rpL27a they were phenotypically equivalent. These data support the hypothesis that the role of MINA53 in autophagy (and cancer) is independent of rpL27a hydroxylation, and argue therefore for the existence of a novel substrate.

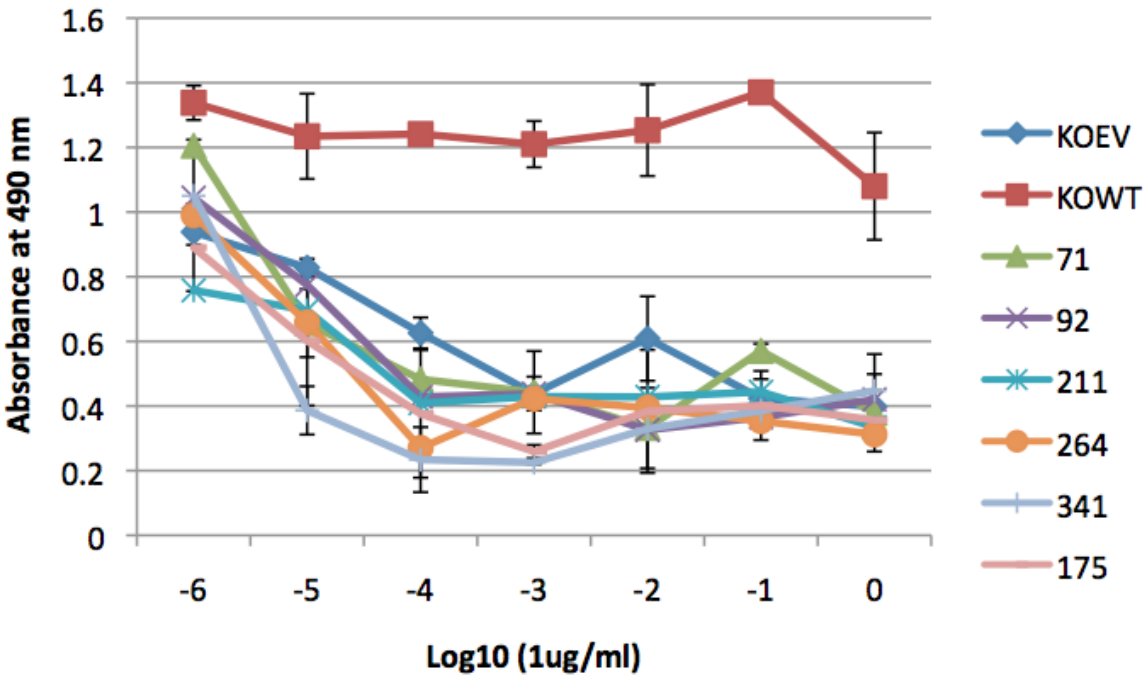


Figure 6.30: Cancer mutants of MINA53 show increased Rapamycin sensitivity compared to WT. MINA53 variants identified in human cancers were expressed in MINA53 KO MEFs. Their sensitivity to Rapamycin was measured by MTS assay. High absorbance at 490 nm indicates higher mitochondrial activity measured by reduction of MTS to formazan and is a measure of cell viability. KOEV: MINA53 KO MEFs expressing empty vector only; KOWT: MINA53 KO MEFs expressing WT MINA53; 71-175: MINA53 KO MEFs expressing MINA53 variants P71H, R92L, V211L, W264L, P341R and G175R. Figure kindly provided by Adam Zayer.

6.8. *In vitro* testing of potential non-ribosomal MINA53 substrates

The existence of an alternative substrate/binding partner of MINA53, which is more sensitive to point-mutations in MINA53 structure, could explain the inconsistency of the mass-spectrometry and phenotypic results.

Mining of the Uniprot database for human proteins which contain an hHbH motif (where 'h' is a hydrophobic, and 'b' is a basic residue), identified candidate hydroxylation sites in zinc-finger homeobox 3 (ZFHX3) (ESAL^{3553}HKHRT) and interferon-related developmental regulator 2 (IFRD2) (SGM^{446}HHHLQNN), known tumour suppressor genes.

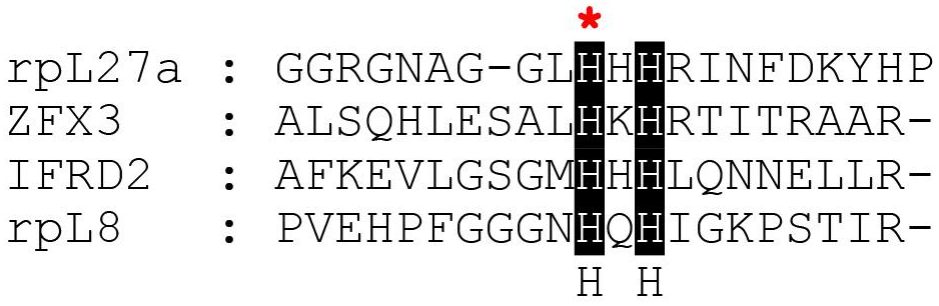


Figure 6.31: Protein fragment sequence alignment of HxH-motif containing proteins rpL27a, ZFHX3, IFRD2 and rpL8. RpL8 is hydroxylated by NO66, but not MINA53 (Dr. Wei Ge, personal communication). The activity of MINA53 against ZFHX3 and IFRD2 has not been determined. Red star indicates the position of hydroxylation substrate of MINA53 on rpL27a. RpL8 is not a hydroxylation substrate for MINA53 (Dr. Wei Ge, personal communication).

In order to test whether these candidates may also be hydroxylated by MINA53, 20-mer fragment peptides were obtained from GL Biochem (China), with the target histidine located in the middle of the sequence, and tested alongside rpL27a peptide fragment. Incubation of an IFRD2 fragment peptide with recombinant MINA53 showed no hydroxylation when 100 μM of peptide were incubated with 1 or 10 μM of MINA53. Under identical conditions, the rpL27a fragment peptide was 57% and 90% hydroxylated, respectively. When 100 μM of each fragment peptide was incubated with 10 μM of MINA53 simultaneously in a competition experiment, only the rpL27a fragment was hydroxylated (75%), whereas the IFRD2 fragment was not hydroxylated at all (Figure 6.32). Overall, this likely excludes IFRD2 as a hydroxylation substrate of MINA53, although it cannot be excluded that MINA53 may hydroxylate IFRD2 on the whole-protein level.

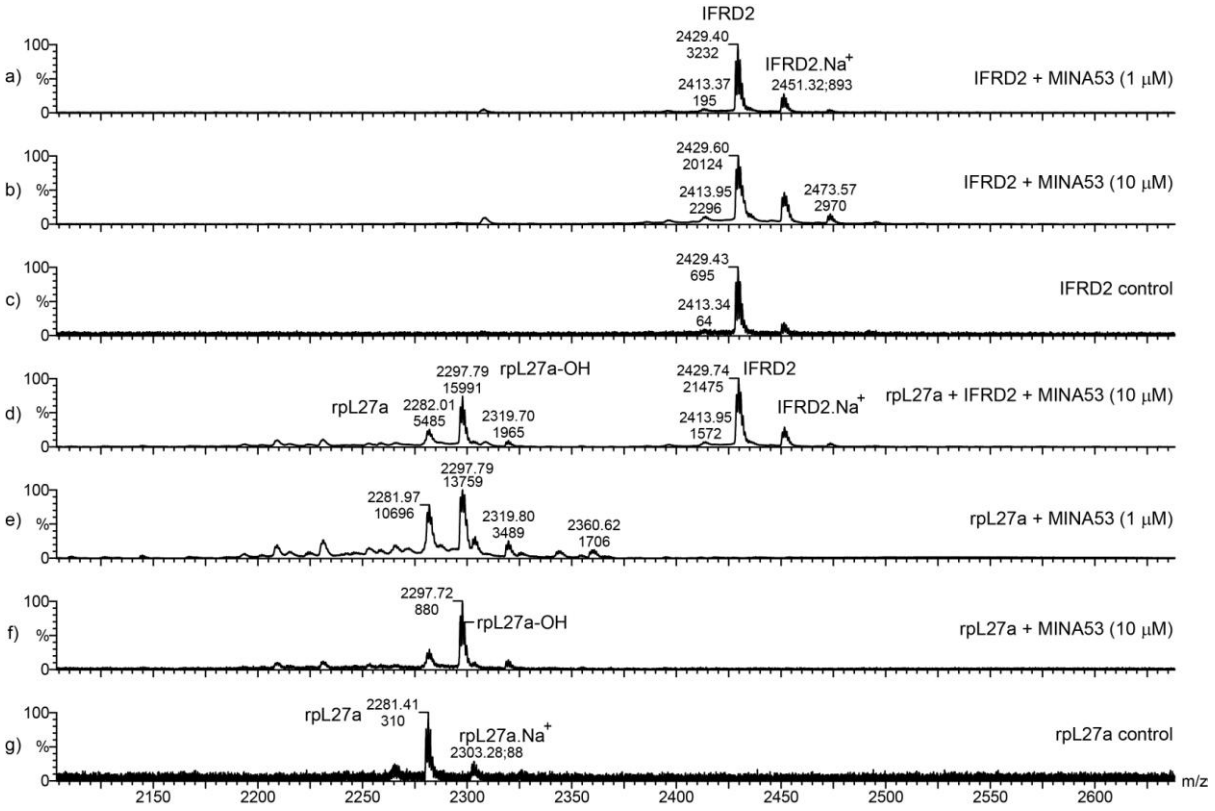


Figure 6.32: MINA53 hydroxylates the rpL27a fragment but not the IFRD2 substrate fragment. Assay mixtures (final volume 20 μL in 50 mM HEPES, 500 mM NaCl, 5 % glycerol, pH 7.5) contained MINA53 (1 or 10 μM), peptide fragments (100 μM), 2OG (200 μM), Fe(II) (100 μM) and ascorbate (1 mM). The reaction mixtures were incubated at 37°C for 30 min, placed on ice, quenched with 10 μl of $HCOOH_{aq}$ (1% vol/vol) The resultant soluble mixture (1 μl) and α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution (1 μl) (LaserBio Labs) were spotted onto a 96 well MALDI sample plate, and analysed using MALDI-TOF MS on a Waters Micromass™ MALDI micro MX™ mass spectrometer in negative ion reflectron mode. Peptide sequences were rpL27a: GRGNAGGLHHHRINFDKYHP, predicted mass: 2283.5 Da), and IFRD2 (AFKEVLGSGMHHHLQNNELLR, predicted mass: 2430.7 Da).

In contrast, the ZFHX3 fragment peptide showed 44% hydroxylation when incubated with 10 μM MINA53. Incubation with 1 μM MINA53 did not lead to hydroxylation (Figure 6.33). Under identical conditions, rpL27a was 80% and 40% hydroxylated, respectively, suggesting that although ZFHX3 is hydroxylated by MINA53, the catalytic activity of MINA53 for the ZFHX3 fragment peptide is lower than for rpL27a and the ZFHX3 peptide is less likely to represent a biologically relevant substrate of MINA53. Consistent with this, an *in vitro* interaction assay suggested that recombinant ZFHX3 protein does not interact with MINA53 (Dr. Mathew Coleman, personal communication). Its large size (404 kDa) renders is difficult to study by whole-protein MS, so it was not pursued further.

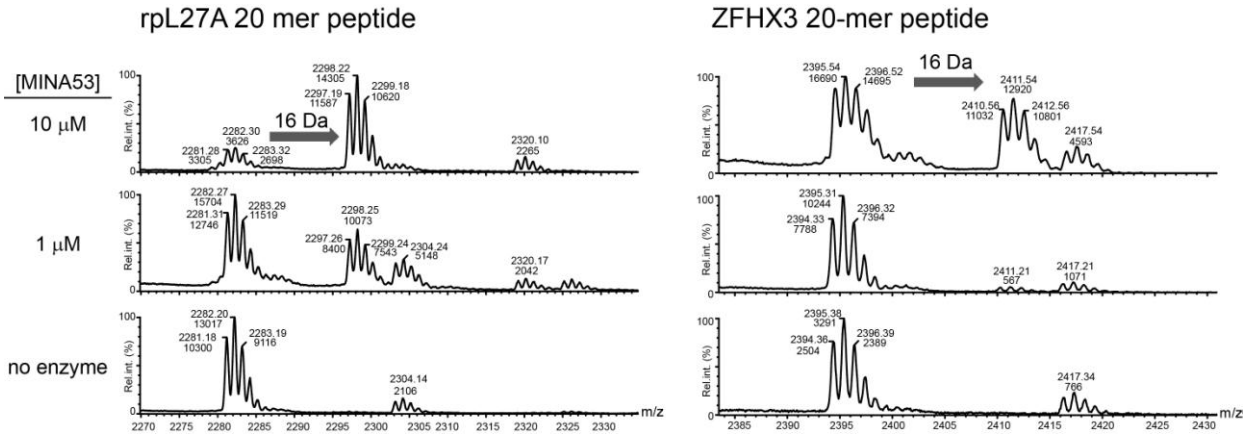


Figure 6.33: MINA53 hydroxylates peptide fragments (20-mers) of rpL27a and ZFHX3. Assay mixtures (final volume 20 μL in 50 mM HEPES, 500 mM NaCl, 5 % glycerol, 1 mM DTT, pH 7.5) contained MINA53 (1 or 10 μM), peptide fragments (100 μM), 2OG (200 μM), Fe(II) (100 μM) and ascorbate (1 mM). The reaction mixtures were incubated at 37°C for 30 min, placed on ice, quenched with 10 μl of $HCOOH_{aq}$ (1% vol/vol), and the mixture centrifuged to separate any insoluble material that precipitated during quenching. The resultant soluble mixture (1 μl) and α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution (1 μl) (LaserBio Labs) were double spotted onto a 96 well MALDI sample plate, and analysed using MALDI-TOF MS on a Waters Micromass™ MALDI micro MX™ mass spectrometer in negative ion reflectron mode. Peptide sequences were (GRGNAGGLHHHRINFDKYHP, rpL27a, predicted mass: 2283.5 Da), and ALSQHLESALHKHRTITRAAR (ZFHX3, predicted mass: 2396.7 Da).

6.9. Conclusions

In this chapter, the LC-MS methodology for analysis of ribosomal proteins was extended to mouse and human cells to demonstrate that MINA53 hydroxylates rpL27a in mammalian cells. In collaboration with Adam Zayer (Wellcome Trust for Human Genomics), it was demonstrated that human and mouse rpL27a are hydroxylated in a manner that depends on the catalytic activity of MINA53. RpL27a hydroxylation is wide-spread and found in all mouse organs analysed, as well as in cancer cell lines and in clinical non-Hodgkin lymphoma samples. Hypoxic treatment of tumour cell lines resulted in small (<15%) reduction of rpL27a hydroxylation, and prolonged exposure of mice to mild hypoxia (10% O_2 for 7 days) did not reduce rpL27a hydroxylation, suggesting that MINA53 has high affinity for O_2, or that unhydroxylated rpL27a is not efficiently incorporated into the ribosome. A partial or complete reduction of rpL27a hydroxylation was observed in a number of clinically identified *MINA53* mutations from the COSMIC database of cancer mutations. Intriguingly, these apparently active mutants were completely inactive in cellular phenotypic assays, therefore suggesting the existence of an alternative substrate with important tumour suppressor functions downstream of MINA53. On the peptide level, two candidate proteins containing a hHbH motif, showed lower levels of turnover by MINA53 compared to rpL27a; IFRD2 was not hydroxylated at all, whereas ZFHX3, a

tumour suppressor protein, was hydroxylated 50% less than rpL27a under identical conditions, suggesting that they are not likely to represent biologically relevant substrates of MINA53.

Future work could focus on characterising the biochemical activity of all sequenced cancer-associated mutations of *MINA53* and compare the results to autophagy phenotypic assays in order to confirm or overthrow the hypothesis that the role of *MINA53* as a tumour suppressor is mediated by ribosomal protein hydroxylation.

Importantly, the work described in this chapter provided a demonstration that cancer tissues may differ from normal tissues in the status of post-translational hydroxylation of their ribosomal proteins. Regardless of the role of ribosome hydroxylation in tumour suppression, the identification of clinical deactivating *MINA53* mutations provides a potential biomarker, and a chemical 'handle' that could be exploited therapeutically in a sub-section of cancers carrying deactivating *MINA53* mutations. As described in Chapter 3, it is principle possible to inhibit selectively modified ribosomes; therefore the work presented in this chapter provides a rationale to commence a search for a selective inhibitor of unhydroxylated rpL27a, which would be active against a sub-set of tumours with inactivating mutation(s) of *MINA53*, but not normal tissue. The method would rely on DNA sequencing of *MINA53*, which should indicate whether deactivating mutations are present in the gene. A selective inhibitor of unhydroxylated ribosomes could then be applied. It is notable that a commonly used inhibitor of the eukaryotic ribosome, cycloheximide, binds in the vicinity of H39 on rpL27a (24). Given its chemical tractability and a defined binding site, cycloheximide may present a suitable starting point for medicinal chemistry efforts to develop a selective inhibitor of unhydroxylated ribosomes.

6.10. References

1. Tsuneoka, M., Koda, Y., Soejima, M., Teye, K., and Kimura, H. (2002) A novel myc target gene, mina53, that is involved in cell proliferation. *J Biol Chem* **277**, 35450-35459
2. Eilbracht, J., Kneissel, S., Hofmann, A., and Schmidt-Zachmann, M. S. (2005) Protein NO52--a constitutive nucleolar component sharing high sequence homologies to protein NO66. *European Journal of Cell Biology* **84**, 279-294
3. Eilbracht, J., Kneissel, S., Hofmann, A., and Schmidt-Zachmann, M. S. (2005) Protein NO52 - a constitutive nucleolar component sharing high sequence homologies to protein NO66. *European Journal of Cell Biology* **84**, 279-294
4. Ishizaki, H., Yano, H., Tsuneoka, M., Ogasawara, S., Akiba, J., Nishida, N., Kojiro, S., Fukahori, S., Moriya, F., Matsuoka, K., and Kojiro, M. (2007) Overexpression of the myc target gene Mina53 in advanced renal cell carcinoma. *Pathol Int* **57**, 672-680
5. Tsuneoka, M., Fujita, H., Arima, N., Teye, K., Okamura, T., Inutsuka, H., Koda, Y., Shirouzu, K., and Kimura, H. (2004) Mina53 as a potential prognostic factor for esophageal squamous cell carcinoma. *Clin Cancer Res* **10**, 7347-7356
6. Fukahori, S., Yano, H., Tsuneoka, M., Tanaka, Y., Yagi, M., Kuwano, M., Tajiri, T., Taguchi, T., Tsuneyoshi, M., and Kojiro, M. (2007) Immunohistochemical expressions of Cap43 and Mina53 proteins in neuroblastoma. *J Pediatr Surg* **42**, 1831-1840
7. Kuratomi, K., Yano, H., Tsuneoka, M., Sakamoto, K., Kusukawa, J., and Kojiro, M. (2006) Immunohistochemical expression of Mina53 and Ki67 proteins in human primary gingival squamous cell carcinoma. *Kurume Med J* **53**, 71-78
8. Ogasawara, S., Komuta, M., Nakashima, O., Akiba, J., Tsuneoka, M., and Yano, H. (2010) Accelerated expression of a Myc target gene Mina53 in aggressive hepatocellular carcinoma. *Hepatol Res* **40**, 330-336
9. Zhang, Q., Hu, C. M., Yuan, Y. S., He, C. H., Zhao, Q., and Liu, N. Z. (2008) Expression of Mina53 and its significance in gastric carcinoma. *Int J Biol Markers* **23**, 83-88
10. Teye, K., Arima, N., Nakamura, Y., Sakamoto, K., Sueoka, E., Kimura, H., and Tsuneoka, M. (2007) Expression of Myc target gene mina53 in subtypes of human lymphoma. *Oncol Rep* **18**, 841-848
11. Komiya, K., Sueoka-Aragane, N., Sato, A., Hisatomi, T., Sakuragi, T., Mitsuoka, M., Sato, T., Hayashi, S., Izumi, H., Tsuneoka, M., and Sueoka, E. (2010) Expression of Mina53, a novel c-Myc target gene, is a favorable prognostic marker in early stage lung cancer. *Lung Cancer* **69**, 232-238
12. Teye, K., Tsuneoka, M., Arima, N., Koda, Y., Nakamura, Y., Ueta, Y., Shirouzu, K., and Kimura, H. (2004) Increased expression of a Myc target gene Mina53 in human colon cancer. *Am J Pathol* **164**, 205-216
13. Komiya, K., Sueoka-Aragane, N., Sato, A., Hisatomi, T., Sakuragi, T., Mitsuoka, M., Sato, T., Hayashi, S., Izumi, H., Tsuneoka, M., and Sueoka, E. (2010) Mina53, a novel c-Myc target gene, is frequently expressed in lung cancers and exerts oncogenic property in NIH/3T3 cells. *J Cancer Res Clin Oncol* **136**, 465-473
14. Wang, Y., Ngo, V. N., Marani, M., Yang, Y., Wright, G., Staudt, L. M., and Downward, J. (2010) Critical role for transcriptional repressor Snail2 in transformation by oncogenic RAS in colorectal carcinoma cells. *Oncogene* **29**, 4658-4670
15. Morgenstern, J. P., and Land, H. (1990) Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. *Nucleic Acids Res* **18**, 3587-3596
16. Guo, J. Y., Chen, H. Y., Mathew, R., Fan, J., Strohecker, A. M., Karsli-Uzunbas, G., Kamphorst, J. J., Chen, G., Lemons, J. M., Karantza, V., Coller, H. A., Dipaola, R. S., Gelinas, C., Rabinowitz, J. D., and White, E. (2011) Activated Ras requires autophagy to maintain oxidative metabolism and tumorigenesis. *Genes Dev* **25**, 460-470
17. Kim, M. J., Woo, S. J., Yoon, C. H., Lee, J. S., An, S., Choi, Y. H., Hwang, S. G., Yoon, G., and Lee, S. J. (2011) Involvement of autophagy in oncogenic K-Ras-induced malignant cell transformation. *J Biol Chem* **286**, 12924-12932

18. Cuervo, A. M. (2004) Autophagy: many paths to the same end. *Mol Cell Biochem* **263**, 55-72
19. Lu, Y., Chang, Q., Zhang, Y., Beezhold, K., Rojanasakul, Y., Zhao, H., Castranova, V., Shi, X., and Chen, F. (2009) Lung cancer-associated JmjC domain protein mdig suppresses formation of tri-methyl lysine 9 of histone H3. *Cell Cycle* **8**, 2101-2109
20. Ge, W., Wolf, A., Feng, T., Ho, C. H., Sekirnik, R., Zayer, A., Granatino, N., Cockman, M. E., Loenarz, C., Loik, N. D., Hardy, A. P., Claridge, T. D., Hamed, R. B., Chowdhury, R., Gong, L., Robinson, C. V., Trudgian, D. C., Jiang, M., Mackeen, M. M., McCullagh, J. S., Gordiyenko, Y., Thalhammer, A., Yamamoto, A., Yang, M., Liu-Yi, P., Zhang, Z., Schmidt-Zachmann, M., Kessler, B. M., Ratcliffe, P. J., Preston, G. M., Coleman, M. L., and Schofield, C. J. (2012) Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol* **8**, 960-962
21. Lancaster, D. E., McNeill, L. A., McDonough, M. A., Aplin, R. T., Hewitson, K. S., Pugh, C. W., Ratcliffe, P. J., and Schofield, C. J. (2004) Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J* **383**, 429-437
22. Ehrismann, D., Flashman, E., Genn, D. N., Mathioudakis, N., Hewitson, K. S., Ratcliffe, P. J., and Schofield, C. J. (2007) Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem J* **401**, 227-234
23. Liu, L., and Simon, M. C. (2004) Regulation of transcription and translation by hypoxia. *Cancer Biol Ther* **3**, 492-497
24. Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011) Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* **334**, 941-948

7. Conclusions

Contents

7.1. Aims of the work described in this thesis 270

7.2. Mass spectrometric and functional studies on ycfD 270

7.3. *YcfD* mediates resistance to peptidyl transferase ribosomal antibiotics 271

7.4. Biochemical characterisation of a thermophilic 2OG oxygenase $ycfD_{RM}$ 271

7.5. Structural studies on $ycfD_{RM}$ 272

7.6. Biochemical characterisation of MINA53 activity in cells 273

7.7. Future work 274

7.1. Aims of the work described in this thesis

The 2OG oxygenases comprise a superfamily of ferrous iron dependent dioxygenases with multiple biological roles, including in hypoxia sensing, transcriptional control, and splicing control. It was recently proposed that 2OG oxygenases catalyse the hydroxylation of ribosomal proteins in prokaryotes (ycfD) and in humans (NO66 and MINA53), raising the possibility that 2OG oxygenases also control translation. The work described in this thesis concerned investigations on the biochemical and functional aspects of prokaryotic and mammalian ribosomal protein hydroxylases (ROX) *in vitro* and in cells.

The thesis set out to investigate the potential role of *ycfD* as a ribosomal protein hydroxylase in cells, including its specificity and oxygen dependence, and its functional role in *E. coli*. Structural aspects of ribosomal protein hydroxylation were investigated using protein crystallography, and the activity of the human homologue of ycfD, MINA53 was investigated in cells, using the methodology developed on the prokaryotic system.

7.2. Mass spectrometric and functional studies on *ycfD*

It was demonstrated that out of 52 *E. coli* ribosomal proteins identified, only L16 is oxidatively modified by ycfD, that the modification is located on R81, and that hydroxylation of L16 varies with pO_2, suggesting that ycfD has the potential to act as an oxygen sensor. Functional studies revealed that *ycfD* deletion results in growth phenotype at low temperature, and in minimal medium; furthermore, *ycfD* deletion results in decreased global rate of translation. However, *ycfD* deletion does not affect translational accuracy or ribosome assembly. Consistent with recent reports that the ycfD substrate residue R81 assists the translocation step of tRNA from the A- to the P-site, the results presented in this thesis are consistent with ycfD catalysing a modification required for survival under conditions of nutrient stress and at low temperatures, which is linked to translation. Given the capacity of ycfD to act as an oxygen/nutrient sensor, it may act to fine-tune the rate of translation under conditions of nutrient deprivation.

7.3. *YcfD* mediates resistance to peptidyl transferase ribosomal antibiotics

Chapter 3 described the role of ycfD-catalysed hydroxylation of the ribosomal protein L16 in increasing resistance to the PTC inhibitors chloramphenicol and clindamycin/lincomycin, which interfere with binding/positioning of the tRNA. The PTC inhibitors were shown to selectively inhibit the growth of *E. coli ycfD* knock-out, but not wild-type strain. The sensitivity of the *E. coli* wild-type strain to chloramphenicol increased by inhibiting the activity of ycfD by removing co-factors required for the ycfD catalytic activity (Fe(II) and O_2), and by using a ycfD inhibitor, IOX1, which inhibited ycfD with IC_{50} of 38 μM *in vitro*. The finding that *ycfD* deletion increases chloramphenicol sensitivity is significant, because oxygen dependence of chloramphenicol sensitivity *via* a 2OG oxygenase-mediated modification has not previously been reported. The therapeutic potential of a post-translational modification mediating antibiotic resistance provides an opportunity for medicinal targeting of ribosome-modifying enzymes, thereby affecting the activity of the ribosome. In co-treatment with an existing antibiotic, such as chloramphenicol, a small molecule inhibitor of a ribosome-modifying enzyme would achieve a potentiated antibiotic effect. The clinical implication of inhibition of modified ribosomes extends beyond microbiology; the existence of human homologues of ycfD, NO66 and MINA53, suggests that human ribosomes may be unhydroxylated under certain conditions.

7.4. Biochemical characterisation of a thermophilic 2OG oxygenase ycfD$_{RM}$

Prior work in the Schofield Lab suggested that the *E. coli* ycfD.L16 complex is not amenable to crystallisation, therefore alternative means were pursued, including a bioinformatic search for a thermophilic homologue of ycfD, which was predicted to possess improved crystallographic properties for crystallisation in complex with the substrate.

A ycfD homologue, named ycfD$_{RM}$, was identified in the thermophilic bacterium *Rhodothermus marinus*. YcfD$_{RM}$ was characterised as a thermophilic 2OG oxygenase catalysing the hydroxylation of R82 on the ribosomal protein L16$_{RM}$. Intact protein MS analyses revealed that in cells, the L16$_{RM}$ hydroxylation levels decreased linearly with increasing temperature, and that the absolute extent of hydroxylation depended on oxygenation of growth medium, which is limited by oxygen solubility at high temperatures. The results of thermal denaturation experiments, including DSF and CD, as well as MALDI-MS based kinetic assays, demonstrated that protein stability, as well as Fe(II) and 2OG stability, are not major limiting factors for ycfD$_{RM}$ activity below 84°C. In contrast, the $K_m^{app}(O_2)$ for ycfD$_{RM}$ was 100±20μM, which is in the range of oxygen solubility at 70°C, consistent with an oxygen-sensing role of ycfD$_{RM}$ which leads to the temperature-dependent decrease in levels of L16$_{RM}$ hydroxylation observed between 60-75°C in *R. marinus* cells.

The identification of a thermophilic oxygenase whose activity in cells is limited by oxygen availability raised the possibility that oxygen sensing mechanisms in bacteria, in particular in thermophiles, extends beyond the currently established [4Fe-4S] cluster- and heme-containing oxygen sensors to encompass members of the 2OG-dependent oxygenases. YcfD$_{RM}$ is the first thermophilic 2OG oxygenase described; its increased thermostability relative to the mesophilic homologue ycfD$_{EC}$ provides evidence that 2OG oxygenases could in principle be engineered to function at high temperatures, which may be of industrial importance for 2OG oxygenase-catalysed transformations.

7.5. Structural studies on ycfD$_{RM}$

YcfD$_{RM}$, biochemically characterised as a thermophilic homologue of ycfD$_{EC}$ that functions as a L16 hydroxylase in *R. marinus*, was then use for crystallographic studies. Crystallographic studies on ycfD$_{RM}$.L16$_{RM}$ complex led to the determination of the first substrate complex crystal structure for a ROX, determined to 3.0 Å. The structure revealed the key interactions required for binding of the L16$_{RM}$ fragment, which include a substrate-binding pocket lined by residues D118, M120, Y137 and S208. A substantial conformational rearrangement of the side chains of substrate-binding residues was observed upon substrate binding, suggesting an induced-fit mechanism. In particular, the side chain of

R212 was observed to form a key salt-bridge interaction with the $L16_{RM}$ residue E80 (2.6 Å), ~8.2 Å away from its position observed in the *apo* structure, providing structural rationale for the substrate selectivity of $ycfD_{RM}$ for E80 described in Chapter 4.

Importantly, the work on the thermophilic ROX has informed structural studies on evolutionarily related ROX, NO66 and MINA53, for which substrate structures have been obtained since the completion of the work. The structural understanding of ROX catalysis was applied to the finding, described in Chapter 6, that MINA53 hydroxylates rpL27a in mammalian cells.

7.6. Biochemical characterisation of MINA53 activity in cells

The work described in Chapter 6 demonstrated that mouse 60S ribosomal protein rpL27a is hydroxylated in mammals in a manner that depends on the catalytic activity of MINA53. RpL27a hydroxylation was found in all mouse organs analysed, as well as in cancer cell lines and in clinical non-Hodgkin lymphoma samples, and did not vary significantly with hypoxic treatment. However, reduction of rpL27a hydroxylation was observed in a number of clinically identified *MINA53* mutations from the COSMIC database of cancer mutations. Structural analysis suggested that mutations occur more frequently at structurally important regions of MINA53, including the βIV-βV insert in the core fold of MINA53. This provided a demonstration that cancer tissues may differ from normal tissues in the status of post-translational hydroxylation of their ribosomal proteins.

The identification of inhibiting clinical mutations suggested that rpL27a hydroxylation levels could be used as a cancer mark, and in the future for selective inhibition by ribosomal antibiotics. The work presented herein demonstrates that it is possible to selectively inhibit modified ribosomes; an inhibitor of unhydroxylated rpL27a could therefore, at least in principle, be active against the sub-set of tumours with inactivating mutation(s) of *MINA53*, but not normal tissue. Future work should therefore focus on identifying a selective inhibitor of unhydroxylated eukaryotic ribosomes which could be applied for treatment of cancers harbouring deactivating *MINA53* mutations. The same approach could be applied to other ribosome modifications (to rRNA, ribosomal proteins, and ribosome-associate factors) that are different cancer in compared to normal cells.

7.7. Future work

The work described in this thesis identified ycfD as a ribosomal protein hydroxylase. Preliminary characterisation of the role of ycfD in translation was described on the global translation level. A more precise role of L16 hydroxylation in translation could in the future be characterised by performing detailed kinetic studies on isolated ribosomes, focusing on the individual steps of the translation cycle (initiation, elongation, termination).

Furthermore, oxygen dependence of ycfD-catalysed L16 hydroxylation could be further investigated by performing detailed mass spectrometric analyses of L16 hydroxylation in *E. coli* cells grown at different pO_2 in order to establish the statistical significance of the capacity of ycfD to function as an oxygen sensor in cells.

A more selective inhibitor of ycfD will be required to further investigate the link between L16 hydroxylation and chloramphenicol sensitivity using a chemical probe. The therapeutic potential of a post-translational modification mediating antibiotic resistance is very appealing, because it provides an opportunity for medicinal targeting of ribosome-modifying enzymes, for example ycfD, which may be more amenable to small-molecule inhibition than the ribosome itself. The work presented in this thesis identified a potentially useful hydroxyquinoline scaffold for achieving ycfD inhibition in cells. The C2 position on the hydroxyquinoline scaffold of IOX1 was identified as a potential anchor point for further chemical modifications without disrupting the binding to ycfD. Future work could focus on medicinal chemistry validation of hydroxyquinoline scaffold as a lead for ycfD inhibition in cells.

Ultimately, a selective inhibitor of unhydroxylated human ribosomes will be required to validate the proposal that growth of cancer tissues harbouring deactivating mutations in *MINA53* gene could be selectively inhibited over normal tissue. Cycloheximide was identified as a small molecule which binds in the proximity of the hydroxylation site on rpL27a. Its activity in the *MINA53* KO MEF model could be investigated, and medicinal chemistry efforts invested in developing analogues which would achieve selective inhibition of cell growth for *MINA53* KO, but not WT cells.

8. Materials and methods

Contents

8.1. Reagents for molecular biology 276

8.2. Growth media 277

8.3. Polyacrylamide-Gel Electrophoresis (PAGE) 278

8.4. Standard procedures 280

8.5. Molecular biology procedures 281

8.6. Cloning procedures 282

8.7. Microbiological techniques 289

8.8. Protein production 291

8.9. Protein purification 293

8.10. $YcfD_{RM}$ crystallisation 295

8.11. Biophysical techniques 296

8.11.1. Differential scanning fluorimetry 296

8.12. Ribosome Purification 298

8.13. Functional studies on $ycfD_{EC}$ 301

8.14. Liquid chromatography-mass spectrometric studies of ribosomal proteins 305

8.15. LC-MS/MS protein analysis 308

8.16. Peptide synthesis 310

8.17. Enzyme activity studies 311

8.1. Reagents for molecular biology

8.1.1. 6 × DNA loading buffer

Reagent	Per 10 mL
Bromophenol Blue	25 mg
Xylene cyanol FF	25 mg
Glycerol	3.3 mL
ddH_2O	6.7 mL

8.1.2. 50 × TAE running buffer

Reagent	Per 1 L
Tris-HCl	242 g
Acetic acid (glacial)	57.0 mL
EDTA (0.5 M)	100 mL

8.1.3. DNA electrophoresis gel (1.0% w/v)

Reagent	Per gel
Agarose	0.4 g
TAE buffer (1 ×)	40 mL
SYBR Safe™ DNA gel stain	4.0 μL

8.1.4. Competent cells used for transformation

Strain	Phenotype
XL10-Gold	Tet^rΔ(mcrA)*183* Δ (*mcrCB-hsdSMR-mrr)173 endA1supE44 thi-1 recA1 gyrA96 relA1 lac* Hte [F' *proAB* $lacI^qZ\Delta M15$ Tn*10* (Tet^r) Amy Cam^r]
BL21-Gold(DE3)	*E. coli* B F^- ompT hsdS(r_b^- m_b^-) *dcm*+ Tet^r *gal* Δ (DE3) *endA* Hte

8.1.5. Other *E. coli* strains used

Strain	Phenotype
BW25113*	*E. coli* F⁻ *Δ(araD-araB)567 ΔlacZ4787*(::rrnB-3) *λ⁻ rph-1 Δ(rhaD-rhaB)568 hsdR514*
JW1114*	*E. coli F⁻ Δ(araD-araB)567 ΔlacZ4787*(::rrnB-3) *λ⁻ ΔycfD788::kan rph-1 Δ(rhaD-rhaB)568 hsdR51*

8.2. Growth media

8.2.1. Miller Luria Bertani (LB) medium*

Reagent	Per 1 L
Bacto Tryptone	10.0 g
Yeast Extract	5.0 g
Sodium Chloride	10.0 g

*To make LB Agar plates, 15 g of agar was added prior to autoclaving.

8.2.2. 2 × Tryptone/Yeast extract (2TY) medium

Reagent	Per 1 L
Bacto Tryptone	16.0 g
Yeast Extract	10.0 g
Sodium Chloride	5.0 g

8.2.3. M9 minimal salts

Reagent	Per 1 L
$Na_2HPO_4 \cdot 7H_2O$	64.0 g
KH_2PO_4	15.0 g
NaCl	2.5 g
NH_4Cl	5.0 g

8.2.4. M9 minimal medium

Reagent	Per 1 L
ddH_2O	700 mL
M9 minimal salts	200 mL
$MgSO_4$ (1 M)	2 mL
Glucose (20%)	20 mL
$CaCl_2$ (1 M)	100 µL

8.2.5. Antibiotics

Unless stated otherwise, the final concentrations of antibiotics were: Kanamycin 30 µg/mL, Ampicillin, 100 µg/mL.

8.3. Polyacrylamide-Gel Electrophoresis (PAGE)

8.3.1. SDS-PAGE separating gel (12%)

Reagents	Per gel
Acrylamide (30% w/v)	2 mL
Tris-HCl (1.5 M, pH 8.8)	1.25 mL
SDS (10% w/v)	50 µL
H_2O	1.57 mL
Ammonium persulphate	50 µL
TEMED	4 µL

8.3.2. SDS-PAGE stacking gel

Reagent	Per gel
Acrylamide (30% w/v)	0.5 mL
Tris-HCl (0.5 M, pH 6.8)	1.25 mL
SDS (10% w/v)	50 µL
H_2O	3.14 mL
Ammonium persulphate	50 µL
TEMED	5 µL

8.3.3. 2× SDS-PAGE loading buffer

Reagent	Per 20 mL
Tris-HCl (0.5 M, pH 6.8)	2 mL
Bromophenol blue	0.10 g
SDS (10 %)	4 mL
Glycerol	2.4 mL
β-mercaptoethanol	1 mL
ddH_2O	10.6 mL

8.3.4. 10× SDS-PAGE running buffer

Running Buffer	Per 1 L
Tris-HCl	34.0 g
Glycine	144.0 g
SDS	10.0 g

8.3.5. Native-PAGE separating gel

Reagents	Per gel
Acrylamide (30% w/v)	1.67 mL
Tris-HCl (0.75 M, pH 8.8)	3.33 mL
ddH_2O	0.83 mL
Ammonium persulphate	30 µL
TEMED	10 µL

8.3.6. Native-PAGE stacking gel

Reagents	Per gel
Acrylamide (30% w/v)	0.3 mL
Tris-HCl (0.5 M, pH 6.8)	0.7 mL
H_2O	1.0 mL
Ammonium persulphate	15 µL
TEMED	5 µL

8.3.7. 2 × Native-PAGE Loading Buffer

Reagents	Per 10 mL
Tris-HCl (0.5 M, pH 6.8)	1.0 mL
Bromophenol blue	0.05 g
Glycerol	1.2 mL

8.3.8. 10 × Native-PAGE running buffer

Reagents	Per litre
Tris-HCl	30.3 g
Glycine	144.0 g

8.4. Standard procedures

8.4.1. Polyacrylamide gel electrophoresis

The purity of proteins was analysed by PAGE. Gels were prepared using 70 mm × 100 mm glass plates, with 0.75 mm spacers. TEMED and freshly prepared ammonium persulphate were added to gel mixures immediately prior to casting. The separating gel was cast first with the addition of a separate layer of water to ensure a level surface. When polymerized, water was removed and the stacking gel was cast. Samples were prepared by mixing equal volumes with sample loading buffer (2×). For denaturing PAGE analysis, samples were incubated at 95˚C for 3 min, and gels run on a mini-PROTEAN Tetra Electrophoresis System (Bio-Rad) at a constant potential of 180 V. Native gel electrophoresis was conducted at 4°C at constant voltage of 80 V, unless otherwise indicated. Following electrophoresis, the gels were stained for 15 min with InstantBlue™ Coomassie® stain (Expedeon).

8.4.2. pH Measurements

The pH of solutions was determined using a Jenway pH Meter 3305, with an Aldrich glass/calomel combination electrode. Calibration was carried out between pH 4.0-7.0 or 7.0-10.0 immediately

before use, with buffer solutions of phthalate (pH 4.0), phosphate (pH 7.0) and borate (pH 10.0) from Fischer Scientific. Electrodes were stored in a 4 M potassium chloride solution.

8.4.3. Centrifugation

Samples of volume less than 1.5 mL were centrifuged at room temperature using an accuSpin™ benchtop centrifuge (Fischer Scientific) at 3,400 rpm for 5 min. Cell growth media and cell lysates were centrifuged in a Beckman Avanti J-25 (8,500 rpm: JA-10 rotor; 15,000 rpm: JA-16.250 rotor; 24,000 rpm: JA-25.50 rotor). Proteins were concentrated using a Beckman Allegra 21R centrifuge (4,000 rpm: SX4250 swing bucket) at 4°C. Ultracentrifugation was performed in SW28 and SW40Ti rotors (Beckman Coulter) in an Optima L-90 K ultracentrifuge (Beckman Coulter).

8.4.4. Dot blotting

1 μL of individual UPLC-separated protein fractions, collected in 1 or 3 min intervals, was spotted onto a nitrocellulose membrane (Hybond ECL, 0.2 μm). The membrane was then blocked using PBS with 0.05% Tween 20 containing 5% defatted milk for 30 min at room temperature. Primary antibody (monoclonal anti-rpS23, MCA3433Z; AbD Serotec), was diluted 1:000 in blocking buffer and the membrane was incubated for 90 min at room temperature. After washing, the membrane was incubated with the secondary antibody for 20 min at room temperature in blocking buffer. HRP-labeled secondary antibody was revealed by the ECL system by 30 min exposure.

8.5. Molecular biology procedures

Standard sterile practices were followed throughout, using a Heraeus® laminar flow hood when necessary. Media and equipment were sterilised by autoclaving at 121°C for 20 min. Solutions of IPTG, antibiotics and other heat-labile compounds were sterilised by filtering through 0.22 μm filters (Minisart®, Sartorius Stedim) instead of autoclaving.

8.5.1. Agarose gel electrophoresis

DNA samples were analysed and purified by gel electrophoresis of the DNA mixture on 1% agarose gels prepared with SYBR Safe™ DNA gel stain (Invitrogen). Samples were loaded after mixing with 6× loading buffer and run in TAE buffer on an M200 electrophoresis system (Bio-Rad) at a constant potential of 90 V until desired fragment separation was achieved. Gels were visualised by a UV transilluminator (Gel logic 200) and the size of the fragments was referenced to either a 1 kb or a 100 bp DNA ladder (New England Biolabs).

8.5.2. DNA recovery from Agarose Gels

After electrophoresis, the DNA was recovered from agarose gels using the QIAEX^{TM} gel extraction kit (Qiagen) following manufacturer's instructions.

8.5.3. DNA Quantification

Concentration of DNA was determined by measurement of absorbance at 260 nm using a NanoDrop ND-1000 UV/Vis Spectrophotometer.

8.5.4. DNA Sequencing

Plasmid DNA sequencing was carried out by the Oxford DNA Sequencing Service Facility in the Department of Biochemistry, Oxford University, managed by Geneservice. T7 promoter and terminator sequence primers were used.

8.6. Cloning procedures

8.6.1. Cloning of ycfD$_{RM}$

The *YcfD$_{RM}$* gene (ID: 8566662) was cloned from genomic DNA (gDNA) of *R. marinus* (strain DSM 4252), extracted using a DNeasy Blood and Tissue Kit (Qiagen), following manufacturer's instructions. The polymerase chain reaction (PCR) was performed in 0.5 mL Eppendorf tubes, using

BIO-X-ACT™ LongMix (Bioline) DNA polymerase with the following components (3 conditions were tested):

	1	2	3
Template (184 ng/μL)	1 μL	1 μL	1 μL
Primers (2.5 μM)	5 μL	5 μL	5 μL
$MgCl_2$	0 μL	0.5 μL	0 μL
DMSO	0 μL	0 μL	1.2 μL
BIO-X-Act	12.5 μL	12.5 μL	12.5 μL
ddH_2O	6.5 μL	6.25 μL	5.3 μL

Thermal cycling was performed in a Techne Genius thermal cycler with the following parameters:

Step	Temperature (°C)	Time (sec)	
Initial denaturation	96.0	30	
Denaturation	96.0	30	40 cycles
Annealing	64.0*	30	
Extension	72.0	75	
Final extension	72.0	10 min	

*Annealing temperatures tested were 60°C, 64°C and 67°C.

Purified PCR products were ligated into an pGEM®-T Easy Vector (Promega) using the following conditions:

Component	Vol (μL)
2× Rapid ligation buffer	5
pGEM®-T Easy Vector	1
PCR Product	4
T4 DNA ligase	1
Total volume	10

Reaction mixtures were mixed by pipetting and incubated for 3 h at room temperature. 1 μL of ligation mixture was transformed into 30 μL of ultra-competent XL10-Gold *E. coli* following the manufacturer's guidelines. Transformed cells were plated onto LB/ampicillin/IPTG/X-Gal plates, using the blue/white screening to identify successful transformants. Single colonies were picked, grown in 2TY medium (5 mL) containing 100 μg/mL ampicillin and their DNA extracted using the standard protocol.

A *ycfD$_{RM}$* clone containing no mutations was used for further PCR reactions in order to attach flanking restriction sites.

The following combinations of restriction sites were created: NdeI/HindIII, NdeI/SalI, NdeI/SacI and BamHI/SalI using the following forward and reverse primers:

Primer	Primer DNA sequence
ycfD$_{RM}$_F_NdeI	ATAAACATATGCAGCTTCCCGA
ycfD$_{RM}$_F_BamHI	TCAAAGGATCCATGCAAGCTTC
ycfD$_{RM}$_R_HindIII	ATCATAATAAGCTTCAGCGTTTGC
ycfD$_{RM}$_R_SalI	GAATAGTCGACTCAGCGTTTGC
ycfD$_{RM}$_R_SacI	AGAATAGAGCTCAGCGTTTGC

The PCR was carried out using the following conditions:

Component	Vol (μL)
ycfD$_{RM}$ (pGEM-T Easy construct, 600 ng/mL)	0.3
Primers (10 μM)	5
BIO-X-Act	12.5
ddH_2O	7.2
Total volume	25

Thermal cycling was as follows:

Step	T (°C)	Time (sec)	
Initial denaturation	96.0	30	
Denaturation	96.0	30	40 cycles
Annealing	64.0	30	
Extension	72.0	75	
Final extension	72.0	10 min	

Products of successful PCR reactions were gel-purified (8.5.1&8.5.2) and re-ligated into pGEM®-T Easy Vector. Plasmids carrying inserts with correct DNA sequence, and target plasmids, were separately double-digested using appropriate restriction enzymes using the following conditions:

	Restriction enzymes	NEB Buffer	BSA added*
ycfD$_{RM}$_NdeI/HindIII	NdeI/HindIII	2	no
pET28a	NdeI/HindIII	2	no
ycfD_ NdeI/SacI	NdeI/SacI	4	yes
pET28a	NdeI/SacI	4	yes
ycfD$_{RM}$_BamHI/SalI	BamHI/SalI	4	no
pGEX-6-P1	BamHI/SalI	4	no

Component	Quantity
DNA	1 μg
10× NEB Buffer	2.5 μL
100× BSA	0.25 μL*
Restriction enzyme 1	1 μL
Restriction enzyme 2	1 μL
ddH_2O	to 25 μL

*BSA was only added where indicated ('yes')

Vector-insert pairs containing complementary restriction sites were ligated using T4 DNA ligase (NEB) for 3 h at room temperature using the following reaction conditions:

Component	Quantity
10× DNA ligase buffer	2 μL
Vector DNA	50 ng
Insert DNA	to 20 μL
T4 DNA ligase	1 μL

Ligation products (1 μL) were transformed into ultra-competent *E. coli* XL10 Gold (30 μL) and plated onto LB agar plates containing 30 μg/mL kanamycin (pET28a and FLAG-pET28a) or 100 μg/mL ampicillin (pGEX-6P-1).

8.6.2. Cloning of $L16_{RM}$

$L16_{RM}$ (ID: 8567487) was cloned from *R. marinus* gDNA (strain DSM 4525) using the following primer pair, which contained NdeI/XhoI restriction sites:

	Forward primer	Reverse primer	Restriction site
$L16_{RM}$	CATATGTTAATGCCCCGACGTA	CTCGAGTTAGCGTCCCTGC	NdeI/XhoI

The PCR was performed using the following conditions:

Component	Vol (µL)
Template (180 ng/µL)	1
Primers (10 µM each)	5
Bio-X-Act Long	12.5
ddH_2O	6.5

Thermal cycling was as follows:

Step	T (°C)	Time (sec)	
Init. denaturation	96.0	30	
Denaturation	96.0	30	40 cycles
Annealing	63.0	30	
Extension	72.0	30	
Final ext.	72.0	10 min	

$L16_{RM}$ was sub-cloned into the pET28a vector after double digestion of the PCR product and pET28a vector by NdeI/XhoI restriction nucleases followed by ligation using T4 DNA ligase as described above.

8.6.3. Site-directed mutagenesis

Site directed mutagenesis of plasmids involved PCR and was conducted in a Techne Genius thermal cycler programmed for initial hold at 95°C for 2 min before the addition of *Pfu*Turbo® polymerase to the reaction mixture. This ensured denaturation of the double stranded DNA and prevented non-specific binding of primers. Unless otherwise indicated, the reaction mixture consisted of the following components:

Component	1	2	3	4
*Pfu*Turbo buffer (10×)	5 µL	5 µL	5 µL	5 µL
Template (10 ng/µL)	1 µL	5 µL	1 µL	5 µL
Forward Primer (10 µM)	1 µL	1 µL	1 µL	1 µL
Reverse Primer (10 µM)	1 µL	1 µL	1 µL	1 µL
dNTP Mix (25 mM each)	1 µL	1 µL	1 µL	1 µL
DMSO	2.5 µL	2.5 µL	5 µL	5 µL
$MgCl_2$	0.75 µL	0.75 µL	0.75 µL	0.75 µL
ddH_2O	41.75 µL	37.75 µL	39.25 µL	35.25 µL
*Pfu*Turbo	1 µL	1 µL	1 µL	1 µL

DNA primers were diluted to 100 µM in ddH_2O according to manufacturer's instructions, then sub-diluted 1:10 for use in reaction mixtures. For primer pairs with predicted weak/moderate secondary structure, reaction conditions 1-2 were initially tested. For primer pairs with predicted strong secondary structure, reaction conditions 3-4 were initially tested. The following thermal cycling parameters were used:

Step	Temperature (°C)	Time	
Initial denaturation	95.0	2 min	
Denaturation	95.0	30 s	
Annealing	55.0	1 min	16 cycles
Extension	68.0	7 min	

The following primer pairs were used to achieve the desired mutations of $ycfD_{RM}$, $L16_{RM}$, $ycfD_{EC}$ and MINA53 (all in pET28a vector, Novagen, 5.4 kb):

$ycfD_{RM}$	Forward primer	Reverse primer
H133A	GGGCGCATATCGCCAACTACGAC	ACGTCGTAGTTGGCGATATGCGC

$L16_{RM}$	Forward primer	Reverse primer
G84C	CCGAAGTCCGTATGTGCAAAGGGA AGGGCTCG	CGAGCCCTTCCCTTTGCACATACGGAC TTCGG

$YcfD_{EC}$	Forward primer	Reverse primer
D110A	CCGGACTGGCGTATTGATGCGCTG ATGATTTCTTTTTCTG	CAGAAAAAGAAATCATCAGCGCATCA ATACGCCAGTCCGG
D110E	CCGGACTGGCGTATTGATGAACTG ATGATTTCTTTTTCTG	CAGAAAAAGAAATCATCAGTTCATCA ATACGCCAGTCCGG
D110R	CTACCGGACTGGCGTATTGATCGT CTGATGATTTCTTTTTCTG	CAGAAAAAGAAATCATCAGACGATCA ATACGCCAGTCCGGTAG
M112A	GACTGGCGTATTGATGATCTGGCG ATTTCTTTTTCTGTACCCGG	CCGGGTACAGAAAAAGAAATCGCCAG ATCATCAATACGCCAGTC
Y129A	GTCGGCCCGCATCTCGATCAGGCG GACGTGTTTATCATTCAGGG	CCCTGAATGATAAACACGTCCGCCTGA TCGAGATGCGGGCCGAC
Y129F	CGGCCCGCATCTCGATCAGTTTGA CGTGTTTATCATTCAGGG	CCCTGAATGATAAACACGTCAAACTGA TCGAGATGCGGGCCG
R140K	CAGGGTACCGGACGTAAACGCTGG CGAGTGGGCG	CGCCCACTCGCCAGCGTTTACGTCCGG TACCCTG
L160R	CACATCCGGATCTGCGTCAGGTCG ATCCGTTC	GAACGGATCGACCTGACGCAGATCCG GATGTG
S199A	GAAAATGCGATGAACTATGCGGTG GGCTTTCGCGCGCC	GGCGCGCGAAAGCCCACCGCATAGTTC ATCGCATTTTC
I211R	GCCAAATACGCGGGAACTGCGTAG TGGATTTGCCGATTATG	CATAATCGGCAAATCCACTACGCAGTT CCCGCGTATTTGGC
H227C	GTTTATATCCCAGTCACGTTGCGA ACTGGATATCGCGCCG	CGGCGCGATATCCAGTTCGCAACGTGA CTGGGATATAAAC
R251Q	CAAGAGATGGATAAACTGCAGGA GATGATGCTCGAATTG	CAATTCGAGCATCATCTCCTGCAGTTT ATCCATCTCTTG
R251Q_L255R	CTGCAGGAGATGATGCGTGAATTG ATCAACCAGC	GCTGGTTGATCAATTCACGCATCATCT CCTGCAG
R251Q_L255R_N259D	GATGATGCGTGAATTGATCGATCA GCCGGAACACTTTAAGC	GCTTAAAGTGTTCCGGCTGATCGATCA ATTCACGCATCATC
R251Q_L255R_N259D_ P261R	GTGAATTGATCGATCAGCGTGAAC ACTTTAAGCAATGG	CCATTGCTTAAAGTGTTCACGCTGATC GATCAATTCAC

MINA53	Forward primer	Reverse primer
G175R	CCCGCAGGATCTCAGCGTCTGCCG	CATAATGGGGCGGCAGACGCTGAGAT
P341R	CACAGACTCCCCCGTTACTCTGCGG	CTCCCGCAGAGTAACGGGGGAGTCTG

After the PCR reaction was complete, an aliquot was removed from the PCR product (10 μL) and discarded; the remainder of the product (40 μL) was treated with DpnI restriction nuclease (NEB, 1 μL) in the presence of NEB Buffer 4 (4.5 μL) for 3 h at 37°C. The mixture was cooled on ice then transformed (1 μL) into XL10 Gold ultracompetent cells and plating on an Agar plate containing the appropriate antibiotic.

8.7. Microbiological techniques

8.7.1. Transformation of competent cells

For the transformation of competent cells, the cells were thawed and resuspended on ice. Appropriate volume of β-mercaptoethanol was added (2.5% vol/vol) and the cells were incubated on ice with occasional gentle swirling before plasmid DNA (1 μL) was added. The samples were incubated on ice for 30 min before heat shock at 42°C (duration depending on type of competent cell used). After the heat shock, cells were incubated on ice for 2 min, then 900 μL of pre-warmed (37°C) 2TY medium was added, and cells were incubated at 37°C with shaking (200 rpm) for 1-2 h. After incubation, the samples were centrifuged (3,000 rpm for 4 min) and 700 μL of the supernatant was removed, while the pellet was resuspended in the remaining supernatant. 200 μL of resuspended cells was plated on a pre-warmed agar plate containing appropriate antibiotics, and incubated at 37°C overnight.

8.7.2. Incubations

Bacterial plate cultures were prepared using solid LB medium with the appropriate antibiotic. Following inoculation, plates were incubated overnight at 37°C in a Gallenkamp Duostat Incubator. Successful cultures were stored in the dark at 4°C. Liquid cultures of 2TY medium and the appropriate antibiotic were incubated at 37°C, with shaking at 200 rpm in a New Brunswick Scientific Innova® 44 incubator. Medium-scale starter cultures (100 mL) were inoculated from a single bacterial colony or from a glycerol stock, and grown in 500 mL Erlenmeyer flasks.

8.7.3. Blue/white colony screening

Cloning into pGEM®-T Easy vector (Promega) allowed detection of successful ligations based on colour. For this type of positive selection, LB agar was supplemented with 0.5 mM IPTG, 80 μg/mL bromo-chloro-indolyl-galactopyranoside (XGal) and 100 mg/mL ampicillin. White-colonies were picked for further manipulation.

8.7.4. Production of BL21 (DE3) cells containing pET28a constructs

The pET28a constructs containing the gene of interest were used to produce N-terminal His_6-tagged fusion proteins. All constructs were initially transformed into *E. coli* XL10-Gold (Stratagene) and streaked onto LB agar plates containing 30 μg/mL kanamycin. A 5 mL bacterial culture was grown in 2TY media containing 30 μg/mL kanamycin; the plasmid DNA containing the desired construct was purified using GeneJET Plasmid Miniprep Kit (Thermo Scientific) and transformed into *E. coli* BL21 (DE3) cells (Stratagene).

8.7.5. Preparation of glycerol stocks

Bacterial cultures were preserved by the addition of 75% (v/v) glycerol (350 μL) to cell culture (350 μL) with gentle mixing. The glycerol stocks were stored at -80°C.

8.7.6. Growth of *Rhodothermus marinus*

R. marinus (strain *R-10 DSM 4252*, ATCC® 43812™) was grown in #1599 broth (ATCC) . The broth contained yeast extract (2.5 g), tryptone (2.5 g), nitrilotriacetic acid (100.0 mg), $CaSO_4 \cdot 2H_2O$ (40.0 mg), $MgCl_2 \cdot 6H_2O$ (200.0 mg), 0.01 M Ferric citrate (0.5 ml) , NaCl (10.0 g), Trace Element Solution (0.5 ml), Phosphate Buffer Solution (100.0 ml), ddH_2O (900.0 ml). The pH of the solution was adjusted to pH 7.2 with NaOH and autoclaved at 121°C. Phosphate buffer was sterilised by filtration through 0.2 μm filter and added separately to the sterile medium. Trace Element Solution contained nitrilotriacetic acid (12.8 g), $FeCl_2 \cdot 4H_2O$ (1.0 g), $MnCl_2 \cdot 4H_2O$ (0.5 g), $CoCl_2 \cdot 6H_2O$ (0.3 g), $CuCl_4 \cdot 2H_2O$ (50.0 mg), $Na_2MoO_4 \cdot 2H_2O$ (50.0 mg), H_3BO_3 (20.0 mg), $NiCl_2 \cdot 6H_2O$ (20.0 mg), ddH_2O (1000 ml). Phosphate Buffer Solution was made of KH_2PO_4 (5.44 g), $Na_2HPO_4 \cdot 12H_2O$ (43.0 g) and ddH_2O (1000 mL). The pH was adjusted to pH 7.2.

Growth of *Rhodothermus marinus* was induced in the broth from a glycerol stock. The organism was grown in a horizontally-oriented 50 mL Falcon tube in a water bath maintained at 70°C for two days, until OD_{600} reached ~0.7 and no longer increased (measured using a 1 mL sample of bacterial culture compared to a 1 mL reference solution of initial growth medium in a Novaspec II spectrophotometer,

Pharmacia). A starter culture was induced by 1:33 dilution into 100 mL of ATCC #1599 broth, pre-incubated at 70°C. The growth of 100 mL culture ('strong starter culture') was monitored for 4 days.

When the strong starter culture reached maximum OD_{600}, a large-scale growth was induced by 1:100 dilution into ATCC #1599 broth in two types of flasks: 600 mL PYREX® Erlenmeyer flasks and 1 L plastic baffled flasks. They were grown simultaneously in an Innova® 44 shaking incubator (New Brunswick Scientific) overnight. In separate batches, cultures were grown at 60, 65, 70, 75, 77 and 80°C. Before harvest, the cultures were cooled to room temperature, then centrifuged at 8,000 rpm for 10 min at 4°C. The supernatant was decanted and the pellet weighed and frozen at -80°C in re-sealable plastic bags.

8.8. Protein production

8.8.1. Expression trials

Selected clones of genes of interest were transformed into *E. coli* BL21(DE3) and plated onto LB agar plates containing the appropriate antibiotics. A single colony was picked from each and grown in 2TY medium (100 mL) containing the appropriate antibiotics overnight to produce a 'starter culture'.

Starter cultures were induced from colonies from freshly prepared LB-agar. 10 mL of 2TY medium containing 30 μg/mL kanamycin (pET28a-based constructs) or 100 μg/mL ampicillin (pGEX-6P-1-based constructs) was inoculated with 100 μL of starter culture. When the culture reached OD_{600} ~0.6, temperature was adjusted to a desired level and protein expression was induced by a defined concentration of IPTG. The following conditions were tested:

Temperature (°C)	IPTG (mM)		Time (h)
18	0.1	0.5	18
30	0.1	0.5	18
37	0.1	0.5	4

The cultures were then centrifuged for 10 min at 4,000 rpm in a Beckman Coulter Allegra X-30 R centrifuge, and the resulting pellet was resuspended in 1 mL of 50 mM Tris-HCl, pH 7.5 and

sonicated 3 × 30 s with 30 s pauses in a Vibra Cell VCX 500 sonicator with a 13 mm probe, on ice. Lysate was centrifuged at 14,800 rpm for 20 min in a Sigma 1-14 benchtop centrifuge (SciQuip). The supernatant (soluble phase) was removed and stored, while the pellet (insoluble phase) was resuspended in 100 μL 50 mM Tris-HCl, pH 7.5, then diluted 100-times in the same buffer.

Soluble and insoluble phases (5 μL) were diluted in 2×SDS-PAGE loading buffer (5 μL) and 4 μL was loaded on a 15% SDS-PAGE gel and electrophoresed at constant voltage (180 V) for 50 min.

8.8.2. Large scale growths

A 100 mL culture of *E. coli* BL21(DE3) containing the appropriate expression plasmid in 2TY media was produced as a 'starter culture' of large scale growths. Large scale recombinant protein production was performed using 2 L PYREX® narrow mouth graduated Erlenmeyer flasks containing 600 mL of 2TY media with appropriate antibiotics (30 μg/mL kanamycin for pET28a-based constructs). Flasks containing growth medium were inoculated with the starter culture (1:100) and incubated with shaking (180 rpm) at 37°C in an Innova® 44 shaker (New Brunswick Scientific) until an OD_{600} of ~0.6 was reached. The temperature was then readjusted to the desired expression temperature, followed by the addition of the desired amount of IPTG (18°C / 0.5 mM IPTG for $ycfD_{EC}$, and 30°C / 0.1 mM IPTG for $ycfD_{RM}$). The cultures were then incubated for 16 - 18 h post-induction. Cells were harvested by centrifugation at 8000 rpm for 20 min at 4°C (Section 8.4.3). The resulting pellet was weighed and stored in a sealable plastic bag at –80°C until further processing.

8.8.3. Cell lysate preparation for protein purification

The cell pellet was resuspended in an appropriate purification buffer (volume in mL approximately 5-times the mass of cell pellet in grams) to which approximately 1 mg of DNAseI, 5 mg of lysozyme and 1 tablet of protease inhibitor cocktail (Roche) were added. The suspension of cells was placed in an iced-water bath and lysed by sonication, applying ~10 × 30 s bursts interrupted by a 45 s cooling period, using a Vibra Cell VCX 500 sonicator with a 13 mm probe, on ice. The lysate was clarified by

centrifugation at 22,000 rpm for 20 min at 4°C (Section 8.4.3). The supernatant was decanted and sequentially filtered through 0.45 μm and 0.22 μm filters (Omnipore™, Millipore U.K.)

8.9. Protein purification

Fast Protein Liquid Chromatography (FPLC) was used for large-scale protein purifications, using Äkta™ FPLC systems (GE Healthcare). All buffers were freshly prepared and filtered through a 0.22 μm filters (Omnipore™, Millipore U.K.) before use. All protein purification was performed at 4°C. Sample volumes up to 10 mL were loaded using a Superloop™ (GE Healthcare); larger samples were loaded using the FPLC pump. Purification procedures were analysed by monitoring absorbance at 280 nm with a UPC-900 monitor. Samples were collected using a Frac-920 fraction collector connected to the FPLC interface.

8.9.1. Buffers used for protein purification

Purification Buffers	Tris-HCl	NaCl	Imidazole	Glycerol	EDTA	$NiSO_4$	pH
Binding (HisTrap)	50 mM	0.5 M	20 mM	5%	-	-	7.5
Binding (Ni-TED)	25 mM	0.5 M	-	-	-	-	7.5
Elution (HisTrap)	50 mM	0.5 M	0.5 M	5%	-	-	7.5
Elution (Ni-TED)	25 mM	0.5 M	0.5 M	-	-	-	7.5
Strip	-	-	-	-	100 mM	-	-
Charge	-	-	-	-	-	50 mM	-
Size exclusion	50 mM	0.3 M	-	5%	-	-	7.5
Binding (AEX)	25 mM	-	-	-	-	-	8.0
Elution (AEX)	25 mM	1.0 M	-	-	-	-	8.0

8.9.2. Standard purification of His_6-tagged proteins

Purification of His_6-tagged proteins was carried out using either a 1 mL or 5 mL (settled bed volume) HisTrap™ nickel-affinity chromatography column (GE Healthcare). 1 mL columns were handled manually and 5 mL columns were handled *via* FPLC systems. The resins were prepared by washing with 5 column volumes (CV) of MilliQ water, followed by 5 CV of charge buffer and 10 CV of binding buffer. Cell lysate was applied at a 1 mL/min flow-rate. The column was then washed with approximately 20 CV of binding buffer. When all residual products of bacterial lysis were eluted from

the column by binding buffer, the column was washed with 3 CV wash buffer and the protein of interest was eluted with a gradient of elution buffer (0-100% over 6 CV). Eluted proteins were collected and fractionated; fractions containing the highest levels of pure protein (as determined by A_{280} and SDS-PAGE) were concentrated to 2.5 mL using Amicon Ultracell-10k concentrators (Millipore) in a Beckman Coulter Allegra™ X-30R centrifuge using a SX4400 rotor. A PD-10 Column (GE Healthcare) was used to buffer-exchange the protein into an appropriate buffer, the protein concentrated to ~10 mg/mL and stored at -80°C. The column was then washed with 10 CV of MilliQ water and 4 CV of strip buffer to remove Ni(II) ions and stored in EtOH (20% vol/vol) at 4°C.

Alternatively, His_6-tagged proteins were purified using a Protino® Ni-TED Ni(II)-affinity column (Mackerey-Nagel). All procedures were as described above, but using Ni-TED buffer A and Ni-TED elution buffer (washing step was avoided).

8.9.3. Preparative size exclusion chromatography

Further purification was achieved by size exclusion chromatography using a 300 mL Superdex™ 75 size exclusion chromatography column (GE Healthcare) which had been equilibrated with 1 CV of size exclusion buffer. Concentrated fractions (maximum volume 3 mL) were loaded on the column at 1 ml/min and eluted with 1 CV of size exclusion buffer; 5 mL fractions were collected. Eluted fractions were analysed by A_{280} and SDS-PAGE gel electrophoresis; fractions containing protein of >90% purity were concentrated to appropriate concentration (~ 10 mg/mL) using Amicon Ultracell-10k concentrators (Millipore).

8.9.4. Ion exchange chromatography

For crystallography, His_6-ycfD_{RM} was purified by Ni(II) affinity chromatography followed by ion exchange chromatography. After the Ni(II)-affinity chromatography step, the protein was buffer exchanged into Binding Buffer (AEX) using a PD-10 Column (GE Healthcare). A Source 15Q strong anion exchange column (GE Healthcare) was equilibrated with the same buffer. The sample was applied onto the column using a Superloop™ (GE Healthcare) at 3 mL/min. The column was then washed with 10 CV Binding Buffer (AEX) after which, the protein was eluted with a linear gradient

0-75% of Elution Buffer (AEX) over 25 CV. 5 mL fractions were collected. Eluted fractions were analysed by A_{280} and SDS-PAGE gel electrophoresis; fractions containing protein of >90% purity were concentrated to 2.5 mL, and buffer-exchanged into 50 mM HEPES, pH 7.5 using a PD-10 Column (GE Healthcare) and concentrated to ~10 mg/mL using Amicon Ultracell-10k concentrators (Millipore).

8.10. $YcfD_{RM}$ crystallisation

Broad crystallisation screens for $ycfD_{RM}$.Mn(II).FG2216 and $ycfD_{RM}$.Mn(II).NOG.$L16_{RM}$ complexes were conducted using crystallisation screens JCSG, INDEX, MIDAS, PACT Premier, PEG ION, NATRIX and SALT-RX (Hampton Research) adopting a sitting drop technique by the vapour diffusion method (drop size: 200-300 nL) at 20°C in 96-well Intelliplates (Art Robbins). Crystallisation hits were optimised as detailed in Chapter 5. Crystals were cryoprotected by transferring to a solution of mother liquor and harvested (by Dr. Michael A. McDonough) in nylon loops and before cryo-cooling by plunging into liquid N_2. Crystals were screened in-house using a Rigaku FR-E+ Superbright copper rotating anode diffractometer equipped with Osmic HF optics and a Saturn 944+ CCD detector interfaced with the CrystalClear 1.0 software (Rigaku) to identify crystals diffracting to high resolution. High resolution data were collected at 100 K using single crystals at Diamond Light Source beamline I04 with a Pilatus 6M-F detector. Data were then integrated and scaled using HKL3000.

8.10.1. Data processing

Crystal structures of $ycfD_{RM}$.Mn(II).FG2216 and $ycfD_{RM}$.Mn(II).NOG.$L16_{RM}$ were solved by molecular replacement (MR) using the $ycfD_{EC}$ structure as the search model (Schofield Laboratory, unpublished data). Structural refinement was carried out with PHENIX with iterative rebuilding of the models using COOT until R_{work}/R_{free} converged to final values 0.152/0.212 and 0.184/0.221 for $ycfD_{RM}$.Mn(II).FG2216 and $ycfD_{RM}$.Mn(II).NOG.$L16_{RM}$, respectively (data processing carried out by Dr. Michael A. McDonough).

8.11. Biophysical techniques

8.11.1. Differential scanning fluorimetry

DSF was used to determine T_m values for $ycfD_{RM}$ and $ycfD_{EC}$ and MINA53. The experiment was performed using a MiniOpticon™ Real-Time PCR Detection System (Bio-Rad) connected to MJ Mini Thermal Cycler (Bio-Rad). Sypro® Orange (Invitrogen) (1 μL) was diluted in 2.5 mL of 50 mM HEPES, pH 7.5, then 182.5 μL of this Sypro® Orange solution was mixed with 50 μL of enzyme (10 μM). 50 μL of the mixture was added to individual wells in 6×8 white, non-skirted PCR plates (StarLab) and mixed thoroughly with a pipette. If inhibitors were added, 49 μL of mixture was mixed with 1 μL of inhibitor in DMSO (at 50× final concentration). Plates were sealed with StarSeal polyolefin sealing tape. The chamber was cooled to 4°C before inserting the sample plate and the readings of fluorescence were taken between 4-95°C, increasing the temperature linearly in 1°C steps (at 1°C/min). The FAM (492 nm) and ROX (610 nm) filters were used for excitation and emission, respectively. Software automatically performed global minimum subtraction. The T_m was calculated using the Boltzmann equation $y = LL + \frac{UL - LL}{1 + e^{\frac{T_m - x}{a}}}$, where LL and UL correspond to minimum and maximum intensities and *a* corresponds to the slope of the curve within T_m.

8.11.2. Circular dichroism

For investigation of protein stability using circular dichroism (CD), measurements were acquired using a Chirascan CD spectrometer (Applied Photophysics) with a Peltier temperature-controlled cell holder. All experiments were performed in a 0.1 cm path length cuvette using 0.1-0.2 mg/ml protein in 10 mM sodium phosphate buffer (pH 8.0). Data were recorded from 240 to 185 nm, at 0.5 nm intervals, and each data point was averaged for 3 s. Spectra were base-line corrected and smoothed using the Savitzky-Golay filter. Data recorded in the 185-240 nm range were analysed using DichroWeb and the CONTIN deconvolution method was used to estimate secondary structural content using reference set 6 (4-5). In thermal denaturation experiments, spectra were recorded every

2-5°C, with a 5 min equilibration time at each temperature. Temperature-dependent changes in secondary structure were monitored by CD at 222 nm, and normalised data fit to a Boltzmann sigmoidal curve in GraphPad Prism (version 5.02) in order to determine T_m values. At the end of the melt, reversibility was determined by returning to the start temperature (10°C) and comparing the CD spectrum with the spectrum obtained prior to denaturation.

8.11.3. MALDI-TOF-MS

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) was carried out using a Waters Micromass™ MALDI micro MX™ mass spectrometer in the positive ion reflectron mode. The following parameters were used: Laser energy 180-220, Pulse 1950, Detector 2350, Suppression 700. The instrument was calibrated using 1 μL PepMix4 (containing Trypsinogen, Enolase and BSA) in 9 μL CHCA matrix. Intensity of the background-subtracted peaks corresponding to the parent $[M+H]^+$and hydroxylated peak $[M+16+H]^+$ was transformed into activity data by determining fraction hydroxylation. 1 μL of the diluted assay mixture was mixed with 1 μL of recrystallized α-Cyano-4-hydroxycinnamic acid (CHCA, Laser BioLabs) and analysed by MALDI-MS.

8.12. Ribosome Purification

8.12.1. E. coli

The following buffers were used for ribosome purification:

***E. coli* lysis buffer**	**Concentration**
Tris-HCl pH 7.5	20 mM
$MgCl_2$	10 mM
NH_4Cl	100 mM
EDTA	0.5 mM
β-Mercaptoethanol	3 mM
Protease inhibitor cocktail (Roche)	1 tablet

***E. coli* sucrose cushion buffer**	**Concentration**
Sucrose	1.1 M
Tris-HCl pH 7.5	20 mM
$MgCl_2$	10 mM
NH_4Cl	500 mM
EDTA	0.5 mM
β-Mercaptoethanol	3 mM
Protease inhibitor cocktail (Roche)	1 tablet

Buffer C	**Concentration**
Tris-HCl pH 7.5	50 mM
$MgCl_2$	5 mM
KCl	25 mM

E .coli cell pellets were resuspended in lysis buffer, lysed by sonication using a Vibra Cell VCX 500 sonicator with a 13 mm probe, on ice (10 s bursts at 10 s intervals for 10 min) and centrifuged at 22,000 rpm for 45 min (section 8.4.3). The top three-quarters of the supernatant were centrifuged again, before the first ultracentrifugation step (38,000 rpm for 3 h in a Beckman SW Ti40 rotor using Optima L-90 K ultracentrifuge). The pellet was resuspended in 1 mL lysis buffer and applied onto a sucrose cushion layer (11 mL). Ribosomes were purified by a single round of ultracentrifugation

through the sucrose cushion (36,000 rpm for 16 h in Beckman SW Ti40 rotor). The ribosomal pellets were resuspended in 0.7 mL Buffer C and the ribosomal protein extraction step was performed immediately.

In order to extract ribosomal proteins for mass spectrometry, to 0.7 mL of resuspended ribosomal pellet, 5 μL 1 M $MgCl_2$ and 0.7 mL of EtOH were added to precipitate the ribosomes, then centrifuged at 7,000 rpm for 2 min at 4°C in a tabletop centrifuge (Microfuge® 22R, Beckman Coulter). The pellet was resuspended in 250 μL of buffer C; 25 μL of 1 M $MgCl_2$ followed by 550 μL of glacial acetic acid were added, before the solution was incubated at 4°C for 45 min with occasional mixing by inversion. RNA was pelleted by centrifugation at 10,000 rpm for 10 min (Microfuge® 22R, Beckman Coulter), while the supernatant was retained for acetone precipitation of ribosomal proteins. A solution formed by the addition of 4 volumes of acetone to the supernatant was incubated at -20°C for at least 2 h, then centrifuged for 5 min in a microcentrifuge. The pellet was washed twice with ice-cold acetone and dried by exposure to air overnight.

8.12.2. R. marinus

***R. marinus* lysis buffer**	**Concentration**
HEPES pH 7.5	20 mM
$MgCl_2$	10 mM
NH_4Cl	100 mM
EDTA	0.5 mM
β-Mercaptoethanol	6 mM
Benzamidine	0.1 mM
Phenyl-methyl-sulfonyl fluoride (PMSF)	0.5 mM

***R. marinus* sucrose cushion buffer**	**Concentration**
Sucrose	1.1 M
HEPES pH 7.5	20 mM
$MgCl_2$	10 mM
KCl	100 mM
EDTA	0.5 mM
β-Mercaptoethanol	6 mM
Benzamidine	0.1 mM
Phenyl-methyl-sulfonyl fluoride (PMSF)	0.5 mM

R. marinus cell pellet was resuspended in lysis buffer and lysed using the Precellys 24 bead-beater tissue homogeniser (Bertin Technologies). 6 cycles of 6,500 rpm for 20 sec, followed by a cooling period of 1 min on ice using glass beads (Precellys Glass Kit 0.1/0.5 mm in 7 mL container) was used to lyse the cells.

The lysate (2 mL per centrifuge tube) was applied onto sucrose cushion (11 mL) and centrifuged in SW40Ti rotor at 38,000 rpm for 18 h. Ribosomal proteins were then extracted as described above.

8.12.3. Human tissue culture samples

Lysis buffer A	Concentration
Tris-HCl pH 8.0	20 mM
$MgCl_2$	10 mM
NaCl	100 mM
KCl	20 mM
Sucrose	5 % (w/vol)
β-Mercaptoethanol	6 mM
Triton-X 100	1 % (vol/vol)
Protease Inhibitor Cocktail (Roche)	1 tablet

Human tissue culture cells (HeLa and HEK 293T cell types) were grown in standard DMEM medium (Sigma Aldrich) supplemented with foetal bovine serum (10%), streptomycin, penicillin, and glutamax solution. Cells were grown to ~90% confluency before harvesting. The medium was aspirated and cells washed with ice-cold PBS. The cells were detached with trypsin and centrifuged at 4,500 rpm in a Beckman Allegra 21R centrifuge (SX4250 swing bucket rotor) at 4°C. Trypsin solution was removed and the cell pellet washed again with ice-cold PBS, before lysis using a Dounce homogeniser in buffer A. The solution was centrifuged at 22,000 rpm for 20 min (Section 8.4.3). The supernatant was centrifuged again at the same speed before ultracentrifugation for 3 h at 38,000 rpm. Ribosomal protein extraction procedure was as for prokaryotic samples.

8.12.4. Mouse liver samples

Mus musculus liver samples (kindly provided by Dr. T. Bishop, Wellcome Trust Centre for Molecular Physiology, University of Oxford) were homogenised in lysis buffer A using a Ultra-Turrax® T25 high-performance dispenser (IKA®, Staufen, Germany). When the homogenate reached appropriate consistency, it was centrifuged in a Beckman Allegra 21R centrifuge (4,500 rpm, SX4250 swing bucket rotor) for 5 min at 4°C and processed as above.

8.13. Functional studies on $ycfD_{EC}$

8.13.1. Growth assays in rich and minimal media

For growth comparisons of BW25113 and JW1114, standard 2TY medium or minimal medium M9 (supplemented with 0.2% glucose) were used at 37°C; *E. coli* colonies were grown on LB agar plates, then a bacterial slab was resuspended in medium and the OD_{600} adjusted to 0.05. 150 µl of bacterial cultures were transferred to a clear, flat bottomed 96-well plate 3370 (Costar). Growth of JW1114 and BW25113 at 37°C was monitored by measuring absorbance at 584 nm (A_{584}) in a Polarstar Optima (BMG Labtech) plate reader at 1 h intervals for 20 h.

8.13.2. Bulk translation rate assays

For global translation rate assays, JW1114 and BW25113 strains were grown in triplicate in minimal M9 media as described above, before labelling with 10 µCi/ml ^{35}S-Methionine (PerkinElmer) at 37°C without shaking. At the times indicated, 1 mL samples were removed from cultures and placed on ice before the addition of 400 µg BSA and 200 µl of CCl_3COOH_{aq} (50% vol/vol, Sigma). Samples were mixed and incubated on ice for 30 min prior to centrifugation at 13,000 rpm for 10 min. Protein pellets were washed twice in cold ethanol, before partially drying and solubilising in 200 µl of 8M Urea. Technical duplicates of 5 µl were then added to 2 mL OptiPhase scintillant (PerkinElmer) before measuring ^{35}S-Methionine incorporation by scintillation counting.

8.13.3. Antibiotic sensitivity studies

Glycerol stocks of BW25113 and JW1114 were used to inoculate LB-agar plates containing no antibiotic (BW25113) or kanamycin (30 μg/mL, JW1114). Slabs of bacterial lawns were used to inoculate starter cultures, which were adjusted to OD_{600} 0.05±0.005, then diluted 20-fold in 2TY medium. Starter cultures (200 μL) were placed in a clear, flat bottomed 96-well plate 3370 (Costar) and antibiotics were added to desired final concentration (1 μL each). Plates were sealed with clear adhesive film (Costar) and A_{584} was measured in a PolarStar Optima (BMG Biotech) for 18 h with vigorous shaking (200 rpm) at 400 s intervals.

8.13.4. Hypoxic antibiotic sensitivity studies on agar plates

Anaerobic growth of *E. coli* BW25113 and JW1114 was performed in an $Invivo_2$ Hypoxic Workstation (Ruskinn life sciences), attached to a Ruskinn gas mixer module to control pO_2. Agar plates containing colonies of BW25113 and JW1114 were used for liquid starter cultures. OD_{600} of starter cultures was adjusted to 0.06 and a serial 10-fold dilution (7 dilutions and a blank) was plated on LB agar plates containing 2% glucose and a desired concentration of chloramphenicol. The plates were transferred into the hypoxic chamber which had been incubated at 0.1% O_2 overnight. Colonies were grown overnight at 37°C with an active humidifier module.

8.13.5. Hypoxic growth of BW25113 in liquid culture

Cultures were grown in an $Invivo_2$ Hypoxic Workstation (Ruskinn life sciences), attached to a Ruskinn gas mixer module to control pO_2; all media was pre-equilibrated overnight for each pO_2 (0.1, 5, 10, 20% O_2) condition. Bacterial colonies on an agar plate were introduced into the workstation *via* an airlock system and a single colony was used to inoculate 100 mL of 2TY. An overnight starter culture was produced, then used to inoculate a growth culture (600 mL) by 1:100 dilution into 2TY medium. By this two-fold passage of cells through pre-equilibrated media, the introduction of O_2 leakage into the chamber was minimised. Cells were grown to OD_{600} 0.6-0.7, then removed from the

chamber and harvested by centrifugation at 8,000 rpm for 10 min (8.4.3). The pellets were frozen at -80°C until further processing.

8.13.6. Ribosome profiles

For the comparative study of BW25113 and JW1114 ribosome profiles, 5 g of bacterial cell pellet was dissolved in lysis buffer D in the presence of lysozyme (~ 1 mg). Cells were sonicated for 10 min (5 s bursts with 10 s pause at 60% amplitude) in a Vibra Cell VCX 500 sonicator with a 13 mm probe, on ice, and centrifuged at 23,000 rpm for 20 min (Section 8.4.3). Sucrose gradients were prepared by layering sucrose buffers of decreasing density (50-5%) with intermittent freezing. Lysate (5-30 A_{260} units) was loaded onto a 10-25/10-50% (wt/vol) sucrose gradients prepared in buffer A and ultracentrifuged in SW28 rotor at 28,000 rpm at 4°C for 3 h (section 8.4.3).

Lysis buffer D	Concentration
Tris-HCl, pH 7.4	10 mM
KCl	70 mM
$MgCl_2$	10 mM

To study the effects of chloramphenicol on ribosome assembly of BW25113/JW1114, the *E. coli* strains BW25113 and JW1114 were grown overnight, then sub-diluted into fresh 2TY medium and grown at 37°C in 100 mL 2TY medium until the OD_{600} reached 0.2. Chloramphenicol (7 μg/ml) was added. The cultures were incubated for a further 1 h at 37°C. The cells were harvested by centrifugation at 9,000 rpm for 10 min at 4°C (8.4.3) and resuspended in 1 mL lysis buffer E; lysozyme and DNase I (Amresco, GE Healthcare) were added to final concentrations of 1 mg/ml and 20 U/ml, respectively. The cells were lysed by three freeze-thaw cycles; cells were incubated at -80°C for 15 min, then thawed in ice-cold water for 30 min. The freeze-thaw cycle was repeated twice, followed by centrifugation at 4°C for 20 min. The supernatant was diluted twofold with lysis buffer F. Lysate was loaded onto a 10-25% (wt/vol) sucrose gradient prepared in lysis buffer F and ultracentrifuged at 28,000 rpm in an SW28 rotor (8.4.3) for 3 h at 4°C. Gradients were analysed on a Teledyne ISCO UV analyser using upward displacement with 60% sucrose solution (in water). Chart recorder speed was 60 cm/h, sensitivity 1, noise filter 0.5.

Lysis buffer E	Concentration
Tris-HCl, pH 8.0	50 mM
KCl	60 mM
NH_4Cl	60 mM
$MgCl_2$	6 mM
β-Mercaptoethanol	6 mM

Lysis buffer F	Concentration
Tris-HCl, pH 8.0	10 mM
KCl	60 mM
NH_4Cl	60 mM
$MgCl_2$	12 mM
β-Mercaptoethanol	6 mM

8.13.7. LC-MS Analysis of TCA cycle intermediates in *E. coli*

Starter cultures of BW25113 and JW1114 were grown from single colonies grown on agar plates containing appropriate antibiotics one day before conducting the experiment; antibiotics were added to the starter culture as appropriate. Bacterial growth was induced by inoculating 2TY medium (100 mL) with 2 mL of starter culture; growth was monitored by removing 1 mL aliquots and measuring the OD_{600}. Aliquots were removed at indicated time points and centrifuged (4,200 rpm for 15 min), 8.4.3), before discarding the supernatant and storing the cell pellet at -80°C until further processing.

Cells were lysed by resuspension in 80% MeOH, and sonication in an iced ultrasonic bath XB3 (Grant) for 20 min in 30 sec pulses. After sonication the samples were centrifuged and supernatant was collected. The solvent was evaporated using a Concentrator Plus speedvac (Eppendorf). Solid residue was resuspended in mobile phase (0.1% FA, 50µL) and analysed by UPLC-Quattro MS (Waters). MS analysis was performed by Khalid al-Qahtani.

8.14. Liquid chromatography-mass spectrometric studies of ribosomal proteins

8.14.1. HPLC studies on *E. coli* BW25113 ribosomal proteins

Intact masses of ribosomal proteins, purified as described above, were recorded by liquid chromatography / mass spectrometry (LC/MS) using a ZMD (Waters®) single quadrupole mass spectrometer, interfaced with a Hewlett Packard® Series 1050 liquid chromatography and sample handling system. The protein sample (7 µl) was injected onto a Grace Vydac 214TP C4 (250 mm × 4.6 mm) reverse-phase HPLC column and eluted at 1 mL/min using a gradient system from Solvent A (0.1 % $HCOOH_{aq}$) to Solvent B (0.1% $HCOOH_{aq}$ in MeCN), as detailed in the table below. Samples were injected directly into the mass spectrometer. The following MS parameters were used: Polarity: ES+; Capillary voltage: 3,000 V; Sample cone voltage: 35 V; Extraction cone voltage: 1V; Desolvation temperature: 250°C; Source temperature: 100°C; Cone gas flow: 100 L/h; Desolvation gas flow (N_2): 830 L/h. Sodium formate was used to calibrate the instrument. Spectra were processed using MassLynx™ 4.0 (Waters Corporation) with the Maximum Entropy method (MaxEnt1) employed for deconvolution of intact mass (0.1 Da resolution).

Time (min)	%B	Time (min)	%B
0.0	15.0	**60.0**	38.5
10.0	23.0	**80.0**	41.0
12.0	25.0	**100.0**	42.5
15.0	26.0	**120.0**	45.0
20.0	30.0	**130.0**	47.0
25.0	33.0	**190.0**	95.0
30.0	34.0	**200.0**	95.0
40.0	35.0	**201.0**	15.0
45.0	35.5	**210.0**	15.0
50.0	36.5		

8.14.2. UPLC-ESI-MS analysis of ribosomal proteins

Samples of crude ribosomal proteins, purified as described above (section 8.12), were analysed by reversed phase ultra-performance liquid chromatography (RP-UPLC) and electrospray ionisation time-of-flight mass spectrometry (ESI-TOF MS) using a Water Acquity UPLC system connected

directly to a Waters LCT Premier or Synapt ESI-TOF-MS instruments. The method used a BEH C4 reversed phase column (2.1 × 50 mm, 1.7 μm particle size, 300 Å pore size, Waters), and a flow rate of 0.3 mL/min. The column was held at 40°C throughout the method. The column was equilibrated with solvent A (0.1% $HCOOH_{aq}$). Typically, 10 μL of ribosomal protein sample was injected onto the column and proteins eluted using a stepped gradient from solvent A to solvent B (0.1% $HCOOH_{aq}$ in MeCN). Optimization of the UPLC gradient used the conditions detailed in the table below.

The following MS parameters were used: Polarity: ES+; Capillary voltage: 3,000 V; Sample cone voltage: 35 V; Extraction cone voltage: 2.5 V; Desolvation temperature: 250°C; Cone gas flow rate: 10 L/h; Desolvation gas flow (N_2): 500 L/h. The mass spectra were acquired from 400 to 2400 *m/z* using MassLynx 4.1 software (Waters) and protein spectra deconvoluted using MaxEnt 1 with a range 3 to 30 kDa (0.1 Da resolution). Masses were confirmed using manual component analysis. Sodium formate was used for external calibration of the instrument. Leucine enkephalin (0.6 μg/mL, Sigma) was used as an internal calibrant for on-line mass correction using the Lock-Spray module (MassLynx).

Method A			**Method B**			**Method C**		
Time (min)	**% A**	**% B**	**Time (min)**	**% A**	**% B**	**Time (min)**	**% A**	**% B**
0	95	5	**0**	95	5	**0**	95	5
2.5	85	15	**2.5**	92	8	**2.5**	85	15
5	82	18	**5**	90	10	**5**	82	18
10	74	26	**10**	88	12	**10**	80	20
20	66	34	**20**	85	15	**20**	77	23
40	65	35	**40**	82	18	**30**	74	26
50	64	36	**50**	75	25	**40**	70	30
85	50	50	**65**	40	60	**50**	65	35
100	5	95	**66**	5	95	**60**	40	60
101	95	5	**70**	5	95	**70**	20	80
108	95	5	**71**	95	5	**71**	95	5
			80	95	5	**75**	95	5

8.14.3. Tryptic digestion of *E. coli* ribosomal proteins

Samples containing a mixture of ribosomal proteins were run on an SDS-PAGE gel (12 %), stained for 1 h using Coomassie stain and destained overnight. The relevant protein bands were cut out with a clean scalpel. Bands were cut into smaller pieces (1-2 mm^3) and placed in a 1.5 ml tube. 200 µl of the wash solution was added and the gel pieces rinsed overnight at room temperature while shaking. The wash solution was then removed and another 200 µl of wash solution added. The gel pieces were rinsed for 2-3 h at room temperature and centrifuged; the wash solution was removed. 200 µl acetonitrile was added and the gel pieces dehydrated for 5 mins at room temperature. Acetonitrile was removed from the sample and the gel pieces dried in a vacuum centrifuge for 2-3 min. 30 µl of DTT solution was added to gel pieces and left to reduce protein samples for 30 min at RT. The solution was subsequently removed. 30 µl of iodoacetamide solution was added and the samples were left to alkylate at RT for 30 min. The solution was then removed. 200 µl acetonitrile was added and the gel pieces were dehydrated for 5 min at RT, then the acetonitrile was removed and samples dried in a vacuum centrifuge. The gel pieces were rehydrated using 200 µl ammonium bicarbonate (100 mM) for 10 min at RT, and the solution was removed. 200 µl acetonitrile was added and the gel pieces dehydrated for 5 min at RT, before removing the solution and drying the gel pieces in a vacuum centrifuge (Concentrator plus, Eppendorf).

The trypsin reagent was prepared by adding 1 mL of ice-cold ammonium bicarbonate (50 mM) to 20 µg of trypsin to make a final concentration (20 ng/µL). 30 µL of the trypsin solution was added to the sample and rehydrated on ice for 10 min with occasional gentle mixing. The gel pieces were centrifuged and the excess trypsin solution removed. 5 µL of ammonium bicarbonate (50 mM) was added to the sample and the digestion carried out overnight at 37°C. After overnight incubation, 50 µL of ammonium bicarbonate buffer (50 mM) was added to the tube containing the gel pieces, which were left to incubate for 10 min with occasional gentle mixing. Supernatant was collected in a 1.5 mL tube. 50 µL of Extraction buffer I was added to the tube containing the gel pieces and incubated for 10

min with occasional gentle mixing. Supernatant was added to the collection tube. Another 50 µl of Extraction buffer I was added to the tube containing the gel pieces and left to incubate for 10 min with occasional gentle mixing. The supernatant was again added to the collection tube. The sample was then dried completely in a vacuum centrifuge, resuspended in 20-40 µl Extraction buffer II and analysed by MS/MS.

Wash solution	Concentration
MeOH	50 % (vol/vol)
CH_3COOH (glacial)	5 % (vol/vol)

Alkylating/reducing agents	Concentration
Dithiothreol (DTT)	10 mM in 25 mM ammonium bicarbonate
Iodoacetamide	50 mM in 25 mM ammonium bicarbonate

Extraction buffer I	Concentration
Acetonitrile	50 % (vol/vol)
HCOOH	5 % (vol/vol)

Extraction buffer II	Concentration
Acetonitrile	2 % (vol/vol)
HCOOH	5 % (vol/vol)

8.15. LC-MS/MS protein analysis

8.15.1. *E. coli* ribosomal proteins

The trypsin-digested *E. coli* ribosomal proteins were subjected to nano-ultra performance liquid chromatography tandem MS analysis (nano-UPLC-MS/MS) using a 75 µm-inner diameter × 25 cm C_{18} nanoAcquity™ UPLC column (1.7 µm particle size; Waters) and a 90 min gradient of 2–45% solvent B (solvent A: 0.1%$HCOOH_{aq}$; solvent B: 99.9% MeCN, 0.1% $HCOOH_{aq}$) on a Waters nanoAcquity UPLC system (final flow rate: 250 nL/min; 7000 p.s.i.) coupled to a Q-TOF Premier

tandem mass spectrometer (Waters) run in positive ion mode. MS analysis was performed in data-directed analysis (DDA) mode (MS to MS/MS switching at precursor ion counts greater than 10 and MS/MS collision energy 14 dependent on precursor ion mass and charge state). All raw MS data were processed using the PLGS software (version 2.3) including deisotoping and deconvolution (converting masses with multiple charge states to m/z = 1). The mass accuracy of the raw data was corrected using Glu-fibrinopeptide (200 fmol/µl; 700 nl/min flow rate; 785.8426 Da $[M + 2H]^{2+}$) that was infused into the mass spectrometer as a lock mass calibrant during analysis. MS and MS/MS data were calibrated at intervals of 30 s. MS/MS spectra (peak lists) were searched against the UniProtKB/Swiss-Prot database (Version 2010.07.16; 518,415 sequences) database using Mascot version 2.3.01 (Matrix Science) and the following parameters: peptide tolerance, 0.2 Da; ^{13}C = 1; fragment tolerance, 0.1 Da; missed cleavages, 2; instrument type, ESI-Q-TOF; fixed modification, carbamidomethylation (C); and variable modifications, deamidation (N,Q) and oxidation (M,D,K,N,P,R).

8.15.2. *R. marinus* ribosomal proteins

The *R. marinus* ribosomal protein samples were not separated by SDS-PAGE but directly digested with trypsin (37°C for either 15 min or overnight) and analysed using an Orbitrap Elite (Thermo Fisher Scientific, DE) connected to a UHPLC Proxeon EASY-nLC 1000 and an EASY-Spray nano-electrospray ion source. Peptides were trapped on an Acclaim PepMap® trapping column (100 µm i.d. × 20 mm, 5 µm C18) and separated on an EASY-spray Acclaim PepMap® analytical column (75 µm i.d. × 500 mm, RSLC C18, 2 µm, 100 Å). Solvent A consisted of 0.1% $HCOOH_{aq}$ and solvent B of MeCN/0.1% $HCOOH_{aq}$. Peptides were separated using a 3 h gradient of 7%-30% solvent B at a flow rate of 200 nL/min. Full scan MS spectra were acquired in the Orbitrap (350-1500 m/z, resolution 120,000, AGC target 1e6, maximum injection time 250 ms). CID and ETD spectra were acquired in the Ion Trap (resolution 7500, AGC cation target 3e4, AGC Anion target 2e5, maximum injection time 100 ms). The 20 most intense peaks in the full MS scan were selected for fragmentation using a data-dependent decision tree (DDDT). ETD fragmentation was used for charge states 3, 4 and

5 with m/z less than 750 and for all charge states greater than 5. CID fragmentation was used for all other peptides.

The raw data files generated were processed using MaxQuant software (Version 1.4.1.2), integrated with the Andromeda search engine. The maximum number of missed cleavages was set to be 4. Hydroxylation/oxidation of methionine-, lysine- and arginine-residues was allowed as a variable modification. All assignments of potential post-translational modifications were validated by manual inspection of spectra.

8.15.3. Non-denaturing mass spectrometry on ycfD-L16 complex

To investigate the stoichiometry of ycfD.L16 complex, the samples of the $ycfD_{EC}$(H277C)-$L16_{RM}$ (G84C) protein-protein complex were buffer exchanged into 200 mM ammonium acetate, pH 7.4 using a Micro Bio-Spin column (Bio-Rad) at 4°C to a final concentration 10 mg/mL.

Mass spectrometry measurements were performed on a modified Synapt G1 HDMS instrument (Waters) with the travelling-wave ion mobility cell replaced by an 18-cm drift cell with radial RF ion confinement and a linear voltage gradient to direct ions along the axis of transmission to the time-of-flight (TOF) mass analyser. The mass spectrometer was typically set to a source pressure of 5-7 mBar, capillary voltage of 1.4-1.7 kV, capillary nanoflow of 0.03-0.2 mBar and argon as collision gas with flow rate set to 5-8 mL/min (~6.6-6.7 e^{-2} mBar). Collision voltage ranged from 5 V. The sample and extraction cone and trap bias voltages, quadrupole profile, and collision gas pressure were optimized for maximal ion intensity of the target protein complex. The drift cell gas (He) was set to a pressure of ~1.6-1.8 Torr (40 ml/min flow rate). Pusher time, pulse width and TDC inhibit were set to 180, 7 and 7 µs, respectively.

8.16. Peptide synthesis

8.16.1. Synthesis of Linear peptides

Peptides were synthesised using a CS-Bio CS336X automated peptide synthesizer, using PL-rink-amide MBHA resin (CS Bio). Fluorenylmethyloxycarbonyl- (Fmoc)-protected amino acids were

activated using *N,N'*-diisopropylcarbodiimide (DIC)/1-hydroxybenzotriazole (HOBt) as the coupling agent. After cleavage of the *N*-terminal Fmoc-protecting group, the resin was filtered off and treated with ~4 mL of deprotection solution (95 % (v/v) CF_3COOH_{aq}, 2.5 % triisopropylsilane, 2.5 % water). After 3 h the volume was reduced to 1 mL under a nitrogen stream and the peptides were precipitated with ice-cold Et_2O. The peptides were pelleted by centrifugation at 4,000 rpm in Beckman Allegra 21R centrifuge (SX4250 swing bucket rotor) for 10 min at 4°C and the supernatant discarded. The resulting pellet was dissolved in 0.1% CF_3COOH_{aq} and purified by HPLC.

8.16.2. Synthesis of L16 cyclic peptide

For the synthesis of the $L16_{RM}$ cyclic peptide, the linear precursor was synthesised as described above. After cleavage of the *N*-terminal Fmoc-protecting group a solution of 150 mg of chloromethylcarbonyloxysuccinimide (ClAc-OSu) in 4 mL DMF was added to the resin and the mixture shaken for 3 h. The resin was filtered off and subsequently treated with 4 mL of deprotection solution (95 % (v/v) CF_3COOH_{aq}, 2.5 % triisopropylsilane, 2.5 % water). After 3 h the volume was reduced to 1 mL under a nitrogen stream and the peptides were precipitated with ice-cold Et_2O. The mixture was then centrifuged and the supernatant discarded. The solid was taken up in 1.5 mL of triethylammoniumacetate buffer (1 M, pH 8.5) and the pH readjusted to > 8 if necessary. In a microwave (Biotage Initiator) the mixture was heated to 80°C for 10 min and purified by HPLC.

8.16.3. Peptide purification

Preparative HPLC purifications were carried out using a Dionex Ultimate 3000 system with a Grace Vydac 218TP101522 column (0-45 % MeCN in 45 min, 0.1 % CF_3COOH_{aq}). Peptides were analysed using an Agilent 1200 series LC-MS system (6120 quadrupole MS) with a Waters Sunfire column.

8.17. Enzyme activity studies

8.17.1. Buffer Optimisation

In order to determine the optimal buffer conditions for $ycfD_{EC/RM}$ activity assays, a range of buffers were screened at their optimal pH values. All reaction components were initially diluted in 50 mM

Tris-HCl (pH 7.5), then diluted into a desired buffer to their final concentration. Fe(II) solution was prepared from $(NH_4)_2Fe(SO_4)_2$ dissolved in 20 mM HCl to make 400 mM stock solution, which was then diluted to the final concentration using MilliQ water. $ycfD_{EC/RM}$ were incubated with Fe(II), ascorbate, 2OG (concentrations detailed below) and substrate peptide fragment (10/50 μM, $ycfD_{EC/RM}$, respectively), diluted in the above buffers (50 mM) at the specified pH values, at 37°C/55°C ($ycfD_{EC/RM}$, respectively) for 30 mins. Reactions were quenched with 1% CF_3COOH_{aq} (20 μL) and analysed by MALDI-MS.

Component	[Final]	
Enzyme	2 μM	Enzyme mix
Fe^{2+}	100 μM	
Ascorbate	1 mM	
2OG	200 μM	Peptide mix
Peptide	10/50 μM	
Buffer	50 mM	

Buffer	Concentration	pH
MES	50 mM	6.0
BIS-TRIS	50 mM	6.5
BES	50 mM	7.0
HEPES	50 mM	7.0
HEPES	50 mM	7.5
HEPES	50 mM	8.0
EPPS	50 mM	8.0
TAPS	50 mM	8.5
TAPS	50 mM	9.0
CHES	50 mM	9.5
CAPSO	50 mM	10.0

8.17.2. $ycfD_{EC}$ single-point inhibition assays

$YcfD_{EC}$, ascorbate, 2-OG and $L16_{EC}$ peptide fragment (20-mer) solutions were made up in 50 mM HEPES buffer, pH 8.0. Enzyme mix containing the enzyme, Fe^{2+}, ascorbate and buffer was made up and aliquoted in 96-well plates. Inhibitors (1 μL) were dispensed in each well. Peptide mix containing

2OG, peptide and buffer was then added to the wells to initiate the reaction. After 10 min incubation at 37°C, reactions were quenched in a 1:1 ratio with 1% CF_3COOH_{aq}. When DMSO was present in the sample, TFA quenching was followed by addition of four volumes of 20 mM triammonium citrate.

	[Final]	[Stock]	Volume (µL)	
Enzyme	1 µM	20 µM	1	*Enzyme Mix*
Fe^{2+}	20 µM	200 mM	2	
Ascorbate	2 mM	20 mM	2	
Buffer	50 mM	50 mM	4	
Inhibitor	20/100 µM	400 µM/2 mM	1	
2OG	10 µM	100 µM	2	*Peptide Mix*
Peptide	100 µM	1 mM	2	
Buffer	50 mM	50 mM	6	

8.17.3. $YcfD_{RM}$ activity assays

In vitro hydroxylation assays on $ycfD_{RM}$ were performed using a MALDI-MS based assay. Unless indicated otherwise, the enzyme was incubated at 55°C with Fe(II), sodium ascorbate, 2OG and substrate peptide (^{72}KPVTKKPAEVRMGKGKGSVE^{91}, GL Biochem, China) in 50 mM HEPES buffer (pH 7.5) at concentrations indicated below. Reactions were quenched with an equal volume of 1% (v/v) CF_3COOH_{aq} and analysed by MALDI-MS. Substrate turnover was defined as $\%OH=I_{OH}/(I_{OH}+I_{non\text{-}OH})\times100$, where I_{OH} and $I_{non\text{-}OH}$ correspond to peak intensities of hydroxylated and non-hydroxylated peptides, respectively.

	µM	[Stock]	Vol (µL)	
Enzyme	1 µM	20 µM	1	Enzyme Mix
Fe(II)	100 µM	1 mM	2	
Ascorbate	1 mM	10 mM	2	
Buffer		50 mM	5	
2OG	200 µM	2 mM	2	Peptide mix
Peptide	50 µM	1 mM	1	
Buffer		50 mM	7	

8.17.4. Temperature dependence of ycfD$_{RM/EC}$ and PHD2 activity

The temperature dependence of ycfD$_{RM}$, ycfD$_{EC}$ and PHD2 was analysed by performing hydroxylation reactions in a GENIUS PCR cycler (Techne). Standard reaction conditions were used. L16$_{EC(71-90)}$, L16$_{RM\ (72-91)}$ and HIF CODD$_{(556-574)}$ peptides were used (at 10, 50 and 100 μM, respectively) to test the activity of ycfD$_{EC}$, ycfD$_{RM}$ and PHD2, respectively. The thermal cycler was pre-equilibrated at a desired temperature, then the enzyme mix and peptide mix were incubated separately for 3 min before being mixed together at the desired temperature. After 3 min, the reaction was quenched by addition of 20 μL 1% CF_3COOH_{aq} (PHD2 reactions were quenched by flash-freezing in liquid N_2). Results were analysed by MALDI-TOF MS (positive ionisation mode for ycfD$_{RM/EC}$ and negative ionisation mode for PHD2).

8.17.5. Substrate selectivity of ycfD$_{RM}$

Substrate selectivity of ycfD$_{RM}$ (10 μM) was analysed by performing the hydroxylation reaction with *R. marinus* and *E. coli* L16 fragment peptides (20-mer, 100 μM) at 70°C for 15 min in the presence of Fe(II), 2OG and ascorbate (at 100 μM, 200 μM and 1 mM, respectively). Reactions were performed in triplicate in the Thermomixer *comfort* (Eppendorf). Reactions were quenched by addition of equal volume of 1% CF_3COOH_{aq} and analysed by MALDI-MS.

8.17.6. O_2 dependence of recombinant ycfD$_{RM}$

To investigate the O_2 dependence of ycfD$_{RM}$, the assays were performed at 37°C in sealed vials which contained 500 μM peptide in 50 mM HEPES pH 7.5 buffer, pre-equilibrated using mass-flow controllers (Brooks Instruments, UK) under N_2/O_2 gas mixtures with variable O_2 content (from 0 to 50%). Other components of the reaction were added using a Hamilton syringe prior the assay to final concentrations of 100 μM Fe(II), 1 mM ascorbate, 500 μM 2OG and 2 μM ycfD$_{RM}$. Reactions were quenched with 1% $HCOOH_{aq}$ at 3 min. Substrate turnover was assessed by MALDI-MS.

8.17.7. K_m^{app} Determination Studies

Peptide name	Sequence
$L16_{RM}$ WT	KPVTKKPAEVRMGKGKGSVE
$L16_{EC}$ WT	KPITEKPLAVRMGKGKGNVE
$L16_{EC}$ A79E	KPITEKPLEVRMGKGKGNVE
CP1	KKPAEVRMGKGKGdC

K_m^{app} values for $ycfD_{RM}$ co-factors/co-substrates were determined for 2OG, Fe(II), 20-mer peptide fragments of L16 variants ($L16_{RM}$ WT, *E. coli* WT L16, $L16_{EC}$ A79E) and the cyclic $L16_{RM}$ WT (CP1) cyclic peptide. K_m^{app} values were determined for an individual cofactor by varying its concentration while keeping all other cofactors/co-substrates at saturating concentrations. Reactions were carried out in triplicate in HEPES buffer (50 mM, pH 7.5/8.0 for $ycfD_{RM/EC}$, respectively) at 55˚C/37°C ($ycfD_{RM/EC}$, respectively) and quenched after either 1 min (K_m^{app} for 2OG), or 2 min (K_m^{app} for Fe(II)).

In order to determine K_m^{app} values for peptide substrates with $ycfD_{RM}$, reactions were quenched after 30 seconds for peptide concentrations (20-100 μM), or 1 min peptide concentrations (100-2000 μM), in order to ensure that hydroxylation reactions were quenched in the linear part of the hydroxylation, whilst ensuring that the signal was at least 10% above baseline, by the addition of 20 μL CF_3COOH_{aq} (0.1%). In order to determine K_m^{app} values for L16 peptide fragment variants with $ycfD_{EC}$, reactions were carried out at 37°C and quenched after 1 min (peptide concentrations 50-300 μM), 2 min (peptide concentrations 300-1000 μM), or 4 min (peptide concentrations 1000-2000 μM).

8.17.8. IC_{50} value determination studies

For IC_{50} determination of $ycfD_{EC}$ inhibitors, the enzyme mixture, containing enzyme, Fe(II), ascorbate and buffer, was made up and aliquoted in 96-well plates. Inhibitors (1 μL) were dispensed into each well. The peptide mix, containing 2OG (20 μM), peptide (50 μM) and buffer, was then added to the wells to initiate the reaction. After incubation at 37˚C for 15 min, reactions were quenched using a equal volume of 1 % CF_3COOH_{aq} (vol/vol) and analysed by MALDI-MS. The log[inhibitor] *vs.*

inhibition—variable slope equation (GraphPad Prism version 5.02) was used to fit the data and obtain IC_{50} values.

8.17.9. Co-factor stability studies

To study the effect of temperature on the stability of co-factors and co-substrates, stock solutions were prepared for Fe(II), 2OG, ascorbate and $ycfD_{RM}$ at 1 mM, 2 mM, 10 mM and 20 μM, respectively. An aliquot (100 μL) of each was individually incubated at 70°C or 95°C in PCR cycler for the indicated lengths of time. Reaction mixtures were prepared (divided into the enzyme mix and the substrate mix as indicated below), systematically omitting each component.

	Component	[Final]	Volume	**1** (Enz)	**2** (Fe(II))	**3** (Asc)	**4**
Enzyme Mix	**Buffer**	50 mM	4 μL	+	+	+	+
	Enzyme	1 μM	2 μL	-	+	+	+
	Fe(II)	100 μM	2 μL	+	-	+	+
	Ascorbate	1 mM	2 μL	+	+	-	+
Substrate mix	**2OG**	200 μM	2 μL	+	+	+	-
	Peptide	100 μM	2 μL	+	+	+	+
	Buffer	50 mM	6 μL	+	+	+	+

The heated component was added to either the enzyme mix or the substrate mix and reaction mixtures were pre-incubated at 55°C before the mixtures were combined and allowed to react at 55°C for 5 min. Each reaction was conducted in triplicate. Reactions were quenched with 20 μL CF_3COOH_{aq} (1% vol/vol) and analysed by MALDI-MS.

8.17.10. Kinetic studies on MINA53 WT, G175R and P341R

Assay mixtures (final volume 20 μL in 50 mM HEPES, 500 mM NaCl, 5 % glycerol, 1 mM DTT, pH 7.5) contained MINA53 (1 or 10 μM), peptide fragments (100 μM), 2OG (200 μM), Fe(II) (100 μM) and ascorbate (1 mM). The reaction mixtures were incubated at 37°C for 30 min, then quenched with 20 μl of 1% $HCOOH_{aq}$. Results were analysed by MALDI- MS. Peptide sequences tested were i) rpL27a (GRGNAGGLHHHRINFDKYHP, predicted mass: 2283.5 Da), ii) ZFX3ALSQHLESALHKHRTITRAAR (predicted mass: 2396.7 Da) and iii) IFRD2: AFKEVLGSGMHHHLQNNELLR,) predicted mass: 2430.7 Da.

9. Appendix

Contents

Table 9.1: Library of 2OG mimetics screened against ycfD$_{EC}$ and ycfD$_{RM}$. 318

Table 9.2: Compounds from the BB-series of compounds tested against ycfD$_{EC}$ 328

Table 9.3: T_m shift analysis of ycfD$_{RM}$ stabilization by 2OG mimetics (plate 2OG1)..... 329

Table 9.4: T_m shift analysis of ycfD$_{RM}$ stabilization by 2OG mimetics (plate 2OG2)..... 330

Figure 9.1: Vector map of pET28a vector. 331

Figure 9.2: Vector map of pTRC99a vector 331

Figure 9.3: Phenotype MicroArray Single Panel Comparison Display. 332

Table 9.5: Refinement statistics of ycfD$_{RM}$ crystal structures..... 333

Table 9.6: Intra-chain hydrophobic contacts in ycfD$_{RM}$. 334

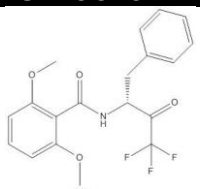
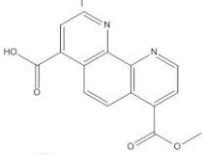
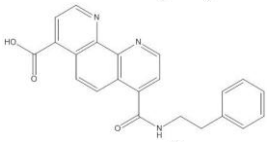
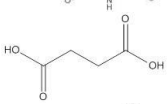
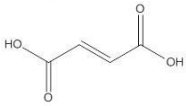
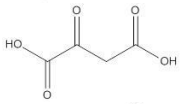
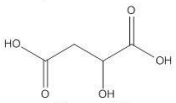
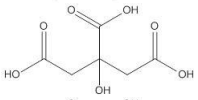
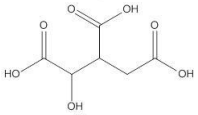
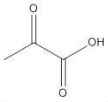
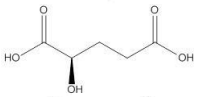
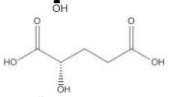
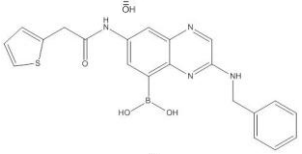
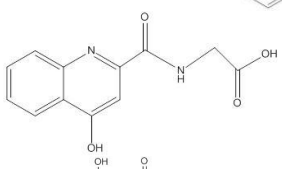
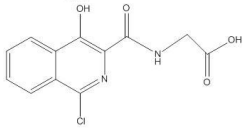
Table 9.7: Intra-chain hydrophobic interactions in ycfD$_{EC}$..... 338

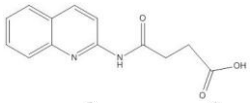
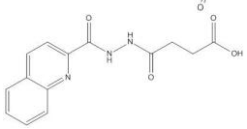
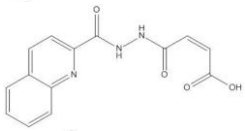
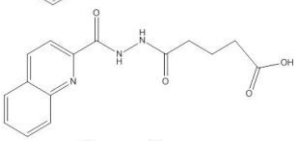
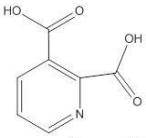
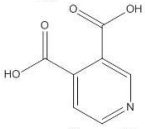
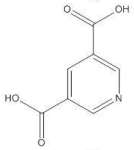
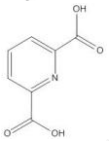
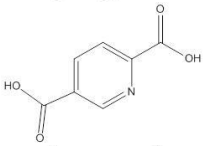
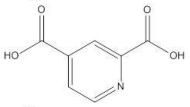
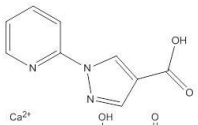
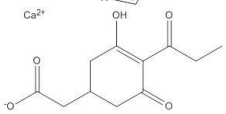
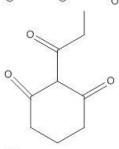
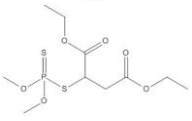
Table 9.8: List of salt-bridge interactions between protomers A and B in ycfD$_{RM}$ and ycfD$_{EC}$. 341

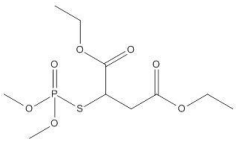
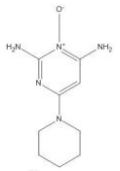
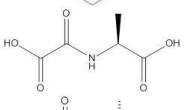
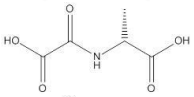
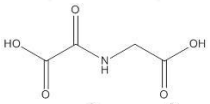
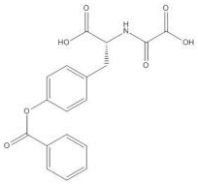
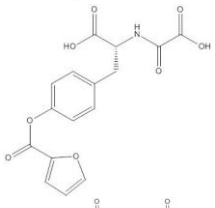
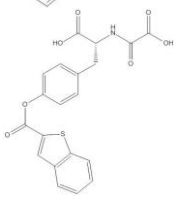
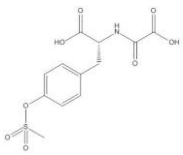
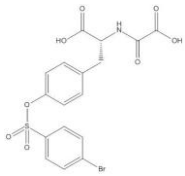
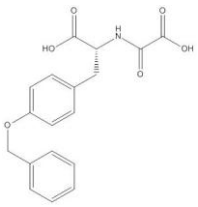
Table 9.9: List of hydrogen-bonding interchain interactions in ycfD$_{RM}$ and ycfD$_{EC}$..... 342

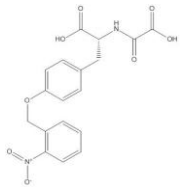
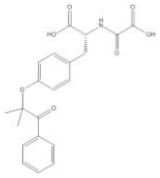
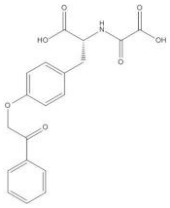
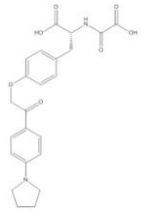
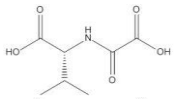
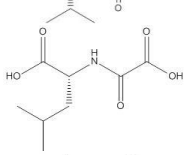
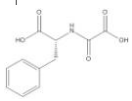
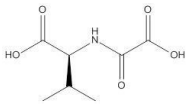
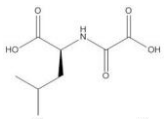
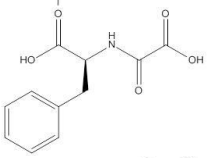
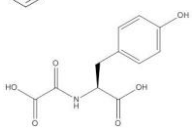
Table 9.10: Salt-bridge interactions within molecule A of ycfD$_{RM}$..... 343

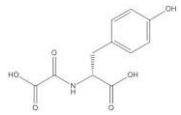
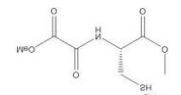
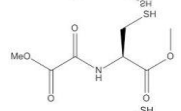
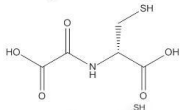
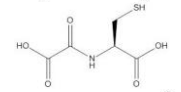
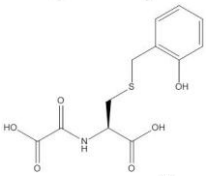
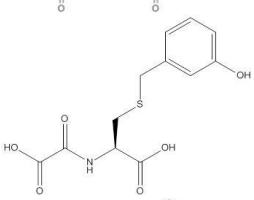
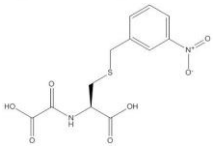
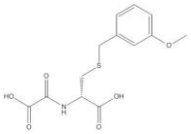
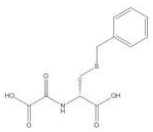
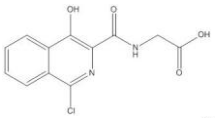
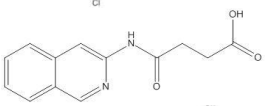
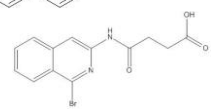
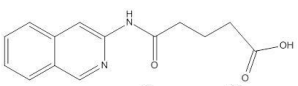
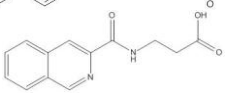
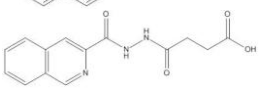
Table 9.1: Library of 2OG mimetics screened against $ycfD_{EC}$ and $ycfD_{RM}$.

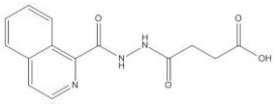
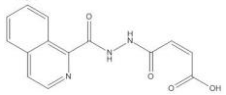
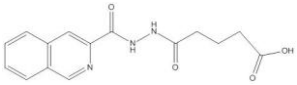
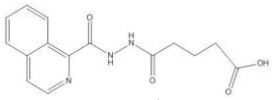
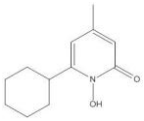
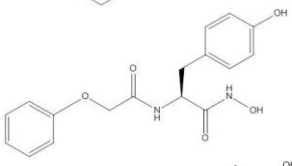
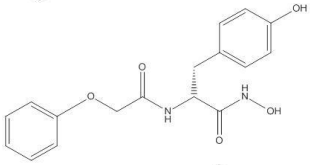
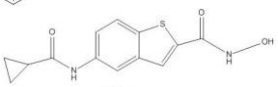
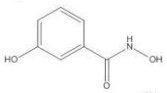
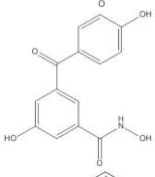
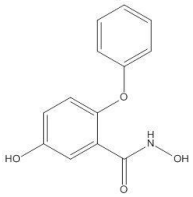
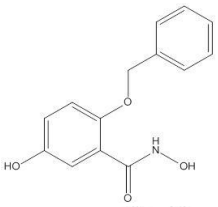
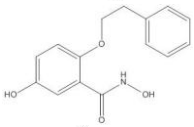
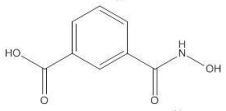
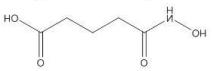
STRUCTURE	Name	Position
	EW-T4	2OG1-A2
	OC0353	2OG1-A3
	OC0358	2OG1-A4
	succinate	2OG1-A5
	fumarate	2OG1-A6
	oxaloacetate	2OG1-A7
	malate	2OG1-A8
	citrate	2OG1-A9
	isocitrate	2OG1-A10
	pyruvate	2OG1-A11
	D-2-hydroxyglutarate	2OG1-B3
	L-2-hydroxyglutarate	2OG1-B4
	EW-TNB	2OG1-B5
	aCG-180a	2OG1-B6
	FG2216	2OG1-B7

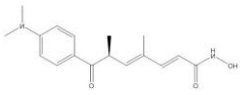
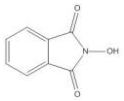
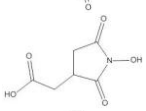
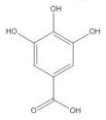
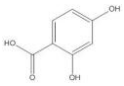
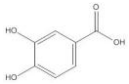
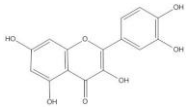
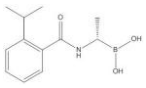
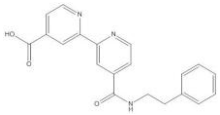
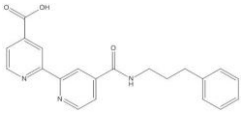
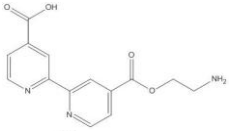
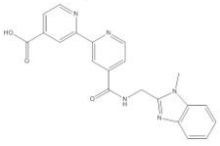
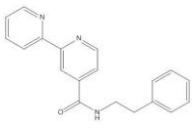
	BB-286	2OG1-B8
	JM104	2OG1-B9
	JM106	2OG1-B10
	JM103	2OG1-B11
	2,3-PDCA	2OG1-C1
	3,4-PDCA	2OG1-C2
	3,5-PDCA	2OG1-C3
	2,6-PDCA	2OG1-C4
	2,5-PDCA	2OG1-C5
	2,4-PDCA	2OG1-C6
	BMRL08012	2OG1-C7
	Prohexadione	2OG1-C8
	BB212	2OG1-C9
	Malathion	2OG1-C10

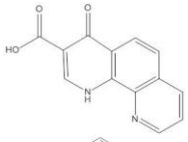
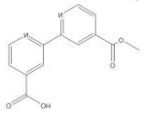
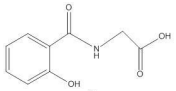
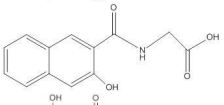
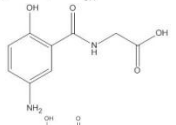
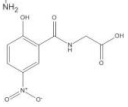
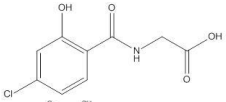
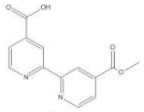
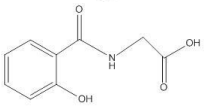
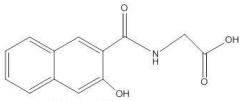
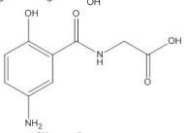
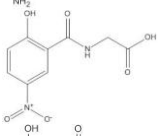
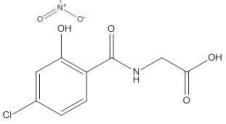
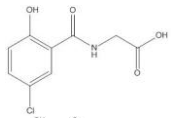
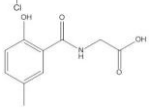
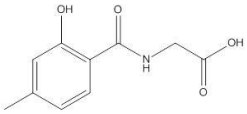
	Malaoxon	2OG1-C11
	Minoxidil	2OG1-D1
	IS70/NOAL	2OG1-D2
	IS71/NOAD	2OG1-D3
	NOG	2OG1-D5
	CO-A2-5a	2OG1-D6
	CO-A1-6d	2OG1-D7
	CO-A7-6h	2OG1-D8
	CO-B10-7a	2OG1-D10
	CO-C3-7f	2OG1-D11
	CO-C10-10a	2OG1-E1

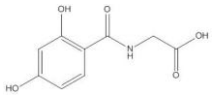
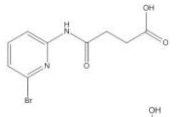
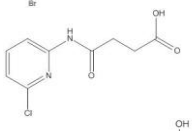
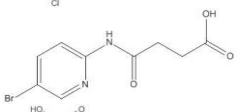
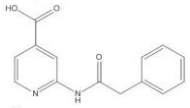
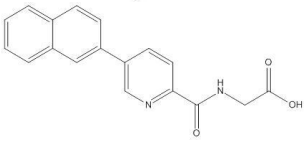
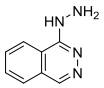
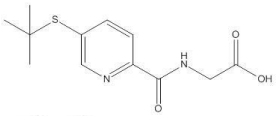
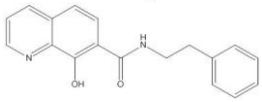
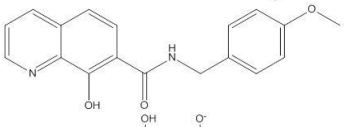
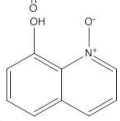
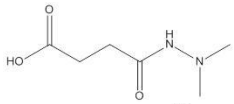
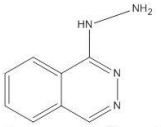
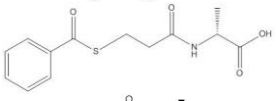
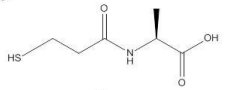
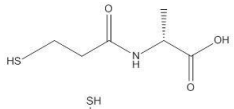
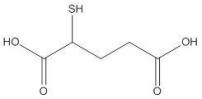
	CO-D8-10k	2OG1-E2
	CO-D9-12	2OG1-E3
	CO-C9-8d	2OG1-E5
	CO-D4-8m	2OG1-E6
	NOVD	2OG1-E7
	NOLD	2OG1-E8
	NOFD	2OG1-E10
	NOVL	2OG1-E11
	NOLL	2OG1-F1
	NOFL	2OG1-F2
	NOYL	2OG1-F3

	NOYD	2OG1-F4
	EW 29	2OG1-F5
	EW 30	2OG1-F6
	EW 31 / NOCD	2OG1-F7
	EW 32 / NOCL	2OG1-F8
	EW 33	2OG1-F9
	EW 34	2OG1-F10
	EW 35	2OG1-F11
	EW 36	2OG1-G1
	EW 37	2OG1-G2
	ACG-263	2OG1-G3
	BB-287	2OG1-G4
	BB-294	2OG1-G5
	BB-315	2OG1-G6
	BB-316	2OG1-G7
	JM87	2OG1-G8

	JM110	2OG1-G9
	JM124	2OG1-G10
	JM102	2OG1-G11
	JM109	2OG1-H1
	Ciclopirox Olamine	2OG1-H2
	IS44	2OG1-H3
	IS45	2OG1-H4
	ACG-114c	2OG1-H5
	LBE-332-6-3	2OG1-H6
	LBE-153	2OG1-H7
	LBE-314	2OG1-H8
	LBE-329	2OG1-H9
	LBE-526	2OG1-H10
	LBE096	2OG1-H11
	LBE060	2OG2-A2

	SAHA	2OG2-A3
	TSA	2OG2-A5
	N-hydroxyphthalimide	2OG2-A8
	IS52	2OG2-A9
	gallate	2OG2-A10
	2,4-DHB	2OG2-A11
	2,5-DHB	2OG2-B2
	3,4-DHB	2OG2-B3
	R2/quercetin	2OG2-B4
	EW-B17	2OG2-B6
	OC0248	2OG2-B7
	OC0249	2OG2-B8
	OC0206	2OG2-B9
	OC0336	2OG2-B10
	OC0340	2OG2-B11

	BB103	2OG2-C1
	BMRL08014	2OG2-C3
	ACG-194b	2OG2-C6
	ACG-181a	2OG2-C7
	ACG-255d	2OG2-C8
	ACG-253a	2OG2-C9
	ACG-227a	2OG2-C10
	ACG-223a	2OG2-C11
	ACG-226a	2OG2-D1
	ACG-228a	2OG2-D2
	ACG-237a	2OG2-D3
	BB-307	2OG2-D6
	BB-308	2OG2-D7
	BB-318	2OG2-D8
	BMRL08050	2OG2-D10
	MD228	2OG2-D11

	MD236	2OG2-E1
	MD104	2OG2-E2
	OC0344	2OG2-E3
	OC0345	2OG2-E5
	HQNO	2OG2-E7
	Daminozide	2OG2-E8
	Hydralazine	2OG2-E10
	IS21	2OG2-E11
	IS37	2OG2-F1
	IS38	2OG2-F2
	IS54	2OG2-F3
	IS84	2OG2-F4
	ACG-148a	2OG2-F5
	ACG-109a	2OG2-F6
	ACG-137a	2OG2-F7
	ACG-144a	2OG2-F8
	ACG-138a	2OG2-F9

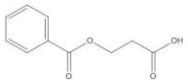
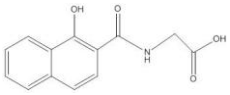
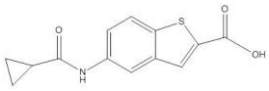
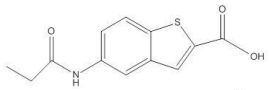
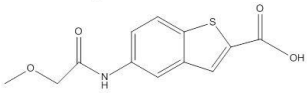
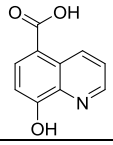
	BB110H	2OG2-F10
	BB169	2OG2-F11
	BB204	2OG2-G1
	IS3	2OG2-G4
	3-hydroxymandelate	2OG2-G5
	5-carboxy-8-hydroxyquinoline	2OG2-G7

Table 9.2: Compounds from the BB-series of compounds tested against ycfD$_{EC}$

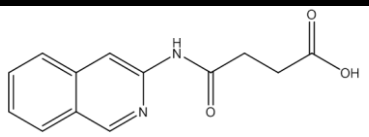
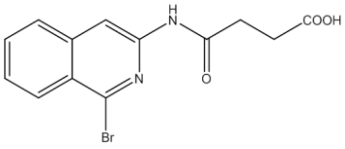
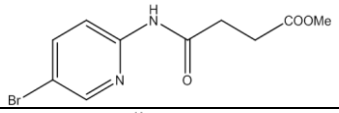
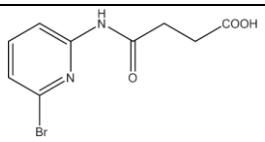
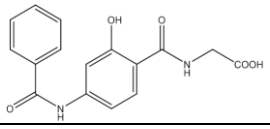
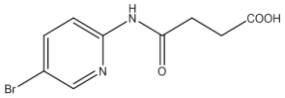
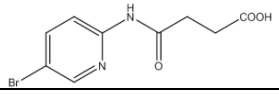
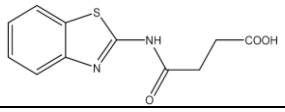
Name	Structure
BB287	
BB294	
BB301	
BB307	
BB309	
BB318	
BB323	
BB324	

Table 9.3: T_m shift analysis of $ycfD_{RM}$ stabilization by 2OG mimetics (plate 2OG1)

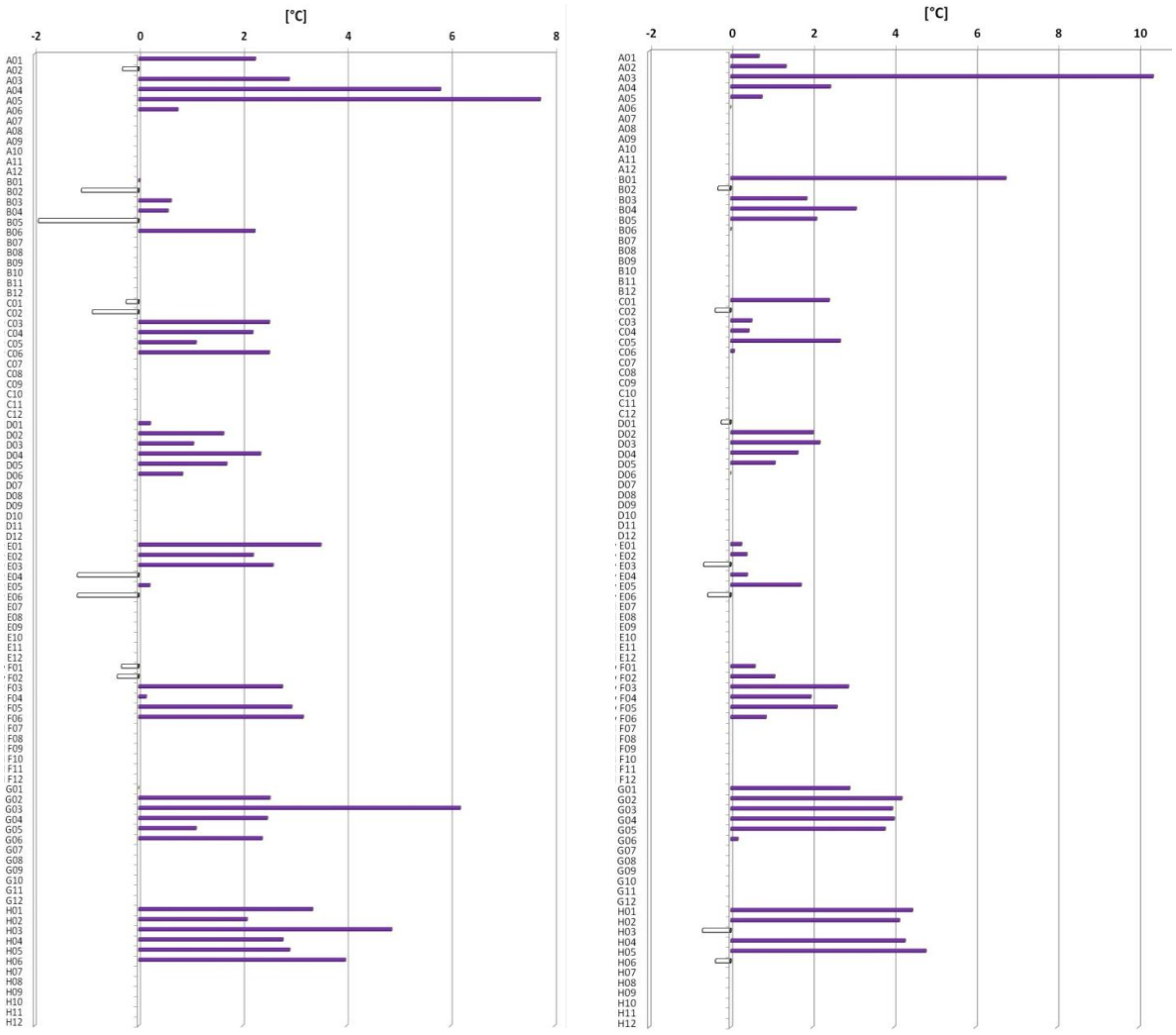
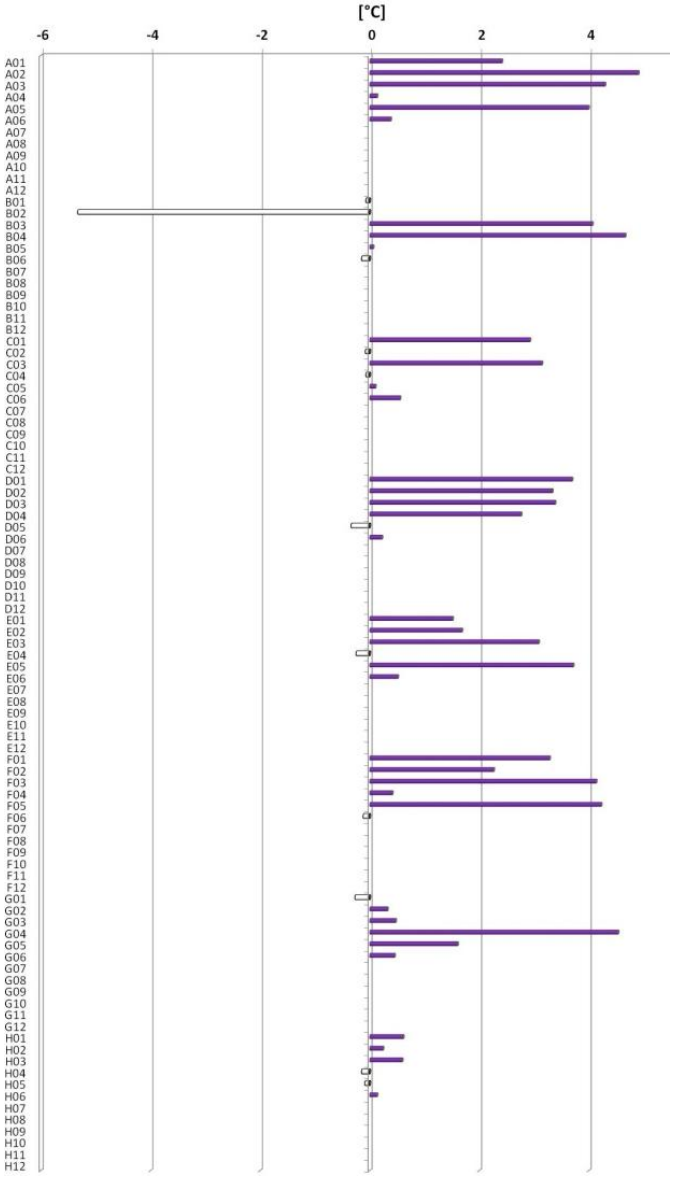


Table 9.4: T_m shift analysis of $ycfD_{RM}$ stabilization by 2OG mimetics (plate 2OG2)



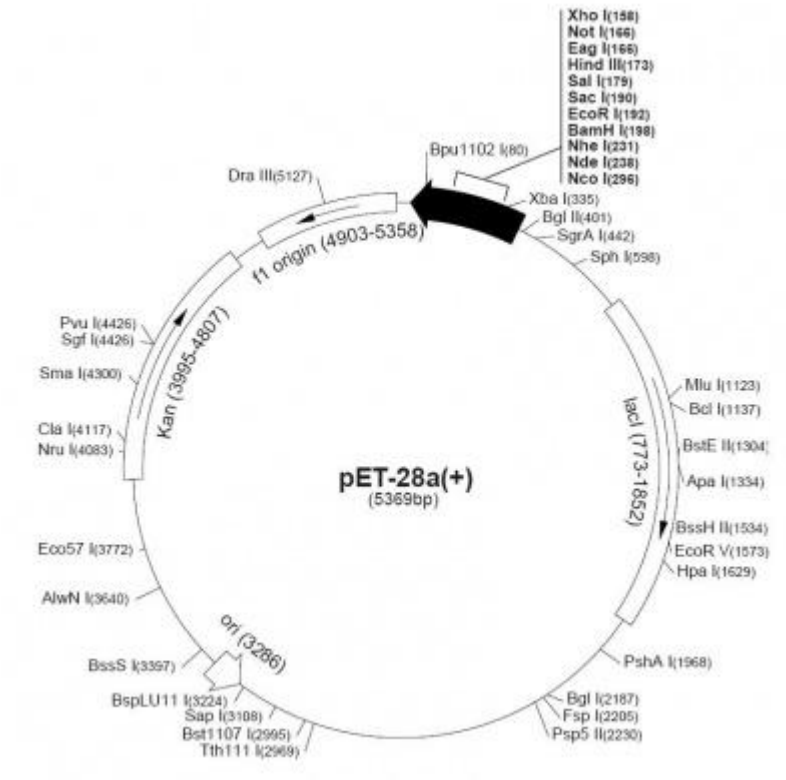


Figure 9.1: Vector map of pET28a vector used for cloning of $ycfD_{RM}$, $ycfD_{EC}$, $L16_{RM}$ and MINA53.

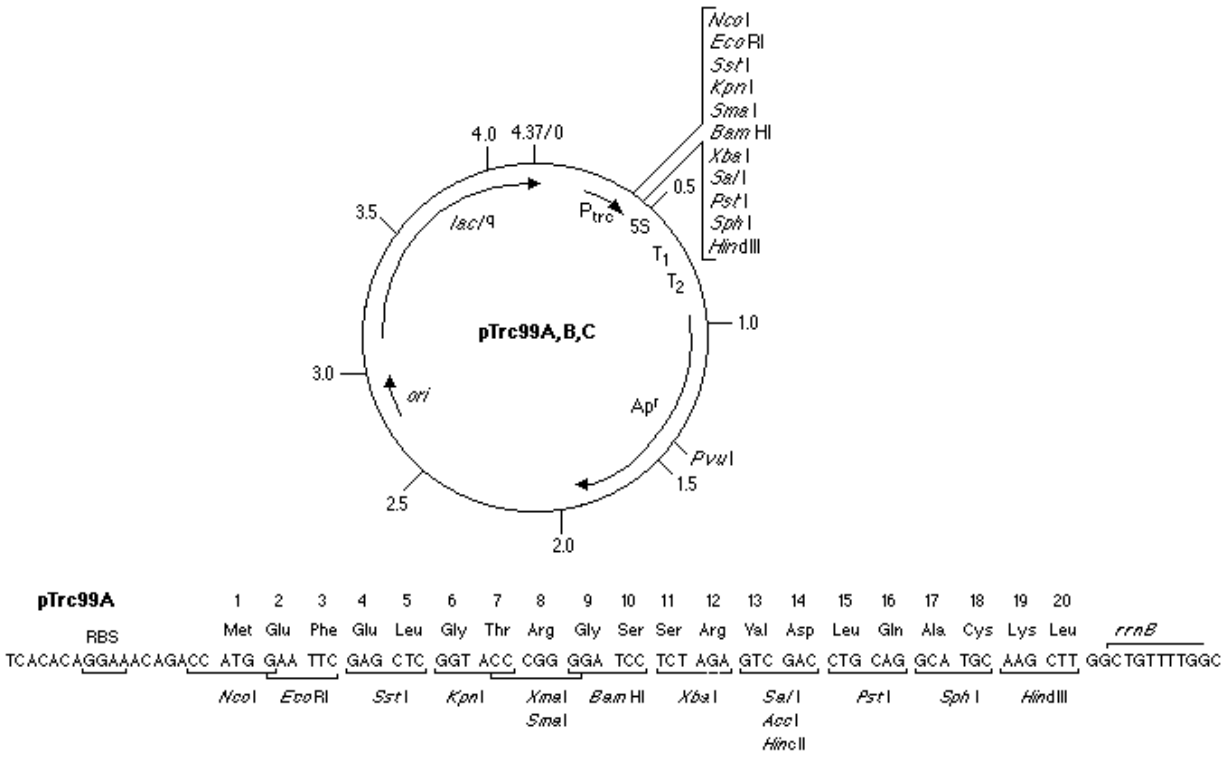


Figure 9.2: Vector map of pTRC99a vector used for over-expression of $ycfD_{RM}$ and $ycfD_{EC}$ in JW1114.

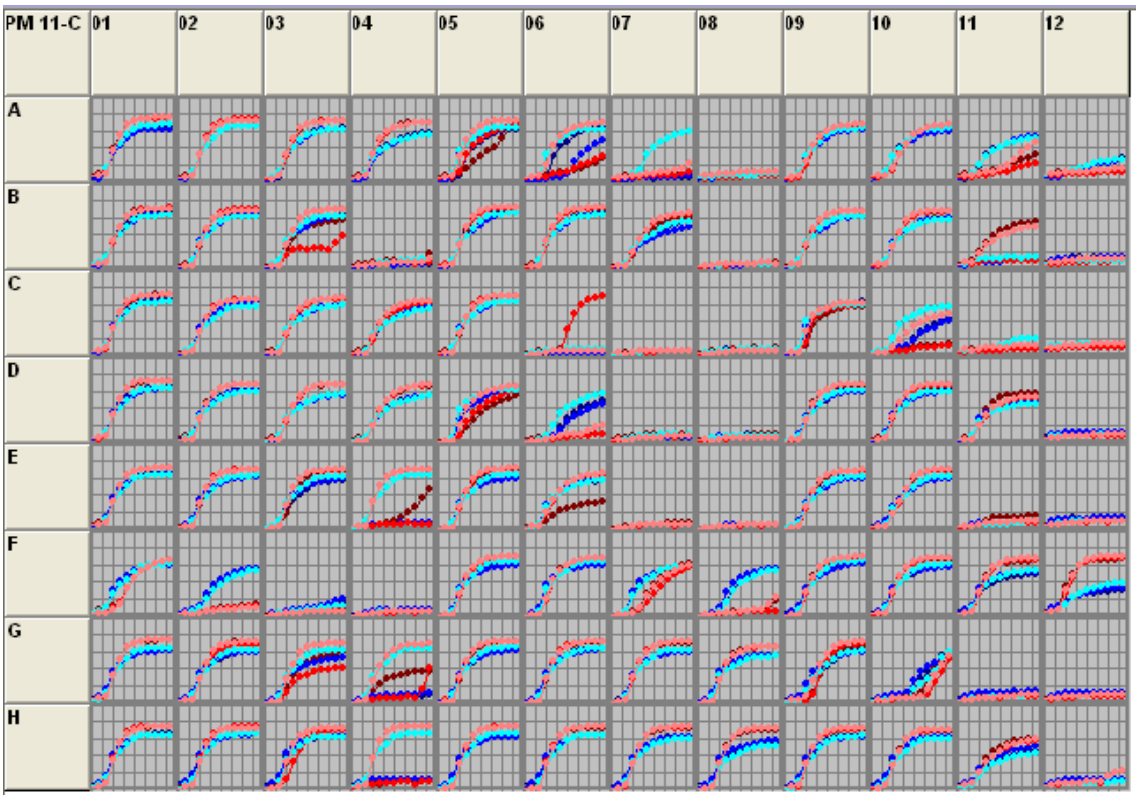


Figure 9.3: Phenotype MicroArray Single Panel Comparison Display showing Cellular Responses of BW25113 and JW1114 for 96 Kinetic Assays on Biolog PM11C MicroplateTM containing ribosomal antibiotics. Four dilutions of each antibiotic were tested against a triplicate of individual strain (Dark blue/cyan: BW25113, red/brown/pink: JW1114). A01-A04: Amikacin, A05-A08: Chlortetracycline, A09-A12: Lincomycin, B01-B04: Amoxicillin, B05-B08: Cloxacillin, B09-B12: Lomefloxacin, C01-C04: Bleomycin, C05-C08:Colistin, C09-C12: Minocycline, D01-D04: Capreomycin, D05-D08: Demeclocycline, D09-D12: Nafcillin, E01-E04: Cefazolin, E05-E08: Enoxacin, E09-E12: Nalidixic acid, F01-F04: Chloramphenicol, F05-F08: Erythromycin, F09-F12: Neomycin, G01-G04: Ceftriaxone, G05-G08: Gentamicin, G09-G12: Potassium tellurite, H01-H04: Cephalothin, H05-H08: Kanamycin, H09-H12: Ofloxacin.

Table 9.5: Refinement statistics of ycfD$_{RM}$ crystal structures

Datasets	ycfD$_{RM}$.Mn(II).FG221	ycfD$_{RM}$.Mn(II).NOG.L16
PDB	Not assigned	Not assigned
Data collection		
Beamline (Wavelength, Å)	Diamond I03 (0.9763)	Diamond I02 (1.0073)
Detector	Pilatus 6M-F	Pilatus 6M-F
Data processing	Xia2	Xia2
Space group	$C2_1$	$P2_12_12_1$
Cell dimensions		
***a, b, c* (Å)**	158.3, 87.69, 90.61	98.22, 117.47, 120.28
***α, β, γ* (°)**	90, 113.10, 90	90, 90, 90
No. of molecules/ ASU	2	2
No. reflections	55430	52236
Resolution (Å)	30.21-2.82 (2.89-2.82)	58.74-2.96 (3.04-2.96)
R_{sym} **or** R_{merge}**	0.122 (0.798)	0.051 (0.766)
$I/\sigma I$	7.6 (2.1)	19.4 (2.6)
Completeness (%)	99.6 (99.9)	99.9 (100)
Redundancy	4.9 (5.0)	6.4 (6.3)
Wilson *B* value (Å^2)	64.47	92.87
Refinement	PHENIX(1)	PHENIX(1)
R_{work}/ R_{free}‡	0.152/0.212	0.184/0.221
No. atomsΨ		
-Enzyme (A/B)	3316 (A),	3343 (A),
-Ligand		320
-Substrate	-	125 (E),
-Water	313	89
B-factorsΨ		
-Enzyme (A/B)	60.12 (A),	82.5 (A),
-Ligand	82.67	95.0
-Substrate	-	130.83 (E), 129.90 (F)
-Water	55.13 (S)	70.3 (S)
R.m.s deviations		
-Bond lengths (Å)	0.011	0.007
-Bond angles (°)	1.315	0.942

Table 9.6: Intra-chain hydrophobic contacts in ycfD$_{RM}$. Residues indicated in the left-hand column interact hydrophobically with residues indicated in the right-hand column.

#	Res.	#	Res.
2	LEU	102	ALA
2	LEU	103	LEU
2	LEU	3	PRO
2	LEU	31	LEU
2	LEU	34	PHE
3	PRO	31	LEU
3	PRO	32	PRO
6	ILE	107	VAL
6	ILE	109	PHE
6	ILE	139	PHE
6	ILE	141	VAL
6	ILE	186	VAL
6	ILE	208	ILE
6	ILE	27	ILE
6	ILE	30	ALA
6	ILE	31	LEU
6	ILE	7	LEU
7	LEU	10	LEU
7	LEU	109	PHE
7	LEU	12	PRO
7	LEU	139	PHE
7	LEU	15	PHE
7	LEU	27	ILE
10	LEU	15	PHE
10	LEU	19	TYR
10	LEU	25	LEU
11	ALA	12	PRO
12	PRO	109	PHE
15	PHE	109	PHE
15	PHE	139	PHE
15	PHE	16	LEU
15	PHE	188	TYR
15	PHE	19	TYR
15	PHE	20	TRP
15	PHE	25	LEU
16	LEU	109	PHE
16	LEU	110	VAL
16	LEU	20	TRP
19	TYR	188	TYR
19	TYR	25	LEU
20	TRP	110	VAL
20	TRP	139	PHE
20	TRP	188	TYR
20	TRP	210	PHE
24	PRO	151	ILE
24	PRO	189	LEU
24	PRO	190	PRO
24	PRO	193	ILE
25	LEU	188	TYR
25	LEU	27	ILE
26	LEU	179	TRP
26	LEU	181	LEU
26	LEU	187	LEU
27	ILE	139	PHE
27	ILE	186	VAL
27	ILE	30	ALA
30	ALA	186	VAL
30	ALA	31	LEU
31	LEU	103	LEU
31	LEU	141	VAL
31	LEU	186	VAL
31	LEU	32	PRO
31	LEU	34	PHE
34	PHE	103	LEU
34	PHE	141	VAL
34	PHE	143	ALA
34	PHE	206	PHE
34	PHE	37	PRO
37	PRO	100	VAL
37	PRO	206	PHE
37	PRO	38	ILE
38	ILE	100	VAL
38	ILE	120	VAL
38	ILE	204	MET
38	ILE	206	PHE
38	ILE	43	LEU
38	ILE	46	LEU
38	ILE	93	VAL
38	ILE	97	VAL
40	PRO	122	TYR
40	PRO	204	MET
43	LEU	120	VAL
43	LEU	122	TYR
43	LEU	46	LEU
43	LEU	88	LEU
43	LEU	90	VAL
43	LEU	93	VAL
44	ALA	78	PHE
44	ALA	88	LEU
46	LEU	90	VAL
46	LEU	93	VAL
46	LEU	96	LEU
46	LEU	97	VAL
47	ALA	54	ALA
47	ALA	73	PHE
47	ALA	78	PHE
47	ALA	88	LEU
47	ALA	90	VAL
52	VAL	286	LEU
52	VAL	54	ALA
52	VAL	72	PRO
52	VAL	90	VAL
52	VAL	96	LEU
54	ALA	73	PHE
54	ALA	88	LEU
54	ALA	90	VAL
56	LEU	58	LEU
56	LEU	73	PHE
56	LEU	78	PHE
56	LEU	81	LEU
56	LEU	86	TRP
56	LEU	88	LEU
57	ILE	129	VAL
57	ILE	167	VAL
57	ILE	66	TRP
57	ILE	68	VAL
57	ILE	89	LEU
58	LEU	81	LEU
58	LEU	82	PRO
58	LEU	86	TRP

#	Res.	#	Res.
63	ALA	64	TYR
64	TYR	65	PRO
65	PRO	66	TRP
66	TRP	165	LEU
66	TRP	167	VAL
66	TRP	169	LEU
68	VAL	167	VAL
68	VAL	70	TYR
73	PHE	78	PHE
73	PHE	88	LEU
78	PHE	81	LEU
78	PHE	88	LEU
79	VAL	80	ALA
81	LEU	82	PRO
81	LEU	86	TRP
82	PRO	83	PRO
82	PRO	86	TRP
86	TRP	122	TYR
86	TRP	124	PRO
86	TRP	88	LEU
88	LEU	122	TYR
88	LEU	90	VAL
89	LEU	119	MET
89	LEU	129	VAL
90	VAL	93	VAL
93	VAL	100	VAL
93	VAL	120	VAL
93	VAL	97	VAL
96	LEU	286	LEU
96	LEU	97	VAL
97	VAL	100	VAL
97	VAL	98	PRO
100	VAL	118	ILE
100	VAL	120	VAL
100	VAL	206	PHE
103	LEU	118	ILE
103	LEU	141	VAL
103	LEU	206	PHE
103	LEU	208	ILE
104	LEU	107	VAL
104	LEU	115	LEU
104	LEU	118	ILE
104	LEU	208	ILE
104	LEU	315	ILE
104	LEU	316	PRO

#	Res.	#	Res.
107	VAL	109	PHE
107	VAL	115	LEU
107	VAL	139	PHE
107	VAL	208	ILE
107	VAL	210	PHE
109	PHE	110	VAL
109	PHE	139	PHE
110	VAL	111	PRO
110	VAL	115	LEU
110	VAL	210	PHE
111	PRO	113	TRP
113	TRP	213	PRO
113	TRP	275	PHE
113	TRP	279	ILE
113	TRP	319	ALA
115	LEU	210	PHE
115	LEU	315	ILE
115	LEU	316	PRO
118	ILE	120	VAL
118	ILE	206	PHE
118	ILE	208	ILE
120	VAL	204	MET
120	VAL	206	PHE
122	TYR	124	PRO
122	TYR	204	MET
123	ALA	124	PRO
123	ALA	199	ALA
131	ALA	170	LEU
131	ALA	173	PHE
131	ALA	196	TYR
133	ILE	156	VAL
133	ILE	170	LEU
133	ILE	194	PRO
136	TYR	212	ALA
138	VAL	140	LEU
138	VAL	189	LEU
139	PHE	141	VAL
139	PHE	186	VAL
139	PHE	188	TYR
139	PHE	208	ILE
139	PHE	210	PHE
140	LEU	149	TRP
140	LEU	187	LEU
140	LEU	189	LEU
141	VAL	186	VAL

#	Res.	#	Res.
141	VAL	206	PHE
141	VAL	208	ILE
143	ALA	144	TRP
143	ALA	204	MET
143	ALA	206	PHE
144	TRP	204	MET
149	TRP	151	ILE
149	TRP	179	TRP
149	TRP	181	LEU
149	TRP	187	LEU
151	ILE	177	ALA
151	ILE	179	TRP
151	ILE	187	LEU
151	ILE	189	LEU
151	ILE	193	ILE
155	PRO	193	ILE
156	VAL	194	PRO
161	LEU	170	LEU
162	VAL	163	PRO
162	VAL	165	LEU
162	VAL	169	LEU
162	VAL	171	ALA
165	LEU	167	VAL
165	LEU	169	LEU
167	VAL	169	LEU
170	LEU	173	PHE
170	LEU	194	PRO
173	PHE	175	PRO
173	PHE	194	PRO
173	PHE	196	TYR
177	ALA	179	TRP
179	TRP	181	LEU
179	TRP	187	LEU
180	ILE	200	LEU
181	LEU	187	LEU
187	LEU	189	LEU
188	TYR	190	PRO
189	LEU	190	PRO
189	LEU	193	ILE
190	PRO	191	PRO
190	PRO	193	ILE
193	ILE	194	PRO
196	TYR	198	VAL
204	MET	206	PHE
208	ILE	210	PHE

#	Res.	#	Res.
212	ALA	213	PRO
213	PRO	218	LEU
213	PRO	275	PHE
213	PRO	279	ILE
218	LEU	274	TRP
218	LEU	275	PHE
218	LEU	278	ILE
221	ALA	274	TRP
221	ALA	278	ILE
222	MET	223	PRO
222	MET	225	MET
222	MET	274	TRP
225	MET	229	LEU
225	MET	264	LEU
225	MET	271	LEU
225	MET	274	TRP
226	ALA	229	LEU
227	ALA	228	TRP
228	TRP	229	LEU
228	TRP	274	TRP
228	TRP	336	ALA
229	LEU	264	LEU
235	TYR	253	ALA
235	TYR	257	ILE
235	TYR	260	LEU
236	ALA	238	PRO
243	ALA	246	PRO
249	ILE	253	ALA
249	ILE	254	LEU
254	LEU	257	ILE
257	ILE	260	LEU
257	ILE	261	LEU
261	LEU	264	LEU
261	LEU	265	ILE
264	LEU	265	ILE
265	ILE	271	LEU
271	LEU	274	TRP
272	ALA	321	VAL
274	TRP	275	PHE
274	TRP	278	ILE
275	PHE	278	ILE
275	PHE	279	ILE
278	ILE	279	ILE
282	PRO	288	PRO
286	LEU	287	PRO

#	Res.	#	Res.
287	PRO	288	PRO
290	PRO	291	PRO
291	PRO	378	ILE
295	LEU	300	LEU
295	LEU	378	ILE
295	LEU	385	LEU
297	ALA	371	LEU
300	LEU	304	LEU
300	LEU	371	LEU
300	LEU	374	LEU
300	LEU	378	ILE
300	LEU	385	LEU
304	LEU	308	ALA
304	LEU	310	LEU
304	LEU	358	ALA
304	LEU	361	LEU
304	LEU	365	LEU
304	LEU	374	LEU
304	LEU	385	LEU
308	ALA	385	LEU
308	ALA	387	PRO
310	LEU	356	LEU
310	LEU	361	LEU
310	LEU	374	LEU
310	LEU	378	ILE
310	LEU	383	LEU
310	LEU	385	LEU
314	ALA	316	PRO
315	ILE	316	PRO
315	ILE	318	LEU
318	LEU	332	ALA
318	LEU	382	ILE
318	LEU	383	LEU
319	ALA	331	PHE
320	TYR	330	LEU
320	TYR	347	ALA
320	TYR	348	PRO
321	VAL	331	PHE
328	ALA	339	LEU
328	ALA	344	ALA
328	ALA	347	ALA
330	LEU	337	TYR
330	LEU	339	LEU
330	LEU	347	ALA
330	LEU	350	LEU

#	Res.	#	Res.
330	LEU	373	LEU
330	LEU	377	LEU
331	PHE	336	ALA
332	ALA	337	TYR
332	ALA	377	LEU
332	ALA	382	ILE
337	TYR	339	LEU
337	TYR	373	LEU
337	TYR	377	LEU
339	LEU	343	LEU
339	LEU	344	ALA
339	LEU	346	VAL
339	LEU	347	ALA
339	LEU	373	LEU
343	LEU	346	VAL
343	LEU	364	TRP
343	LEU	370	PHE
346	VAL	349	LEU
346	VAL	350	LEU
346	VAL	364	TRP
346	VAL	370	PHE
346	VAL	373	LEU
347	ALA	348	PRO
349	LEU	350	LEU
349	LEU	356	LEU
349	LEU	364	TRP
350	LEU	356	LEU
350	LEU	370	PHE
350	LEU	373	LEU
350	LEU	374	LEU
350	LEU	377	LEU
350	LEU	383	LEU
356	LEU	361	LEU
356	LEU	383	LEU
361	LEU	364	TRP
361	LEU	365	LEU
361	LEU	370	PHE
361	LEU	374	LEU
363	PRO	364	TRP
364	TRP	370	PHE
365	LEU	370	PHE
365	LEU	371	LEU
370	PHE	373	LEU
370	PHE	374	LEU
371	LEU	374	LEU

#	Res.	#	Res.
373	LEU	377	LEU
374	LEU	377	LEU
374	LEU	378	ILE
374	LEU	383	LEU

#	Res.	#	Res.
377	LEU	382	ILE
377	LEU	383	LEU
378	ILE	383	LEU
378	ILE	385	LEU

#	Res.	#	Res.
382	ILE	383	LEU
386	ILE	387	PRO
387	PRO	388	ALA

Table 9.7: Intra-chain hydrophobic interactions in ycfD$_{EC}$. Residues indicated in the left-hand column interact hydrophobically with residues indicated in the right-hand column.

#	Res.	#	Res.
3	TYR	28	PHE
3	TYR	31	PHE
3	TYR	34	PRO
3	TYR	95	ALA
3	TYR	96	LEU
3	TYR	99	PRO
5	LEU	100	PHE
5	LEU	178	ILE
5	LEU	28	PHE
5	LEU	7	LEU
5	LEU	9	TRP
5	LEU	99	PRO
7	LEU	12	PHE
7	LEU	132	PHE
7	LEU	24	LEU
7	LEU	9	TRP
9	TRP	10	PRO
9	TRP	100	PHE
9	TRP	103	LEU
9	TRP	12	PHE
9	TRP	13	LEU
9	TRP	132	PHE
12	PHE	103	LEU
12	PHE	13	LEU
12	PHE	132	PHE
12	PHE	17	TRP
12	PHE	180	TYR
12	PHE	22	VAL
12	PHE	24	LEU
13	LEU	103	LEU
13	LEU	17	TRP
17	TRP	103	LEU
17	TRP	132	PHE
17	TRP	180	TYR
17	TRP	202	PHE
21	PRO	144	VAL
21	PRO	169	ILE
21	PRO	179	LEU
21	PRO	181	ILE
21	PRO	182	PRO
21	PRO	185	PHE
22	VAL	180	TYR

#	Res.	#	Res.
22	VAL	24	LEU
23	VAL	169	ILE
23	VAL	173	LEU
23	VAL	179	LEU
24	LEU	132	PHE
24	LEU	178	ILE
24	LEU	180	TYR
28	PHE	100	PHE
28	PHE	134	ILE
28	PHE	178	ILE
28	PHE	31	PHE
28	PHE	96	LEU
28	PHE	99	PRO
31	PHE	134	ILE
31	PHE	175	PRO
31	PHE	198	TYR
31	PHE	34	PRO
31	PHE	96	LEU
34	PRO	198	TYR
34	PRO	35	ILE
34	PRO	92	PRO
34	PRO	96	LEU
35	ILE	113	ILE
35	ILE	198	TYR
35	ILE	40	LEU
35	ILE	86	VAL
37	PRO	115	PHE
40	LEU	113	ILE
40	LEU	115	PHE
40	LEU	81	LEU
40	LEU	83	VAL
40	LEU	86	VAL
41	ALA	44	ALA
41	ALA	71	TYR
41	ALA	81	LEU
43	LEU	49	VAL
43	LEU	86	VAL
43	LEU	89	TRP
44	ALA	68	PHE
44	ALA	71	TYR
44	ALA	81	LEU
44	ALA	83	VAL

#	Res.	#	Res.
49	VAL	83	VAL
49	VAL	89	TRP
53	LEU	68	PHE
53	LEU	71	TYR
53	LEU	79	TRP
53	LEU	81	LEU
54	VAL	122	VAL
54	VAL	61	TRP
54	VAL	63	VAL
54	VAL	82	LEU
68	PHE	71	TYR
68	PHE	81	LEU
71	TYR	81	LEU
79	TRP	115	PHE
79	TRP	117	VAL
79	TRP	81	LEU
81	LEU	115	PHE
81	LEU	83	VAL
82	LEU	122	VAL
83	VAL	86	VAL
86	VAL	111	LEU
86	VAL	113	ILE
96	LEU	111	LEU
96	LEU	134	ILE
96	LEU	198	TYR
99	PRO	100	PHE
100	PHE	108	ILE
100	PHE	132	PHE
100	PHE	134	ILE
100	PHE	178	ILE
100	PHE	200	VAL
100	PHE	202	PHE
103	LEU	104	PRO
103	LEU	132	PHE
103	LEU	202	PHE
104	PRO	106	TRP
106	TRP	205	PRO
106	TRP	268	PHE
106	TRP	272	ILE
106	TRP	312	LEU
108	ILE	200	VAL
108	ILE	202	PHE

#	Res.	#	Res.
111	LEU	113	ILE
111	LEU	198	TYR
111	LEU	200	VAL
113	ILE	198	TYR
115	PHE	117	VAL
117	VAL	118	PRO
124	PRO	165	PHE
126	LEU	162	VAL
126	LEU	186	PRO
129	TYR	204	ALA
131	VAL	133	ILE
131	VAL	181	ILE
132	PHE	178	ILE
132	PHE	180	TYR
132	PHE	200	VAL
132	PHE	202	PHE
133	ILE	142	TRP
133	ILE	179	LEU
133	ILE	181	ILE
134	ILE	178	ILE
134	ILE	198	TYR
134	ILE	200	VAL
142	TRP	173	LEU
142	TRP	179	LEU
144	VAL	168	ILE
144	VAL	169	ILE
144	VAL	179	LEU
144	VAL	181	ILE
144	VAL	185	PHE
157	PRO	160	LEU
165	PHE	167	ALA
165	PHE	186	PRO
168	ILE	169	ILE
169	ILE	179	LEU
173	LEU	179	LEU
179	LEU	181	ILE
180	TYR	182	PRO
181	ILE	182	PRO
181	ILE	185	PHE
182	PRO	183	PRO
182	PRO	185	PHE
185	PHE	186	PRO
190	TYR	192	LEU
191	ALA	195	ALA
200	VAL	202	PHE

#	Res.	#	Res.
204	ALA	205	PRO
205	PRO	210	LEU
205	PRO	268	PHE
205	PRO	271	PHE
205	PRO	272	ILE
210	LEU	267	TRP
210	LEU	268	PHE
210	LEU	271	PHE
211	ILE	258	ILE
214	PHE	217	TYR
214	PHE	218	VAL
214	PHE	257	LEU
214	PHE	258	ILE
214	PHE	264	PHE
214	PHE	267	TRP
217	TYR	218	VAL
217	TYR	223	LEU
217	TYR	267	TRP
218	VAL	219	LEU
218	VAL	223	LEU
218	VAL	257	LEU
223	LEU	257	LEU
228	TYR	250	LEU
233	VAL	234	PRO
234	PRO	235	PRO
234	PRO	243	LEU
243	LEU	244	PRO
255	LEU	258	ILE
257	LEU	258	ILE
264	PHE	267	TRP
264	PHE	268	PHE
267	TRP	268	PHE
267	TRP	271	PHE
268	PHE	271	PHE
268	PHE	272	ILE
268	PHE	314	ILE
271	PHE	272	ILE
272	ILE	312	LEU
272	ILE	314	ILE
279	LEU	281	ILE
279	LEU	306	LEU
279	LEU	309	LEU
281	ILE	283	PRO
281	ILE	306	LEU
281	ILE	368	TYR

#	Res.	#	Res.
282	ALA	283	PRO
283	PRO	284	PRO
284	PRO	287	PRO
284	PRO	288	TYR
284	PRO	364	VAL
286	PRO	287	PRO
288	TYR	293	ILE
288	TYR	364	VAL
288	TYR	371	PHE
290	PRO	357	LEU
290	PRO	358	ALA
290	PRO	361	ALA
293	ILE	297	LEU
293	ILE	360	LEU
293	ILE	361	ALA
293	ILE	364	VAL
293	ILE	371	PHE
294	TYR	351	LEU
294	TYR	357	LEU
296	ALA	371	PHE
297	LEU	303	LEU
297	LEU	344	ALA
297	LEU	347	PHE
297	LEU	351	LEU
297	LEU	360	LEU
297	LEU	371	PHE
302	VAL	304	VAL
302	VAL	341	ALA
303	LEU	342	LEU
303	LEU	347	PHE
303	LEU	360	LEU
303	LEU	364	VAL
303	LEU	369	TRP
303	LEU	371	PHE
304	VAL	340	ILE
304	VAL	341	ALA
304	VAL	370	PHE
306	LEU	309	LEU
306	LEU	370	PHE
309	LEU	311	VAL
309	LEU	368	TYR
309	LEU	369	TRP
311	VAL	318	VAL
311	VAL	320	ALA
311	VAL	325	ILE

#	Res.	#	Res.
311	VAL	337	ALA
311	VAL	369	TRP
314	ILE	319	TYR
318	VAL	325	ILE
318	VAL	333	LEU
318	VAL	337	ALA
320	ALA	325	ILE
320	ALA	368	TYR
320	ALA	369	TRP
325	ILE	333	LEU
325	ILE	336	LEU
325	ILE	363	LEU
325	ILE	369	TRP
332	ALA	347	PHE
332	ALA	350	ALA

#	Res.	#	Res.
332	ALA	356	PHE
333	LEU	336	LEU
333	LEU	356	PHE
335	ALA	342	LEU
336	LEU	342	LEU
336	LEU	347	PHE
336	LEU	356	PHE
336	LEU	360	LEU
336	LEU	369	TRP
340	ILE	341	ALA
342	LEU	347	PHE
342	LEU	369	TRP
344	ALA	351	LEU
347	PHE	350	ALA
347	PHE	351	LEU

#	Res.	#	Res.
347	PHE	356	PHE
347	PHE	360	LEU
350	ALA	356	PHE
351	LEU	357	LEU
354	PRO	357	LEU
356	PHE	360	LEU
357	LEU	358	ALA
360	LEU	363	LEU
360	LEU	364	VAL
363	LEU	364	VAL
363	LEU	368	TYR
363	LEU	369	TRP
364	VAL	369	TRP
368	TYR	369	TRP

Table 9.8: List of salt-bridge interactions between protomers A and B in $ycfD_{RM}$ and $ycfD_{EC}$. Software PISA (available on http://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver) was used for structural analysis.

	$ycfD_{RM}$			$ycfD_{EC}$		
	Molecule A	Molecule B	Dist. [Å]	Molecule A	Molecule B	Dist. [Å]
1	ASP 214[OD1]	ARG 234[NH1]	3.24	ARG 107[NE]	ASP 230[OD1]	3.33
2	ASP 214[OD1]	ARG 234[NH2]	3.08	ARG 107[NH2]	ASP 230[OD1]	2.92
3	GLU 217[OE2]	ARG 234[NH1]	3.59	ARG 107[NE]	ASP 230[OD2]	2.83
4	ASP 237[OD1]	ARG 114[NE]	3.77	ARG 107[NH2]	ASP 230[OD2]	3.85
5	ASP 237[OD1]	ARG 114[NH2]	3.53	ARG 313[NH1]	ASP 241[OD1]	2.93
6	ASP 237[OD2]	ARG 114[NE]	3.15	ARG 313[NH2]	ASP 241[OD2]	3.76
7	ASP 255[OD1]	ARG 268[NH2]	2.49	ARG 313[NH1]	ASP 241[OD2]	3.91
8	ASP 266[OD1]	ARG 262[NH2]	3.11	HIS 238[NE2]	ASP 334[OD1]	3.79
9	ASP 266[OD1]	ARG 262[NE]	3.42	ASP 230[OD1]	ARG 107[NE]	3.49
10	ASP 266[OD2]	ARG 262[NH2]	3.64	ASP 230[OD1]	ARG 107[NH2]	3.23
11	ASP 266[OD2]	ARG 262[NE]	3.07	ASP 230[OD2]	ARG 107[NE]	2.78
12	ARG 234[NH2]	ASP 214[OD1]	3.19	ASP 230[OD2]	ARG 107[NH2]	3.98
13	ARG 234[NH1]	GLU 217[OE2]	3.44	ASP 334[OD1]	HIS 238[NE2]	3.61
14	ARG 234[NH2]	GLU 217[OE2]	3.54			
15	ARG 114[NE]	ASP 237[OD1]	3.71			
16	ARG 114[NH1]	ASP 237[OD1]	2.94			
17	ARG 114[NE]	ASP 237[OD2]	2.89			
18	ARG 114[NH1]	ASP 237[OD2]	3.61			
19	ARG 322[NH2]	GLU 248[OE1]	3.62			
20	ARG 322[NE]	GLU 248[OE2]	3.83			
21	ARG 322[NH2]	GLU 248[OE2]	3.57			
22	ARG 268[NH1]	ASP 255[OD1]	2.24			
23	ARG 262[NH2]	ASP 266[OD1]	3.84			
24	ARG 262[NE]	ASP 266[OD2]	2.98			
25	ARG 262[NH2]	ASP 266[OD2]	3.06			

Table 9.9: List of hydrogen-bonding interactions identified between protomers A and B in ycfD$_{RM}$ and ycfD. PISA software was used for the structural analysis.

	ycfD$_{RM}$			ycfD$_{EC}$		
	Molecule A	Molecule B	Dist. [Å]	Molecule A	Molecule B	Dist. [Å]
1	GLN 21[OE1]	ASP237[H]	2.43	ASP 230[N]	GLN 18[OE1]	3.25
2	ALA 212[O]	ARG234[HH11]	1.92	TYR 228[N]	PRO 205[O]	2.94
3	PRO 213[O]	TYR 235[H]	2.10	ARG 208[NH2]	LEU 223[O]	3.14
4	GLN 215[OE1]	ARG 233[H]	2.36	ARG 208[NE]	LEU 223[O]	3.03
5	ARG 233[O]	GLN 215[H]	2.20	THR 207[N]	GLY 224[O]	3.21
6	ASP 237[OD2]	ARG 114[HE]	2.30	ARG 208[N]	GLY 224[O]	2.84
7	ASP 244[O]	ARG 353[HH21]	1.84	THR 207[N]	ASN 226[O]	3.25
8	PRO 246[O]	ASN 112[H]	1.92	THR 207[OG1]	ASN 226[O]	3.78
9	GLY 247[O]	TYR 320[H]	2.35	THR 207[OG1]	ASN 226[OD1]	2.81
10	ASP 255[OD1]	ARG 268[HH22]	1.87	TRP 17[NE1]	ASP 230[OD1]	3.37
11	GLN 258[OE1]	ARG 268[HH12]	1.94	ARG 107[NH2]	ASP 230[OD1]	2.92
12	LEU 260[O]	GLN 215[HE21]	2.13	ARG 107[NE]	ASP 230[OD2]	2.83
13	ASP 266[OD1]	ARG 262[HH21]	2.29	LYS 19[NZ]	PRO 231[O]	2.98
14	ASP 266[OD2]	ARG 262[HE]	2.34	ASP 105[N]	PRO 239[O]	2.89
15	ASP 266[OD2]	GLN 258[HE22]	2.37	ARG 313[N]	ALA 240[O]	2.98
16	TYR 320[O]	ILE 249[H]	2.15	ARG 313[NH1]	ASP 241[OD1]	2.93
17	ASP 237[H]	GLN 21[OE1]	2.27	TRP 106[NE1]	GLU 246[OE2]	3.11
18	ARG 234[HH11]	ALA 212[O]	1.93	VAL 242[N]	ARG 313[O]	2.93
19	TYR 235[H]	PRO 213[O]	2.20	GLN 18[OE1]	ASP 230[N]	3.09
20	ARG 233[H]	GLN 215[OE1]	2.37	PRO 205[O]	TYR 228[N]	2.98
21	ALA 216[H]	GLY 231[O]	2.33	LEU 223[O]	ARG 208[NH1]	2.79
22	GLN 215[H]	ARG 233[O]	2.37	GLY 224[O]	THR 207[N]	3.20
23	ARG 114[HH12]	ASP 237[OD1]	2.11	GLY 224[O]	ARG 208[N]	2.98
24	ARG 114[HE]	ASP 237[OD2]	2.03	GLY 225[O]	ASN 206[ND2]	3.86
25	LYS 22[HZ2]	PRO 238[O]	2.22	ASN 226[O]	THR 207[N]	3.37
26	ARG 353[HH12]	ALA 243[O]	1.67	ASN 226[OD1]	THR 207[OG1]	2.78
27	ASN 112[H]	PRO 246[O]	1.87	ASP 230[OD1]	TRP 17[NE1]	3.21
28	TYR 320[H]	GLY 247[O]	2.19	ASP 230[OD1]	ARG 107[NH2]	3.23
29	ARG 268[HH11]	ASP 255[OD1]	1.97	ASP 230[OD2]	ARG 107[NE]	2.78
30	ARG 268[HH12]	ASP 255[OD1]	1.93	PRO 231[O]	LYS 19[NZ]	2.83
31	ARG 268[HH22]	GLN 258[OE1]	1.94	PRO 239[O]	ASP 105[N]	2.85
32	GLN 215[HE21]	LEU 260[O]	2.04	ALA 240[O]	ARG 313[N]	2.99
33	ARG 262[HE]	ASP 266[OD2]	2.20	GLU 246[OE2]	TRP 106[NE1]	3.00
34	ARG 262[HH21]	ASP 266[OD2]	2.31	ARG 313[O]	VAL 242[N]	2.93
35	GLN 258[HE21]	ASP 266[OD2]	2.32			

Table 9.10: Salt-bridge interactions within molecule A of ycfD$_{RM}$.

Molecule A	Molecule B	Dist. [Å]
HIS85[ND1]	GLU59[OE1]	3.03
ARG95[NH1]	GLU317[OE1]	2.61
ARG95[NH2]	ASP94[OD1]	3.94
ARG95[NH2]	ASP94[OD2]	3.44
ARG108[NH2]	GLU105[OE1]	3.66
ARG114[NH1]	ASP137[OD1]	3.51
HIS132[NE2]	ASP134[OD1]	3.88
HIS132[NE2]	ASP134[OD2]	3.25
ARG146[NH2]	GLU201[OE1]	2.74
HIS153[ND1]	ASP176[OD1]	3.12
ARG168[NH1]	GLU159[OE1]	3.19
HIS172[NE2]	GLU160[OE1]	3.5
HIS195[NE2]	ASP134[OD2]	3.11
ARG211[NH1]	GLU217[OE1]	3.03
ARG273[NH2]	GLU338[OE1]	3.18
ARG284[NH1]	ASP116[OD1]	3.23
ARG284[NH1]	ASP117[OD2]	3.82
ARG284[NH2]	ASP116[OD1]	3.71
ARG284[NH2]	ASP116[OD2]	3.7
ARG284[NH2]	ASP117[OD2]	3.87
ARG362[NH2]	GLU359[OE1]	3.15
ARG367[NH2]	ASP369[OD1]	3.87
ARG367[NH2]	ASP369[OD2]	2.9

List of Publications

Publications resulting from DPhil project

1) Jayatunga MK, Thompson S, McKee TC, Chan MC, Reece KM, Hardy AP, Sekirnik R, Seden PT, Cook KM, McMahon JB, Figg WD, Schofield CJ, Hamilton AD (2014) Inhibition of the HIF1α-p300 interaction by quinone- and indandione-mediated ejection of structural Zn(II) *Eur J Med* [Epub ahead of print] doi: 10.1016/j.ejmech.2014.06.006.

2) Chowdhury R, Sekirnik R, Brissett NC, Krojer T, Ho CH, Ng SS, Clifton IJ, Ge W, Kershaw NJ, Fox GC, Muniz JR, Vollmar M, Phillips C, Pilka ES, Kavanagh KL, von Delft F, Oppermann U, McDonough MA, Doherty AJ, Schofield CJ (2014) Ribosomal oxygenases are structurally conserved from prokaryotes to humans *Nature* **510,** 422-6. doi: 10.1038/nature13263.

3) Loenarz C, Sekirnik R, Thalhammer A, Ge W, Spivakovsky E, Mackeen MM, McDonough MA, Cockman ME, Kessler BM, Ratcliffe PJ, Wolf A, Schofield CJ (2014) Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy *Proc Natl Acad Sci U S A* **111,** 4019-24. doi: 10.1073/pnas.1311750111.

4) Katz MJ, Acevedo JM, Loenarz C, Galagovsky D, Liu-Yi P, Pérez-Pepe M, Thalhammer A, Sekirnik R, Ge W, Melani M, Thomas MG, Simonetta S, Boccaccio GL, Schofield CJ, Cockman ME, Ratcliffe PJ, Wappner P (2014) Sudestada1, a Drosophila ribosomal prolyl-hydroxylase required for mRNA translation, cell homeostasis, and organ growth *Proc Natl Acad Sci U S A* **111,** 4025-30. doi: 10.1073/pnas.1314485111.

5) Singleton RS, Liu-Yi P, Formenti F, Ge W, Sekirnik R, Fischer R, Adam J, Pollard PJ, Wolf A, Thalhammer A, Loenarz C, Flashman E, Yamamoto A, Coleman ML, Kessler BM, Wappner P, Schofield CJ, Ratcliffe PJ, Cockman ME (2014) OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation *Proc Natl Acad Sci U S A,* **111,** 4031-6. doi: 10.1073/pnas.1314482111.

6) Ge W, Wolf A, Feng T, Ho CH, Sekirnik R, Zayer A, Granatino N, Cockman ME, Loenarz C, Loik ND, Hardy AP, Claridge TD, Hamed RB, Chowdhury R, Gong L, Robinson CV, Trudgian DC, Jiang M, Mackeen MM, McCullagh JS, Gordiyenko Y, Thalhammer A, Yamamoto A, Yang M, Liu-Yi P, Zhang Z, Schmidt-Zachmann M, Kessler BM, Ratcliffe PJ, Preston GM, Coleman ML, Schofield CJ (2012) Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans *Nat Chem Biol* **8,** 960-2. doi: 10.1038/nchembio.1093.

Other publications

7) Masson N, Singleton RS, Sekirnik R, Trudgian DC, Ambrose LJ, Miranda MX, Tian YM, Kessler BM, Schofield CJ, Ratcliffe PJ (2012) The FIH hydroxylase is a cellular peroxide sensor that modulates HIF transcriptional activity *EMBO Rep* **13,** 251-7. doi: 10.1038/embor.2012.9.

8) Sekirnik R, Rose NR, Mecinović J, Schofield CJ (2010) 2-Oxoglutarate oxygenases are inhibited by a range of transition metals *Metallomics* **2,** 397-9. doi: 10.1039/c004952b.

9) Sekirnik R, Rose NR, Thalhammer A, Seden PT, Mecinović J, Schofield CJ (2009) Inhibition of the histone lysine demethylase JMJD2A by ejection of structural Zn(II) *Chem Commun* **14**, 6376-8. doi: 10.1039/b916357c.