Mechanistic and structural study of human aspartate/asparagine-β-hydroxylase



Amelia Brasnett

Jesus College

Michaelmas Term 2023

A thesis submitted to the Board of the Faculty of Mathematical, Physical & Life Sciences in partial fulfilment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Abstract

The 2-oxoglutarate (2OG) dependent oxygenases are a superfamily of enzymes that utilise an Fe(II) cofactor and 2OG co-substrate for the catalysis of oxidative reactions. It is widely considered that an Fe(II)-binding motif comprising two histidine residues and a third acidic residue, thus creating an HXD/E…H triad, is a defining characteristic of 2OG oxygenase active sites.

Aspartate/asparagine-β-hydroxylase (AspH) is a 2OG-dependent oxygenase that catalyses the hydroxylation of β-carbons of aspartate and asparagine residues in epidermal growth factor-like domains. By contrast with most other 2OG-dependent hydroxylases, AspH employs an Fe(II)-binding dyad, consisting of two histidine residues only. The significance of this unusual Fe(II)-binding motif and any effect this may have on the AspH mechanism is currently unknown. Chapter 1 provides an overview of the consensus mechanism for 2OG-oxygenases, and the current understanding of AspH structure and function.

Chapter 2 describes studies investigating the nature of Fe(II)-binding at the AspH active site through the use of site-directed mutagenesis studies. The discovery that AspH retains significant levels of hydroxylation activity with a single active site Fe(II)-ligating histidine residue was investigated in detail.

Chapter 3 describes work which explores the nature of 2OG-binding in AspH through the use of clinically relevant AspH variants. An unambiguous link between impaired AspH activity and pathogenic mutations in the AspH gene is established by characterisation of these AspH variants.

Chapter 4 outlines work that establishes an AspH microcrystal system that could be used for future time-resolved serial X-ray crystallography experiments. A method for generating slurries of AspH microcrystals was developed and the system was tested at a synchrotron beamline. Additionally, a structure of AspH was obtained using serial crystallography at an X-ray free electron laser (XFEL) beamline.

Attempts to recombinantly express a predicted 2OG-dependent oxygenase, Aspartate-β-hydroxylase domain-containing protein 1 (AspHD1), are outlined in Chapter 5. AspHD1 is predicted to possess a highly unusual RXG…H active site Fe(II) binding motif.

Overall, as described in Chapter 6, this work characterises the unusual AspH Fe(II)-binding dyad and provides further insight into the mechanism of AspH catalysis. It is hypothesized that the side chains of AspH substrate aspartate residues may have a role in metal templating at the AspH active site. This hypothesis requires further consideration, but could have implications for the assignment of metal-ligating protein active sites.

Acknowledgements

Firstly, I would like to thank Prof Chris Schofield for being a fantastic supervisor. It has been a pleasure to work in the CJS group under his supervision and I am grateful for the guidance and encouragement I have received throughout my PhD.

I would like to thank Dr Lennart Brewitz for introducing me to the world of AspH and for advice during my PhD. My thanks go to Dr Patrick Rabe for assistance with protein preps, crystallography, and for approaching every experiment with impressive levels of enthusiasm. I would also like to thank Dr Stephen Marshall for his patience with answering my endless biochemistry and crystallography questions, for proofreading, and for proving that tea time chats can be highly productive.

I am indebted to so many other colleagues who have shared their time and expertise with me. In particular, Dr Anthony Tumber for support with the RapidFire, Dr Adam Hardy for ensuring the lab is run smoothly, and Dr Mark Allen for the regular supply of biscuits. My thanks also go to Dr Allen Orville and members of the XFEL Hub at Diamond Light Source including Dr Pierre Aller, Dr Agata Butryn, Dr Anastasiia Shilova, Dr Jos Kamps and Dr Tiankun Zhou. Additionally, I am grateful for the assistance from the groups of Dr Jan Kern and Dr Junko Yano for collaboration with XFEL experiments. Many thanks also to Ms Cynthia Hou for being a great part II student, and Zahn and Gwen for administrative assistance.

I feel incredibly lucky for the support I have had from my friends throughout my PhD. Giorgia and Emily your friendship throughout all the highs and lows has been invaluable to me, Freddie and James thanks for always being happy to help, and Raph, Klem, Ziggy and Tom, and all the other CJS members who have made it a great place to work. Special thanks to Gitu, Lewis and Jamie for sharing another four years of chemistry together, and to Sophie for her continual wisdom.

An enormous thanks goes to my mum, dad and Chris and Minim for their endless encouragement and support throughout my PhD and beyond. And finally, thank you Sean for being amazing and for everything you do.

Declaration and author contributions

I declare that I carried out all work carried described in this thesis, except where stated otherwise in chapter acknowledgements and below:

The plasmid encoding the human gene for $AspH_{315-758}$ was provided by Dr Inga Pfeffer.

Chapter 2

The plasmids encoding the H679A, H725A, G681D and G681E AspH variants were provided by Dr Inga Pfeffer. Dr Lennart Brewitz carried out NMR uncoupled turnover assays and determined IC_{50} values for the H679A and H725A variants. The crystal for the structure PDB 7E6J was harvested by Dr Yu Nakashima and the data were refined by Dr Yu Nakashima. Crystal data for structure PDB 6Q9I was refined by Dr Rashed Chowdhury and crystallisations were performed by Dr Inga Pfeffer.

Chapter 3

Ms Cynthia Hou, a part II student under my supervision, performed site directed mutagenesis to produce the G434V AspH variant plasmid, performed the purifications of the R735Q, R735W, R688Q and G434V AspH variants and determined the kinetic parameters with respect to 2-oxoglutarate, Fe(II) and L-ascorbic acid.

Chapter 4

Dr Stephen Marshall, Dr Patrick Rabe and Ms Mariska De Munnik carried out batch crystallisation of AspH at Linac Coherent Light Source. XFEL data were collected with the assistance of the groups of Dr Jan Kern and Dr Junko Yano. All serial data on AspH was processed by members of the XFEL Hub at Diamond Light Source, including Dr Tiankun Zhou and Dr Anastasiia Shilova.

Chapter 5

The plasmid encoding for the H679R AspH variant was provided by Dr Inga Pfeffer.

Table of Contents

Chapter 1 | Introduction 1

1.1 Metalloproteins 2

1.2 2-Oxoglutarate dependent oxygenases 3

1.2.1 The cupin superfamily 3
1.2.2 The structure and function of 2OG-dependent oxygenases 6
1.2.3 The catalytic mechanism of 2OG-dependent oxygenases 9
1.2.3.1 2OG-Dependent halogenases 11
1.2.3.2 Modification of the Fe(II)-binding motifs in 2OG-dependent oxygenases 12

1.3 Iron-binding motifs in 2OG-dependent oxygenases 13

1.3.1 JARID2 15
1.3.2 PHF2 15
1.3.3 Hairless 16
1.3.4 Aspartate/asparagine-β-hydroxylase 17
1.3.4.1 Epidermal growth factor-like domains 18
1.3.4.2 The clinical relevance of AspH 20
1.3.4.3 AspH structure 21
1.3.4.4 AspH homologues: Aspartate-β-hydroxylase domain-containing proteins 26

1.4 Aims of the work described in this thesis 29

Chapter 2 | Investigation into the mechanism of AspH catalysis using Fe(II)-binding variants 31

2.1 Introduction 32

2.2 Aims of the work outlined in this chapter 34

2.3 Results and Discussion 34

2.3.1 Production and purification of recombinant AspH variants 34
2.3.2 Development of AspH activity assays 36
2.3.3 Initial assessment of the activity of the H679A and H725A AspH variants 37
2.3.3.1 Uncoupled turnover 39
2.3.4 Determination of kinetic parameters of the H679A and H725A AspH variants 41
2.3.5 Inhibition studies on the H679A and H725A variants 44
2.3.6 Crystallographic studies of the H679A and H725A AspH variants 46
2.3.6.1 Optimisation of crystallisation procedures 48
2.3.7 The mechanism of oxygen incorporation into the product 51
2.3.7.1 The pH dependence of the reaction pathway 56
2.3.7.2 The mechanism of oxygen incorporation in AspH 57
2.3.8 Investigation into the incorporation of anions in AspH reaction 61
2.3.9 Study of an AspH catalytic triad 62

2.4 Conclusions and future work 65

2.4.1.1 Future work: engineering of AspH to catalyse halogenation reactions 67

2.5 Acknowledgements 69

2.6 Supplementary information 70

Chapter 3 | Characterisation of clinically relevant AspH 2OG-binding variants 73

3.1 Introduction 74

3.2 Aims of the work described in this chapter 76

3.3 Results and discussion 77

3.3.1 Purification of clinically-relevant AspH variants 77

3.3.2 Characterization of clinically relevant 2OG-binding AspH variants 79

3.3.2.1 Initial activity assays 79

3.3.2.2 Determination of kinetic parameters 80

3.3.2.3 Structural studies 83

3.3.2.4 The effect of clinically relevant 2OG-binding mutations on AspH 94

3.3.3 Characterization of AspH 2OG variant activity with 2OG derivatives 96

3.3.3.1 2OG derivatives can improve the activity of AspH 2OG binding variants 97

3.3.3.2 Structural studies of 2OG derivatives with the R735Q and R688Q variants 101

3.3.3.3 Treatment of FDLAB syndrome with 2OG derivatives 104

3.4 Conclusions and future work 105

3.5 Acknowledgements 106

3.6 Supplementary information 107

Chapter 4 | Towards time-resolved X-ray crystallography studies on AspH 111

4.1 Introduction 112

4.1.1 Methods in structural biology 112

4.1.2 Serial X-ray crystallography 114

4.1.3 Time-resolved serial femtosecond protein crystallography 115

4.2 Aims of the work described in this chapter 119

4.3 Results and discussion 119

4.3.1 Optimisation of the purification of recombinant AspH 119

4.3.2 Investigations into the structure of AspH using macrocrystals 123

4.3.2.1 Aerobic AspH crystallography 123

4.3.2.2 Assessment of AspH activity *in crystallo* 130

4.3.3 Towards time-resolved serial X-ray crystallography on AspH 137

4.3.3.1 Production of AspH microcrystals 137

4.3.3.2 Fixed target serial synchrotron crystallography 139

4.3.3.3 XFEL experiments at Linac Coherent Light Source 143

4.3.4 Serial crystallography on AspH 147

4.4 Conclusions and future work 148

4.5 Chapter acknowledgements 150

4.6 Supplementary information 151

Chapter 5 | AspH homologue: Aspartate-β-hydroxylase domain-containing protein 1 159

5.1 Introduction 160

5.2 Aims of the research described in this chapter 164

5.3 Results and discussion 164
5.3.1 AspHD1 construct design 164
5.3.2 Initial AspHD1 expression trials 166
5.3.3 AspHD1 expression in *E. coli* optimisation: the competent cell strain 168
5.3.4 AspHD1 expression in *E. coli* optimisation: fusion tags 169
5.3.5 Purification of MBP-His_6-$AspHD1_{206-390}$ produced in *E. coli* 171
5.3.6 A model system to assess arginine as a metal binding residue 174

5.4 Conclusions and future work 181

5.5 Chapter acknowledgements 183

5.6 Supplementary information 184

Chapter 6 | Conclusions and future work 185

Chapter 7 | Materials and methods 193

7.1 General procedures 194

7.2 Molecular biology methods 194
7.2.1 General methods 194
7.2.2 DNA concentrations and sequencing 194
7.2.3 Plasmids and genes 195
7.2.4 Primers 195
7.2.5 PCR 197
7.2.6 Agarose gel electrophoresis 198
7.2.7 Minipreps 199
7.2.8 Gibson assembly protocol 200
7.2.9 Insertions, deletions and substitutions of DNA sequences 201

7.3 Protein expression and purification 202
7.3.1 Agar plates 202
7.3.2 Media 202
7.3.3 Transformations 202
7.3.4 Starter cultures 203
7.3.5 Expression trials 203
7.3.6 Large scale protein expression 204
7.3.7 His_6-tagged protein purification 205
7.3.8 MBP-tagged protein purification 206
7.3.9 SDS PAGE analysis 206
7.3.10 Western Blots 207

7.4 Crystallography 209
7.4.1 Crystallisation buffer preparation 209
7.4.2 Seed stock preparation 209
7.4.3 Crystallisation methods 209
7.4.4 Crystal harvesting 211
7.4.5 Data collection and processing 212
7.4.6 Anaerobic crystallography 212

7.5 Analytical methods 213
7.5.1 RapidFire MS 213
7.5.2 Xevo MS 214

7.5.3 NMR 214
7.6 Assays 214
7.6.1 General procedures 214
7.6.2 RapidFire timecourse assays 215
7.6.3 AspH variant inhibition assays 216
7.6.4 2OG derivative assays 217
7.6.5 Oxygen isotope studies 217
Chapter 8 | References 221
Chapter 9 | Appendix 239
9.1 Protein sequences 240
9.2 Additional data 241
9.2.1 The pH dependence of $^{18}O_2$ reactions 241
9.2.2 Time-course data for Chapter 3 242
9.3 ACIE publication 245

Abbreviations

2,4-PDCA	2,4-Pyridinedicarboxylic acid
2OG	2-oxoglutarate
2YT	2x Yeast Extract Tryptone medium
ADE	Acoustic droplet ejector
AIM	Autoinduction media
AlkB	Alkane dioxygenase
AspH	Aspartate/asaparaginyl-β-hydroxylase
AspHD1	Aspartate/asaparaginyl-β-hydroxylase domain-containing protein 1
AspHD2	Aspartate/asaparaginyl-β-hydroxylase domain-containing protein 2
AsqJ	*Aspergillus nidulans* dioxygenase
BHya	3-hydroxy-L-aspartate
BHyn	3-hydroxy-L-asparagine
CAS	Clavaminic acid synthase
DAH	Dechloroacutumine halogenase
DAO	Diamine oxidase
DAOCS	Deacetyloxycephalosporin C synthase
DMSO	Dimethylsulfoxide
DSBH	Double-stranded beta-helix
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EGFD	Epidermal growth factor-like domain
EM	Electron microscopy
ESI	Electrospray ionisation
FAS	Ferrous ammonium iron sulphate
FDLAB	Facial dysmorphism lens dislocation
FIH	Factor inhibiting HIF
GDVN	Gas dynamic virtual nozzle
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hF	Human coagulation factor
hFX	Human coagulation factor X
hFX-Asp_{103}	Synthetic AspH substrate based on the sequence of hFX
hFX-Asn_{103}	Synthetic AspH substrate based on the sequence of hFX with an Asn substitution
HIF	Hypoxia-Inducible Factor
HPLC	High Performance Liquid Chromatography
IC_{50}	Half maximum inhibitory concentration
JmjC	Jumonji C
JMJD1C	Jumonji domain containing 1C
JHDM	Jumonji-domain histone demethylase
Kan	Kanamycin
k_{cat}	Turnover number
KDM	Lysine demethylase
K_m	Michaelis constant
LAA	L-ascorbic acid
LB	Luria-Bertani broth
LCLS	Linac coherent light source
LC-MS	Liquid-chromatography-mass spectrometry
MBP	Maltose-binding protein

MES	2-(N-morpholino)ethanesulfonic acid
MS	Mass spectrometry
MW	Molecular weight
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nm
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Poly-ethylene glycol
PEG 3350	Poly-ethylene glycol with an average molecular weight of 3350
PHD	Plant homeodomain
PHD2	HIF prolyl hydroxylase domain-2
PHF2	PHD finger protein 2
ppm	Parts per million
RMSD	Route mean square deviation
RNA	Ribonucleic acid
SACLA	Spring-9 Angstrom Compact free electron LAser
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SEC	Size exclusion
SFX	Serial femtosecond X-ray
SPE-MS	Solid phase extraction mass spectrometry
SSX	Serial synchrotron X-ray
SUMO	Small ubiquitin modifying protein
SyrB2	*Pseudomonas syringae* halogenase
TauD	Taurine dioxygenase
TB	Terrific broth
TfdA	2,4-dichlorophenoxyacetic acid dioxygenase
TPR	Tetratricopeptide repeat
Tr	Time-resolved
Tris	Tris(hydroxymethyl)aminomethane
Tr-SFX	Time resolved serial femtosecond X-ray
UV	Ultra-violet
v/v	percentage volume
vis	visible
V_{max}	Maximum velocity of an enzyme reaction
WelO5	*Hapalosiphon welwischii* halogenase
XFEL	X-ray free electron laser

Standard 3 letter or single letter abbreviations for amino acids, standard SI unit abbreviations and chemical element symbols are used throughout this thesis.

Chapter 1 | Introduction

1.1 Metalloproteins

The diversity of roles carried out by proteins across Nature is facilitated by the use of cofactors. Cofactors, such as small organic molecules, metal ions or non-metal and metal clusters, expand the functional capabilities of proteins beyond what would be achievable from simply using amino acid side chains.[1]

Metalloproteins contain at least one metal ion cofactor, which is bound to the protein by prosthetic groups and/or side chain residues.[2] Metalloproteins perform a wide range of functions that are fundamental to life, including redox sensing, electron transfer, gas sensing and transport, and catalysis of a variety of reactions.[3–11] Metal ions such as Ca(II), Zn(II), or Mg(II) frequently adopt structural roles in stabilising quaternary protein structures or can enhance the nucleophilicity of coordinating groups, thus catalysing a reaction.[12–14] Redox active metals are employed as cofactors in enzyme-catalysed oxidative reactions, with Fe, Cu, Mo, Ni and Co commonly used for this purpose.[14–18] Use of dioxygen for incorporation of oxygen into a substrate is thermodynamically favourable, but kinetically slow due to an electron spin pairing mismatch.[19] As such, biology has evolved to involve catalytic transition metal centres, such as Fe or Cu, for this purpose.[18]

Fe is the fourth most abundant element in the earth's crust and is probably essential in all organisms.[20] As a redox active transition metal, Fe has catalytically useful redox reaction capabilities and also possesses ligand-binding properties that are exploited in a range of systems.[20] In bacteria, Fe is often found in FeS clusters that facilitate electron transport through proteins and are fundamental to biological functions such as nitrogen fixation and metabolic processes involving H_2.[21,22] In higher organisms, Fe is required by proteins that are essential to metabolic pathways, including oxygen transport, DNA synthesis and electron transfer.[23,24]

Proteins that employ Fe cofactors can be subdivided into to two categories: heme proteins that use a prosthetic porphyrin ring to bind Fe, and nonheme proteins that ligate Fe directly using protein side chain residues.[25] Owing to the versatility of the heme group, heme proteins are

involved in a range of systems including oxygen transport and storage, electron transfer, catalysis and control of gene expression.[26,27] The delocalised π system of porphyrin rings means that heme proteins possess characteristic spectroscopic features that enable their purification and characterisation by various spectroscopic methods, such as UV-vis or vibrational spectroscopy.[28–30] The oxygen transport heme proteins myoglobin and haemoglobin have been extensively studied; the solution of structures were pioneering developments in protein structure determination by John Kendrew and Max Perutz in the mid-twentieth century.[31]

Non-heme Fe-dependent enzymes have a diverse range of roles across Nature and are of particular importance in the catalysis of reactions involving dioxygen.[32] The mononuclear non-heme iron enzymes can be classified into two categories; those which use Fe(III) centres for substrate activation, and those which employ an Fe(II) cofactor for oxygen activation in a comparable manner to heme-iron enzymes.[25] The mechanism of reactions catalysed by non-heme iron enzymes can be more challenging to study than that of their heme-containing counterparts, as Fe(II) active site centres are often considered to be spectroscopically "silent".[32] Nonetheless, significant developments have been made in the field of mononuclear non-heme iron enzyme mechanistic studies by use of methods such as Mössbauer spectroscopy, variable temperature magnetic circular dichroism spectroscopy, electron paramagnetic resononance (EPR) spectroscopy, and structural and kinetic studies.[32]

1.2 2-Oxoglutarate dependent oxygenases

1.2.1 The cupin superfamily

The cupin structural superfamily encompasses a broad range of functionally diverse proteins.[33] The cupins have roles ranging from the modification of bacterial cell walls, storage proteins in plants, to transcription regulation in mammals.[34–36] Cupin proteins typically exhibit low amino acid sequence similarity to one another, but have a conserved quaternary structure which can give a better indication of evolutionary relationships between the subfamilies.[37–39]

The name cupin was taken from the Latin 'cupa', meaning small barrel, which aptly describes the characteristic three-dimensional shape adopted by the cupin domain.[38] The barrel-like structure is typically constructed from an eight-stranded double-stranded beta-helix (DSBH) core fold. Proposed consensus sequences for the conserved motif regions are $G(X)_5HXH(X)_{3-4}E(X)_6G$ (motif 1) and $G(X)_{5-7}PXG(X)_2H(X)_3N$ (motif 2).[40] However, analysis of large number of cupin sequences later indicated that the motifs are not as highly conserved as originally proposed.[41]

Metal ions, such as Mn(II), Ni(II) and Fe(II), can be ligated by cupin proteins using residues that are typically located on the β-sheets motifs.[40] The metal binding patterns vary throughout the cupin superfamily, but are often conserved within a subfamily. The archetypal cupin fold protein, germin, employs two His residues and a Glu residue (Motif 1), and a His residue (Motif 2) as Mn(II) binding ligands at the active site, as observed in phosphomannose isomerase.[42] In some isomerases and dioxygenases, one of the Motif 1 His residues is substituted for a Gln, Asp or Glu residue for metal binding.[41] An overview of representative examples of different metal binding motifs found in cupin fold-containing proteins is shown in Figure 1-1 and summarised in Table 1-1.

Histidine is the most commonly employed residue in cupin metal-binding motifs, with other frequently employed metal binding residues including Asp and Glu. Less common motifs involve neutral side chains of amino acid residues such as Gln and Tyr. The coordination environment of a metal ion in cupin proteins is typically (but not always) octahedral overall, at least in the resting states, and the number of protein-based ligands can range from two (SyrB2, a halogenase) to four (CarO, a retinal-forming carotenoid dioxyenase).[43] It is reported that artificial modification of the metal coordination environment can alter reaction selectivity and the stereochemistry of products, or completely eliminate catalytic activity.[44,45]

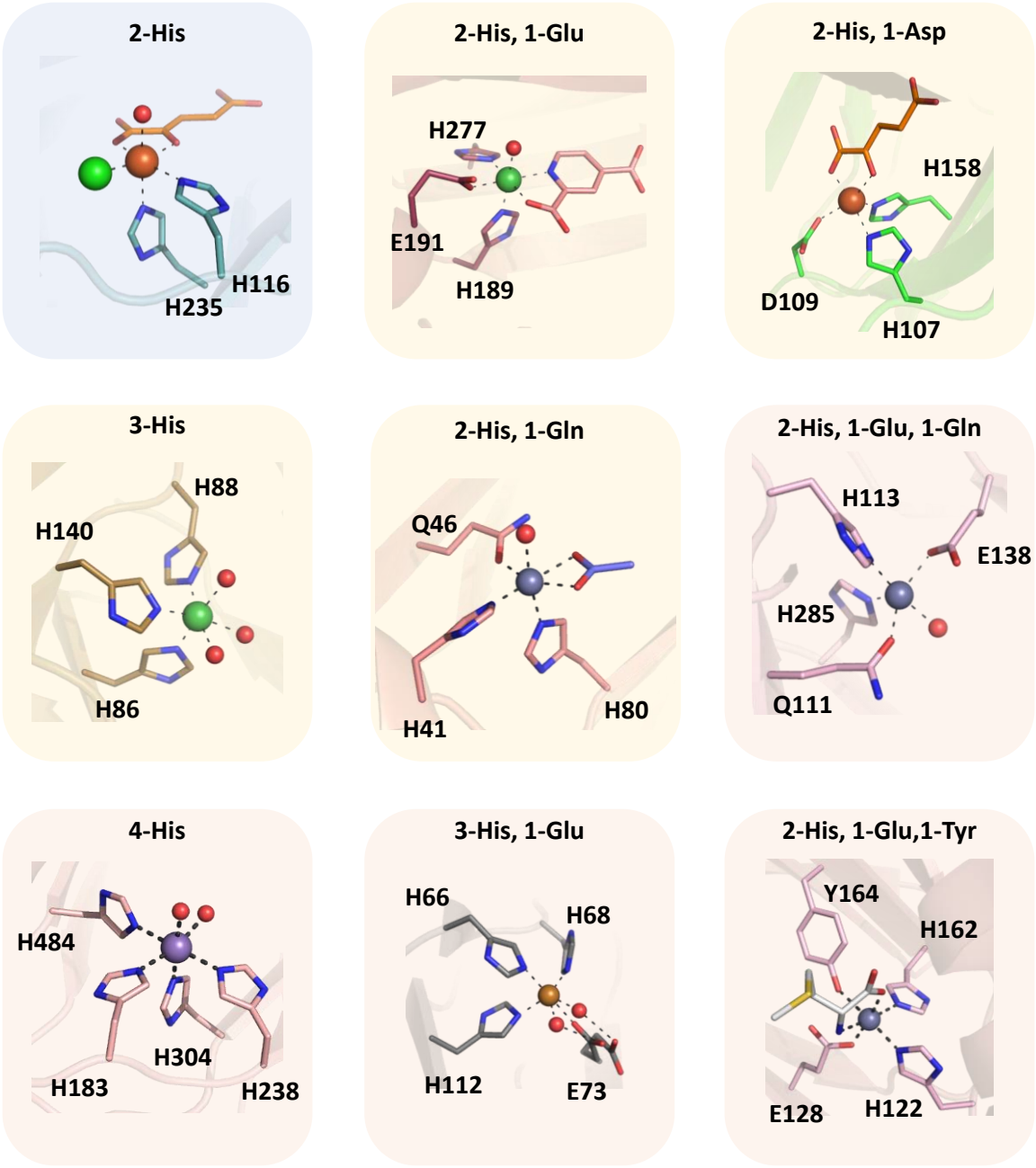


Figure 1-1. Examples of metal binding patterns in cupin fold domain in proteins with 2, 3 and 4 protein residues contributing to metal coordination. 2-His: SyrB2 halogenase (PDB 2FCT);[46] 2-His, 1-Glu: Histone demethylase KDM4A-E (PDB 2W2I);[47] 2-His, 1-Asp: proline-3-hydroxylase (PDB 1E5S);[48] 3-His: Cysteine dioxygenase (PDB 2ATF)[49]; 2-His 1-Gln: Tetracenomycin polyketide synthesis protein (PDB 3H50);[50] 2-His, 1-Glu, 1-Gln: Phosphomannose isomerase (PDB 1PMI); 4-His: Carotenoid oxygenase (PDB 2BIW);[51] 3-His, 1-Asp/Glu – Quercetin 2,3-dioxygenase (PDB 1JUH); 2-His 1-Glu 1-Tyr: *V. cholerae* transcriptional regulator (PDB 1Y9Q).

Table 1-1. Selected examples of metal-binding motifs reported in the cupin family. The list is non-exhaustive with one example given per binding motif.

Number of protein residues	Metal-binding motif	Enzyme	Species	Metal	PDB
2	2-His	SyrB2	*P. syringae pv. syringae*	Fe	2FCT[46]
3	3-His	Cysteine dioxygenase	*M. musculus*	Fe	2ATF[49]
	2-His, 1-Asp	Proline-3-hydroxylase	*S. griseoviridus*	Fe	1E5S[48]
	2-His, 1-Glu	Histone demethylase KDM4A-E	*H. sapiens*	Fe	2W2I[47]
	2-His, 1-Gln	Tetracenomycin polyketide synthesis protein	*X. campestris*	Zn	3H50[50]
4	4-His	Carotenoid oxygenase	*Synechocystis*	Fe	2BIW[51]
	3-His, 1-Glu	Quercetin 2,3-dioxygenase	*A. japonicus*	Cu	1JUH[52]
	3-His, 1-Gln	Bacilysin biosynthesis protein BacB	*B. subtillis*	Mn	3H7J[53]
	2-His, 1-Glu, 1-Gln	Phosphomannonse isomerase	*C. albicans*	Zn	1PMI[54]
	2-His, 1-Glu, 1-Tyr	Transcriptional regulator	*V. cholerae*	Zn	1Y9Q

1.2.2 The structure and function of 2OG-dependent oxygenases

The 2OG-dependent oxygenases are the largest subfamily of proteins in the cupin superfamily. 2OG-dependent oxygenases are non-heme Fe proteins which employ an Fe(II) cofactor and use 2OG and O_2 as cosubstrates to catalyse oxidative reactions.[55] In animals, 2OG oxygenases are reported to only catalyse hydroxylation and closely related reactions, e.g. demethylation via hydroxylation. Across Nature, 2OG-dependent oxygenases catalyse a range of other reactions, including halogenations, demethylations, ring closures and desaturations (Figure 1-2).[55–59] These reactions are integral to processes such as nucleic acid repair, protein biosynthesis, transcriptional regulation, hypoxia sensing and antibiotic synthesis.[60–65]

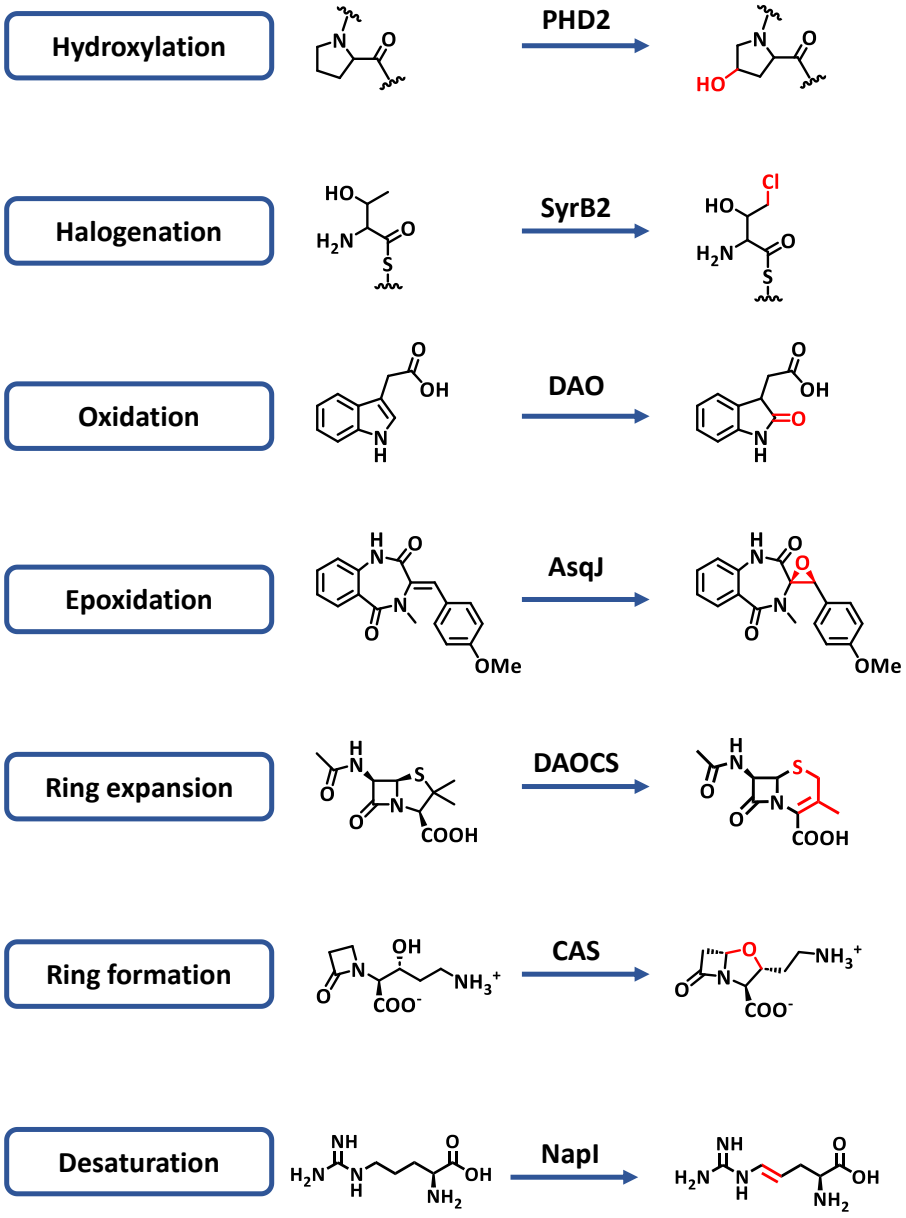


Figure 1-2. Examples of reactions catalysed by 2OG-dependent oxygenases. PHD2 = Hypoxia inducible factor (HIF) prolyl hydroxylase domain-2; SyrB2 = *Pseudomonas syringae pv. syringae* halogenase; DAO = Auxin dioxygenase; AsqJ = *Aspergillus nidulans* dioxygenase; DAOCS = deacetoxycephalosporin C synthase; CAS = clavaminic acid synthase; NapI = L-arginine 4,5-desaturase in the naphthyridinomycin synthesis pathway.

Examination of the crystal structures of 2OG-dependent oxygenases has revealed a conserved cupin fold (double-stranded beta-helix (DSBH) or jumonji domain) is an integral structural feature of the family (Figure 1-3a).[36] The DSBH domain of 2OG-dependent oxygenases consists of antiparallel β-strands that form the characteristic cupin barrel structure.[36] The DSBH core houses the 2OG-dependent oxygenase active site, in which a conserved set of protein residues ligate an

Fe(II) cofactor. Sequence alignments show that Fe(II) is typically ligated by a HXD/E…H triad, with aspartate more commonly occurring than glutamate (Figure 1-3b).[66] In human 2OG-dependent oxygenases, glutamate forms part of the triad in the histone demethylase (KDM) 3, KDM4 and KDM6 subfamilies.[67] 2OG is held at the active site by bidentate coordination to Fe(II) through its C1 carboxylate and C2 keto oxygen, and with electrostatic interactions to key active site residues. The latter usually includes a salt bridge between the C5 2OG carboxylate and a basic protein residue side chain, typically a lysine or arginine residue.[55]

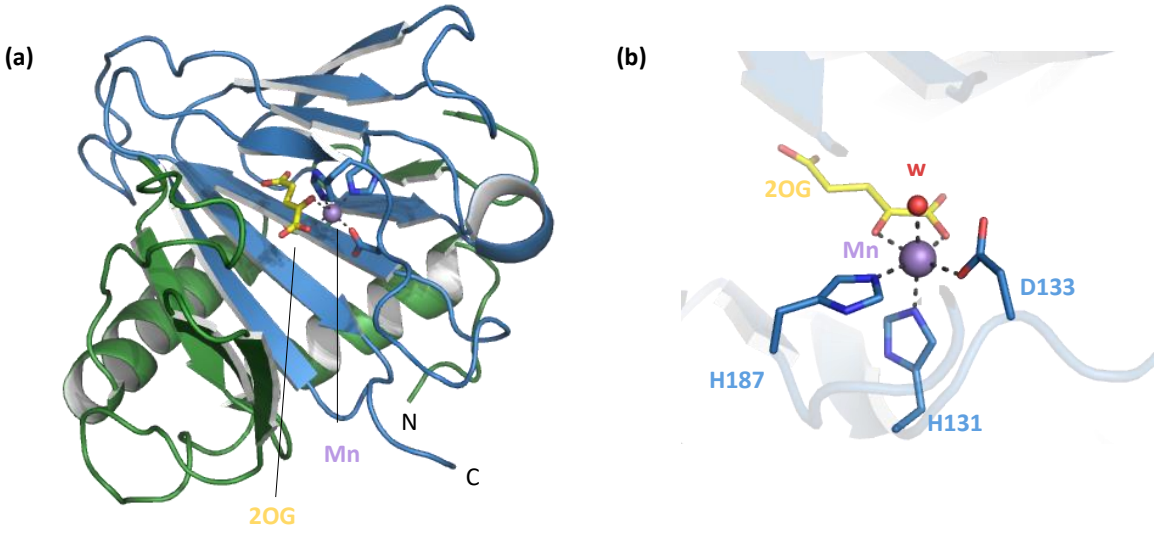


Figure 1-3. Views from a crystal structure of the 2OG-dependent oxygenase AlkB. (a) The overall fold of AlkB. The oxygenase domain including the cupin fold is shown in blue. (b) View from the active site of AlkB. The Fe(II)-binding triad consists of H187, H131 and D133.

The HXD/E…H Fe(II) binding triad is very highly conserved in 2OG-dependent oxygenases, deviations from the pattern do, however, exist. The most common exception to the triad is observed in the 2OG-dependent halogenase subfamily, members of which possess a conserved HXG…H Fe(II)-binding dyad.[46,68–72] The absence of a third acidic residue leaves a vacant coordination site around Fe(II), thus allowing coordination of a halogen ion to Fe(II) during catalysis.[73,74]

1.2.3 The catalytic mechanism of 2OG-dependent oxygenases

The consensus mechanism for 2OG-dependent oxygenases was first based on model hydroxylations and has since been supported by subsequent experimental findings (Figure 1-4).[75,76] Prior to catalysis, the Fe(II) cofactor is coordinated to the enzyme by three residues in the HXD/E…H triad, and typically three water molecules, which complete the octahedral coordination environment around Fe(II) (Figure 1-4(i)).[66] The consensus catalytic mechanism is initiated by 2OG binding bidentately to Fe(II), displacing two water molecules in the process (Figure 1-4 (ii)). Crystal structures indicate that the 2OG C2 keto oxygen typically binds in the coordination site opposite the aspartate/glutamate protein residue of the triad and the C1 carboxylate group coordinates opposite to one of the histidine residues of the triad, although which His this is can vary and some mechanistic studies have proposed more than one 2OG binding configuration occur during catalysis.[77–79] The formation of a protein:2OG:Fe(II) complex is evidenced by changes in the UV-visible spectra, as the 2OG complex typically exhibits a metal-to-ligand charge-transfer band around 530 nm, while the apo-protein is colourless.[80–82]

Substrate binding at the active site in close proximity to the Fe(II) centre initiates the next stage in the catalytic cycle (Figure 1-4 (iii)). Substrate binding is typically accompanied by promotion of the loss of a Fe(II)-coordinated water molecule from the active site and formation of a 5-coordinate square pyramidal Fe(II) complex.[83,84] UV-vis data show that the protein:2OG:Fe(II) metal-to-ligand charge-transfer band at 530 nm is perturbed by the addition of substrate.[80,85]

In the next step of the mechanism, O_2 coordinates, likely in an end-on manner to the Fe(II) in the vacant coordination site created by the loss of a third water molecule, generating a superoxide intermediate (Figure 1-4(iv)). The Fe(III)-superoxo species (Figure 1-4(iv)) is short-lived and has not been observed spectroscopically in a 2OG-oxygenase; however, the intermediate has been observed in the related oxidase isopenicillin N synthase.[75,86] The free oxygen atom can then perform nucleophilic attack on C2 of 2OG to form a cyclic intermediate. Two possible pathways have been proposed for this stage in the mechanism. The first involves the formation of a cyclic

peroxy intermediate (Figure 1-4(v_a)) with subsequent oxidative decarboxylation to release CO_2. Alternatively, a peroxysuccinate intermediate may be directly formed ((Figure 1-4(v_b)), as observed in an arginine hydroxylase mechanism (VioC).[87,88]

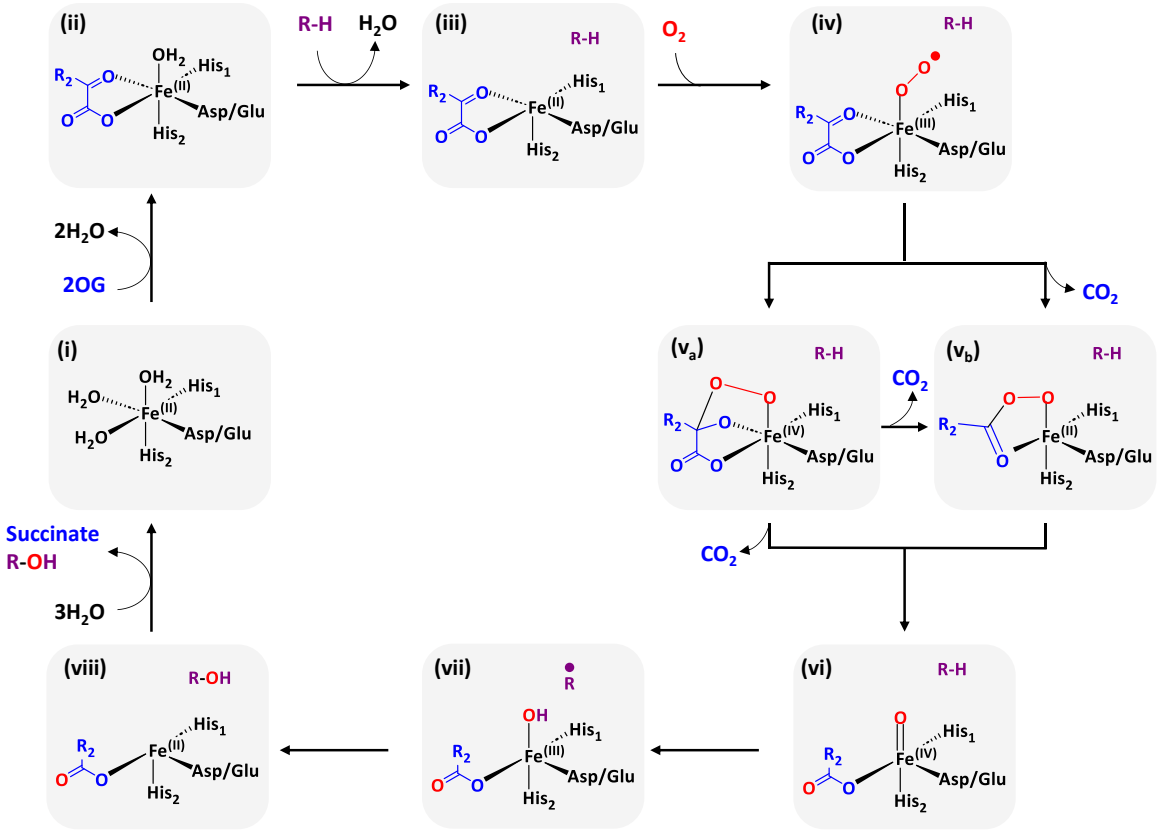


Figure 1-4. The consensus mechanism for hydroxylation catalysed by 2OG-dependent hydroxylases. His_n = histidine residues of the HXD/E…H triad. R = substrate. R_2 = $CH_2CH_2CO_2H$. Variations on this general mechanism may occur.

Subsequent oxidative decarboxylation of the cyclic intermediate results in the cleavage of the O-O bond and formation of co-products CO_2 and succinate (Figure 1-4(vi)). This step also leads to the formation of an Fe(IV) intermediate (the ferryl intermediate) which has been characterised using UV-vis, Mössbauer, extended X-ray absorption fine-structure spectroscopy (EXAFS), and EPR spectroscopy.[89–92] Oxygen-isotope labelling has also confirmed that one oxygen atom in the succinate coproduct originates from O_2.[93]

The ferryl species can then abstract a hydrogen atom from the substrate, forming a substrate radical and an Fe(III)-OH complex (Figure 1-4(vii)).[94] The Fe(III)-OH intermediate has not been

observed by experimental means; however, evidence for the complementary substrate radical has been reported.[95] The Fe(III)-bound hydroxyl then combines with the substrate radical in an "oxygen rebound" step, forming the hydroxylated product (Figure 1-4(viii)). Oxygen labelling experiments supported the hypothesis that substrate hydroxy predominantly originates from O_2.[75] The product and succinate then dissociate from the active site to complete the cycle.

1.2.3.1 2OG-Dependent halogenases

2OG-dependent halogenases do not possess the acidic residue of the HXD/E…H Fe(II) binding triad.[75] This difference creates a vacant coordination site around the Fe(II) centre which can be occupied by a halogen ion during catalysis.[71] Consequently, the mechanism of halogenation was predicted to follow a similar mechanism to that of the 2OG-dependent hydroxylases, with the addition of an initial binding step of a halide ion at the Fe(II) centre prior to 2OG coordination.[96,97] The coordinated halide ion can recombine with the substrate radical in the final step of the mechanism, instead of the hydroxyl group, thereby forming the halogenated product (Figure 1-5).[98]

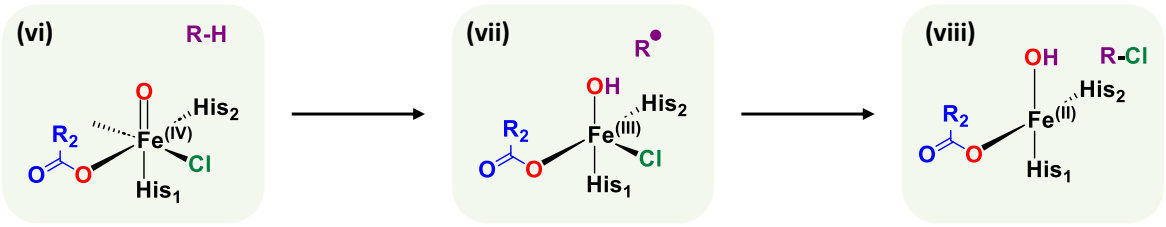


Figure 1-5. The predicted final steps in the mechanism of chlorination catalysed by 2OG-dependent halogenases. Recombination of the substrate radical with either OH or Cl in intermediate (vii), may, in theory, be possible. However, other factors, such as substrate positioning or secondary sphere polar residues, are proposed to direct the reaction pathway to lead to recombination of Cl and not OH. His_n = histidine residues of the HXX…H dyad. R = substrate. $R_2 = CH_2CH_2CO_2H$.

The selectivity of the final rebound step that leads to recombination of the halide ion and not the hydroxyl group is of mechanistic interest and the subject of many investigations (Figure 1-5, the conversion of intermediate (vii) to intermediate (viii)).[96,99–101] This step is proposed to be determined by a number of factors and selectivity is not simply controlled by the coordination environment around the Fe(III) centre.[99,102,103] Secondary sphere active site interactions that direct the orientation of the substrate or interact with the Fe(III)-OH intermediate are postulated to be

influential in directing the selectivity of the final rebound step.[103] It is possible that the mechanism of the final steps differs in halogenase catalysis compared to hydroxylation, with some studies suggesting there may be isomerisation of the haloferryl intermediate (Figure 1-5(vi) or Figure 1-5(vii)), prior to radical recombination of the halogen radical and substrate.[99,104]

The derivatization of inactivated sp^3 C-H bonds under ambient conditions is of interest from a biocatalytic perspective.[105,106] Most, but not all, identified 2OG-dependent halogenases catalyse modifications to protein-tethered substrates, which adds a level of complexity to studying their mechanisms and structures. By contrast, the 2OG-dependent hydroxylases accept a wider variety of substrates, ranging from small molecules to larger protein complexes. Therefore engineering of the hydroxylases to catalyse non-native reactions, including halogenations, may have more potential applications in biocatalytic technology.

1.2.3.2 Modification of the Fe(II)-binding motifs in 2OG-dependent oxygenases

Modifications to the Fe(II)-binding motifs in 2OG-dependent oxygenases have different outcomes depending on the identity of the enzyme. In some cases, substitution of a single metal binding-residue with a non-metal ligating residue, such as alanine or glycine, can eliminate catalytic activity. Susbtitution of any of the residues in the HXD…H Fe(II)-binding triad of 2,4-dichlorophenoxyacetic acid dioxygenase (TfdA) to alanine was reported to produce catalytically inactive TfdA variants only.[107] Similarly, mutagenesis studies on the TauD catalytic Fe(II)-binding triad residues provided evidence that substitution of any of the HXD…H triad residues with alanine resulted in virtually inactive TauD variants. However, unexpected levels of catalytic activity were exhibited by the H255E TauD variant, in which one Fe(II)-binding histidine residue was substituted by glutamate.[108] This is somewhat surprising as the introduction of a second negatively-charged amino acid residue at the active site could potentially disrupt the active site.

In other 2OG-oxygenases, substitutions in the HXD/E…H triad are reported to differently affect catalytic activity. The D201G, but not the D201A, variant of factor inhibiting hypoxia inducible

factor (FIH) was reported to catalyse asparagine hydroxylation; evidence from these studies suggested that two-histidine ligands may be adequate for Fe(II)-binding and catalysis in FIH.[109]

Modification of the Fe(II) binding dyads and triads of halogenases and hydroxylases respectively can affect reaction selectivity. Mutation of the active sites of plant halogenase dechloroacumutmine halogenase (DAH) and WelO5, to introduce a third acidic residue at the active site, produced enzyme variants with an HXD…H triad that were only capable of catalysing hydroxylation reactions.[110,111] By contrast, the inverse mutation of a hydroxylase binding triad to a halogenase dyad does not typically result in selective halogenation activity, as evidenced by mutagenesis studies performed on FIH, TauD and prolyl-4-hydroxylase.[109,112,113] Two hydroxylases, *B. ambifaria* SadA and *S. meliloti* SmP4H, have been reported to catalyse modest levels of halogenation following the mutation of their Fe(II)-binding triads to HXG…H dyads.[114,115] An artificially engineered SadA variant was also reported to catalyse the formation of products containing N_3, NCO and OCN.[116]

Engineering of hydroxylases to selectively perform halogenation is achievable.[103] Substitution of a single secondary sphere valine residue with an asparagine introduced halogenation activity in a lysine hydroxlase.[103] Selectivity for halogenation over hydroxylation was improved following the modification of just thirteen additional secondary sphere residues.[103] The amino acid substitutions were proposed to modify electrostatic interactions from the active site to the hydroxyl radical formed during catalysis, and thus make the halogenation pathway more favourable.[103]

1.3 Iron-binding motifs in 2OG-dependent oxygenases

Structurally informed sequence comparisons imply that the Fe(II) binding triad employed by 2OG-dependent oxygenases is highly conserved throughout the superfamily, although the 2OG-dependent halogenase subfamily exemplify a notable conserved exception to the triad. However, deviations from the triad have been recorded in 2OG-dependent oxygenases that

catalyse reactions other than halogenations, and are predicted in 2OG-oxygenases with unknown functions (Table 1-2).

Table 1-2. Examples of members of the 2OG-dependent oxygenase family predicted or confirmed to contain metal-binding motifs which differ from the conserved HXD/E…H triad. Residues which differ from the conserved binding triad are highlighted in red. Motifs have been confirmed by structural studies or are predicted from sequence alignments. Note that the list is non-exhaustive and halogenases listed are representative of the subfamily.

SUB-GROUP	ORGANISM	ENZYME	BINDING MOTIF	FUNCTION
JARID1	*S. pombe*	sp Spbp4h10	**T**X**S**…H	Transcription factor[117]
	S. pombe	sp Q9UT79	**Q**XD…**Y**	Transcription factor[117]
JARID2	*H. sapiens*	JARID2	**S**XD…**V**	Transcription factor[117]
	D. melanogaster	CG3654	**Y**XD…**T**	Transcription factor[117]
JHDM3/JMJD2	*S. cerevisiae*	Gis1	HXD…**Y**	Transcriptional repressor[118]
KDM3C/JMJD1C	*H. sapiens*	Hairless	**C**XE…H	Histone demethylase or regulatory role[119]
KDM7C	*H. sapiens*	PHF2	HXD…**Y**	Histone demethylation or transcription factor[120–122]
HYDROXYLASE	*H. sapiens*	AspH	HX**G**…H	Aspartate/asparagine hydroxylase[123]
PREDICTED OXYGENASE	*H. sapiens*	AspHD1	**R**X**G**…H	Predicted hydroxylase
	H. sapiens	AspHD2	HX**G**…H	Predicted hydroxylase
HALOGENASES	*H. welwischii*	WelO5	HX**G**…H	Chlorination 12-epi-fischerindole U[111]

In humans, aspartate/asaparginyl-β-hydroxylase (AspH), is possibly the most well biochemically characterised 2OG-dependent hydroxylase which possesses a non-standard HXG…H dyad of Fe(II) binding residues.[124] Only a few other 2OG-oxygenase related proteins with unconventional Fe(II)-binding motifs have been biochemically investigated, including Plant Homeodomain (PHD) Finger

Protein 2 (PHF2), Hairless and Jarid 2, and these are proposed to be catalytically inactive.[117] Thus there is scope to further explore the nature and significance, if any, of the non-standard Fe(II)-binding motifs.

1.3.1 JARID2

Sequence alignments of human JARID2, an important regulator of transcription, suggest that it may contain an unusual SXD…V Fe(II)-binding motif. Cryogenic electron microscopy (cryo-EM) structures of human JARID2 in complex with polycomb repressive complex 2 (PRC2) and cofactor adipocyte enhancer-binding protein (AEBP2) have been reported and further work provided evidence for JARID2 activation of PRC2.[125–127] However, the reported structures used a construct of JARID2 that did not contain the predicted Fe(II)-binding motif, and thus no structural characterisation of the predicted SXD…V motif has been performed.

1.3.2 PHF2

PHF2 is one of three members of the KDM7 2OG-dependent oxygenase subfamily.[128,129] The N-terminal domain of PHF2 is composed of PHD and Jumonji domains, and the PHD domain is involved in binding to histone H3 trimethylated lysine 4 (H3K4me3).[121] The sequence of human PHF2 predicts an HXD…Y motif at the active site in the Jumonki domain and a crystal structure of the enzyme confirmed the capacity of the active site to bind a metal ion using these residues (Figure 1-6).[121] Wen et al. reported western blot-based evidence that PHF2 catalyses the demethylation of monomethylated histone H3 lysine 4 (H3K9me1) *in vivo*.[130] However, subsequent mechanistic studies, including using western blot analyses and co-precipitation studies, disputed these findings and proposed a regulatory role for PHF2.[120,122] It has also been proposed that PHF2 alone is enzymatically inactive, and requires protein kinase A-mediated phosphorylation to induce PHF2 demethylation activity.[131,132]

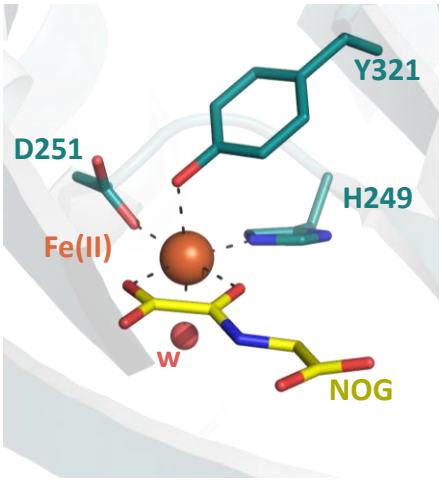


Figure 1-6. View from a crystal structure of the active site of PHF2 in complex with N-oxalylglycine (NOG) and Fe. The active site of PHF2 shows the positioning of Y321 in the active site Fe(II)-binding triad (PDB: 3PU8).[121]

1.3.3 Hairless

Another striking example of a member of the 2OG-dependent oxygenase structural family member with an apparently unusual Fe(II) binding motif is Hairless.[133] The *Hairless* gene encodes for a human protein that is a member of the KDM3 subfamily and was first identified in mice for its link to baldness.[134] In 2006 Klose et al. noted that Hairless "is likely to be deficient in enzymatic activity" owing to its unusual CXE…H binding triad that was presumed to be incapable of ligating a metal.[117] In 2014, in vitro demethylation assays using immunofluorescence staining provided cellular evidence that Hairless can demethylate monomethylated histone H3 lysine 9 (H3K9me1).[119] However, subsequent work performed in 2020 contradicted this finding and did not detect Hairless activity in cell-free demethylase assays using purified methylated H3K9 peptides and validation by Western blot analysis.[135] Thus, the capability of Hairless for catalysis is still under debate. It has been proposed that Hairless only has a regulatory role in cells.[119,135] Cysteine can coordinate metals ions in other biological systems and therefore it is, at least in principle, possible that Hairless is capable of catalysis.[136] However, further work is needed in this area, including on the effect of the sulfur oxidation state, to reveal the role of cysteine at the Hairless active site.[136]

1.3.4 Aspartate/asparagine-β-hydroxylase

AspH catalyses the hydroxylation of β-carbons of aspartate and asparagine residues in epidermal growth factor-like domains (EGFDs) (Figure 1-7). AspH catalysed hydroxylation results in the *erythro* stereoisomers (2*S*,3*R*)-hydroxyaspartate (βHya) or (2*S*,3*R*)-hydroxyasparagine (βHyn). B-hydroxylation of Asp and Asn residues was first identified in the vitamin K-dependent proteins coagulation factor IX (hFIX) and factor X (hFX), and the plasma complement proteins C1r and C1s.[137–140] This was a milestone in the study of post-translation protein modifications, as previously only the modification of collagen domains by atmospheric oxygen had been reported. *In vitro* studies demonstrated that AspH is member of the 2OG-dependent oxygenase family.[141] Characterisation of AspH extracted from bovine and rat livers determined that that Fe(II) and 2OG were a requirement for β-hydroxylation of Asp and Asn residues.[123,142,143]

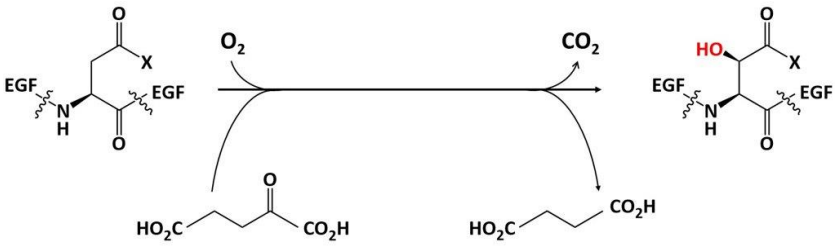


Figure 1-7. A reaction scheme of hydroxylation reactions catalysed by AspH. X = N or OH.

Alternative splicing of the human *ASPH* gene gives rise to multiple AspH isoforms and four functionally distinct proteins: AspH, humbug, junctate and junctin.[144,145] *ASPH* is under the control of at least two promoters, and AspH and humbug are expressed in tissues under the control of the P1 promoter, while junctate and junction are mostly expressed in tissues under the control of the P2 promoter.[146] Humbug is involved in Ca(II) regulation and overexpression of its gene is associated with tumour development.[147–149] Junctate is a membrane-bound Ca(II) binding protein which controls intracellular Ca(II) concentration through transmembrane channels.[150,151] Junctin modulates Ca(II) ion release and entry from the endoplasmic reticulum.[152]

1.3.4.1 Epidermal growth factor-like domains

Human epidermal growth factor (EGF) is a 53 amino-acid protein that regulates the proliferation of epidermal tissues and thus has a significant role in the body.[153–156] EGF-like domains (EGFDs) are conserved domains that are homologous to EGF.[157,158] EGFDs were first characterised in the EGF precursor (pro-EGF), a membrane protein that contains an extracelluar domain with eight EGFDs in addition to EGF itself.[159,160] EGF is proteolytically cleaved from pro-EGF, possibly in a Ca(II)-dependent manner.[161]

EGFDs are components in a variety of proteins including growth factors, extracellular matrix proteins and proteins involved in blood coagulation. Membrane-bound proteins that contain EGFDs include Notch, Jagged, Delta-like and thrombomodulin. EGFDs are also found in secreted proteins such as coagulation factors (hF), vitamin K dependent proteins, fibrillins and uromodulin. EGFDs often occur in repeating units, as observed in Notch and fibrillin, which contain 36 and 47 consecutive EGFDs respectively.[162,163] EGFDs have a range of roles including in facilitating interactions between proteins or structural roles. EGFDs in urokinase and hFXVIII are proposed to mediate binding of the proteins to their receptors.[164–166] Other EGFDs may function as spacer domains; in Notch, EGFDs are postulated to maintain the precise spacing between the Notch ligand-containing domain and the cell membrane to facilitate the correct interactions with a receptor on another cell surface.[162] Mutations in genes encoding for EGFDs are associated with genetic diseases, such as Marfan syndrome and hemophilia B.[167,168]

Both EGF and EGFDs are characterised by six cysteine residues that form a conserved pattern of disulfide bonds, as observed in multiple crystal structures. EGF contains a 1-3, 2-4, 5-6 connectivity pattern of cysteines, which is termed the “canonical” fold and considered to be essential for biological activity (Figure 1-8b and Figure 1-8c).[169] The canonical disulfide pattern is reported to be present in most EGFDs including EGFD-4 of thrombomodulin, EGFDs in hFIX and hFX and EGFDs in transforming growth factor. The non-canonical disulfide pattern, that is a 1-2, 3-4, 5-6 connectivity of disulfides, has also been observed in some EGFD, and has been proposed to be of wider

relevance than has been generally perceived (Figure 1-8a). NMR and crystallographic studies established that the disulfide connectivity in EGFD-5 of thrombomodulin is the non-canonical 1-2, 3-4, 5-6 pattern.

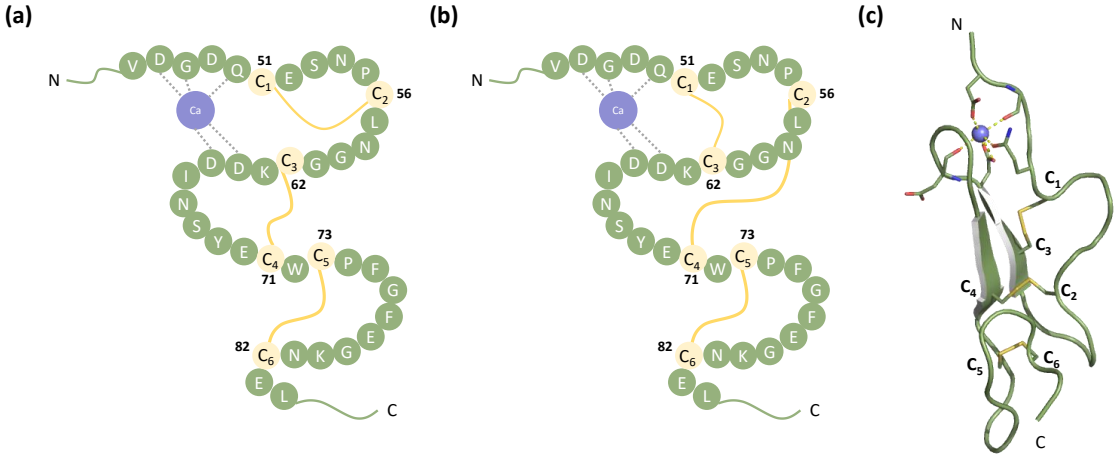


Figure 1-8. View from a crystal structure of an EGFD in human coagulation factor IX. (a) Schematic representation of the structure of one EGFD human coagulation factor IX with a non-canonical 1-2, 3-4, 5-6 connectivity of cysteine disulfide bonds. (b) Schematic representation of the canonical pattern of connectivity of disulfide bonds between cysteine residues (1-3, 2-4, 5-6). (c) View of a crystal structure of $hFIX_{46-84}$ which contains the canonical disulfide pattern (PDB 1IXA).[170] C = C terminus, N = N terminus, C_n = cysteine residue involved in disulfide bonding.

Some human EGFDs are also reported to contain a consensus sequence that is responsible for Ca(II) binding, that is (D/N)X(D/X)(Q/E/D)C_1…C_3X(D/N)XXXX(Y/F)XC_4.[169] EGFDs adopt a fold which is constructed from two adjacent β-sheets, and this structure is reported to be rigidified by Ca(II) binding.[171] In fibrillin EGFDs, it was reported that Ca(II) binding extends the overall length of the subdomain and may assist in protecting fibrillin from proteolysis.[172] Additionally, Ca(II) binding may be required for biological function of EGFDs. Point mutations that disrupt Ca(II) binding in EGFDs in hFX, hFIX and protein C led to a reduction in biological activity of hFX, hFIX and protein C. Correct folding of EGFDs, that is folding that results in the canonical disulfide fold (1-3, 2-4, 5-6), is postulated to be a prerequisite for Ca(II) binding.[173]

EGFDs undergo post-translational modifications which can affect the processing and functional properties of EGFD containing proteins.[174–178] In 1983 McMullen et al. first identified the post-translational hydroxylation of an aspartate in hF X.[138] Further hydroxylated aspartate residues

were consequently discovered in other human proteins, each located within a consensus sequence.[139,140] Hydroxylation of asparatate and asparaginyl residues was later shown to be catalysed by AspH.[123] Indeed, hydroxylated Asp and Asn residues in eukaryotes were first observed in EGFDs.[169]

The consensus sequence for AspH hydroxylation is **C_3**X(**D/N**)XXXX(Y/F)X**C_4**, where D/N denotes the hydroxylated Asp or Asn residue and C_n is a cysteine residue which forms part of the disulfide bonding pattern in EGFDs.[179] The degree of hydroxylation observed in human proteins containing EGFDs is reported to vary between AspH substrates. Purification of native hFIX revealed only 30% contained β-hydroxylation, whereas 100% of hFX EGFDs showed β-hydroxylation.[180] By contrast, purified native hFVIII does not manifest any β-hydroxylation, despite possessing the AspH consensus sequence.

The biological role of AspH-catalysed hydroxylation is unclear. Hydroxylated Asp/Asn residues are not a prerequisite for Ca(II) binding by EGFDs nor EGFD-containing protein activity.[180–182] Notably, in vitro studies using an $AspH_{315-758}$ construct determined that AspH only accepts EGFD substrates with a non-canonical disulfide pattern, that is 1-2, 3-4, 5-6.[124] It was therefore proposed that AspH could be responsible for recognising and labelling mis-folded EGFDs.[124]

AspH overexpression leads to accumulation of Notch, and activation of targets downstream of Notch, such as Hair and Enhancer of Split 1.[183,184] Overexpression of AspH is also reported in many human cancers and AspH-knockout mouse exhibit developmental defects.[185,186] Although the precise reason for AspH-catalysed hydroxylation is unknown, these results clearly indicate that the hydroxylation of specific Asp/Asn residues is essential for maintaining the normal functioning of biological processes.

1.3.4.2 The clinical relevance of AspH

AspH has attracted attention from a biomedical research perspective due to its implication in human disease. Overexpression of AspH has been recorded in 70-90% of human solid tumors

including hepatocellular carcinoma, chaloangiocarcinoma, pancreatic cancer, colorectal carcinoma, breast cancer, glioblastoma, prostate carcinoma, neuroblastoma, gastric cancer, ovarian cancer and colon cancer.[185,187–192] In hepatocellular carcinoma AspH overexpression is reported to correlate with cell motility and tumour invasiveness.[147,184,193,194] Overexpressed AspH is translocated to the cell surface, leaving the C-terminal region/oxygense domain exposed and thus AspH could be considered as a tumour biomarker or future therapeutic target. ASPH expression is reported to negatively correlate with breast cancer mortality.[195]

Early stage therapeutic treatments which target ASPH have been developed and tested in a range of model systems. Small molecules which inhibit AspH activity are reported to supress tumour growth in *in vivo* models, however issues with selectivity of AspH inhibition have been encountered.[196–198] Other therapies have used ASPH-specific antibodies, small interfering RNA and vaccinations to stimulate an immune response against ASPH.[199–205]

A rare genetic disorder has been attributed to pathogenic mutations in the *ASPH* gene. The disease, termed facial-dysmorphism, lens dislocation syndrome (FDLAB), or Traboulsi syndrome, was first identified in a family in 1995.[206] Affected individuals present with distinct ocular defects and facial dysmorphisms, and may exhibit other abnormalities in the cardiovascular and ocular systems.[111,207–209] Whole-exome sequencing confirmed FDLAB syndrome is a distinct disorder with affected *ASPH* gene.[210] Mutations were predominantly found in the region *ASPH* encoding the oxygenase domain and are thus postulated to impair AspH activity.[208,209,211,212] However, no biochemical investigation into the effect of the pathogenic mutations has currently been carried out.

1.3.4.3 AspH structure

Full-length AspH is composed of a cytoplasmic domain linked to a lumenal domain by a single pass type II transmembrane domain (Figure 1-9). The lumenal domain comprises an EF hand Ca(II) binding region, a tetratricopeptide repeat (TPR) and an oxygenase domain with a DSBH fold. The

active site is situated at the open end of the DBSH of the oxygenase domain. The TPR and oxygenase domains are connected by a flexible hinge loop region which is reported to be important in substrate recognition and binding.[124] In vitro assays determined that the TPR domain is essential for catalysis and the induced-fit mechanism of substrate binding requires cooperation from the oxygenase and TPR domains, in order to position the substrate in a productive manner at the active site.[124]

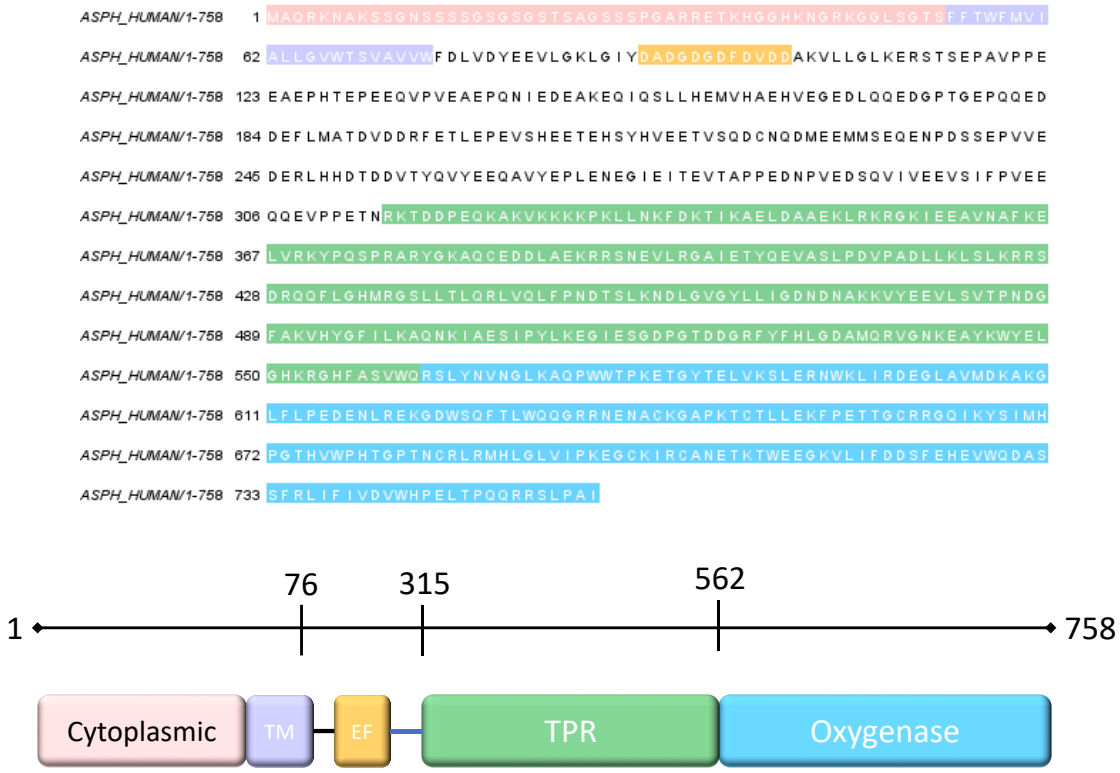


Figure 1-9. The full length human AspH amino acid sequence with major features annotated. Graphical representation of positon of main features in sequence with corresponding colour coding. Cytoplasmic = cytoplasmic domain, TM = single pass transmembrane domain, EF = EF-hand motif, TPR = Tetracotripeptide repeat domain, Oxygenase = Oxygenase domain.

Biochemical characterization of human AspH was previously carried out by Pfeffer et al., who established a construct that could be used for in vitro assays and structural studies.[124] This construct comprises the oxygenase and TPR domains, that is $AspH_{315-758}$ (Figure 1-10a). Full length AspH has not been structurally characterised.

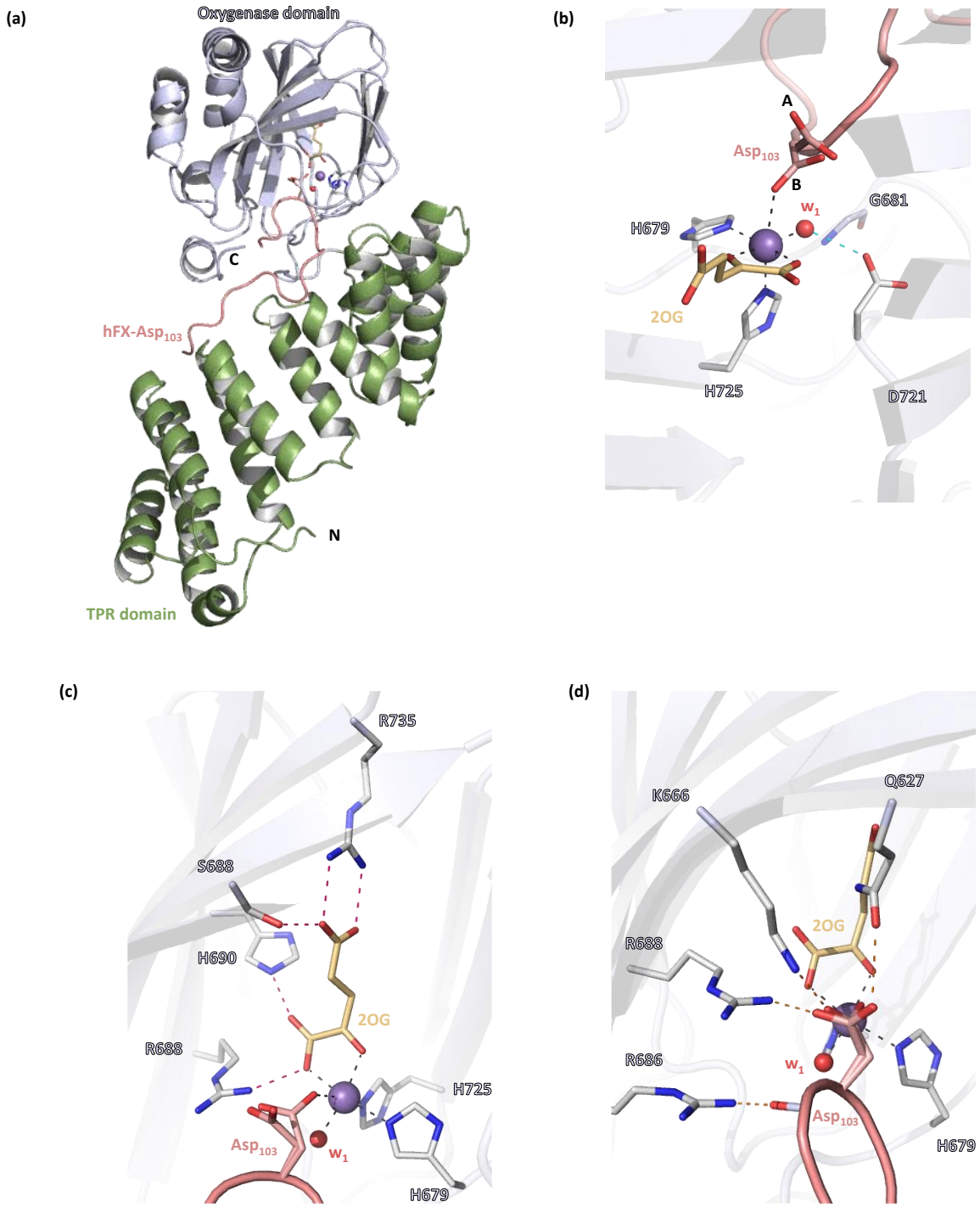


Figure 1-10. View from a crystal structure of the AspH oxygenase and TPR domains in complex with Mn, 2OG and a synthetic peptide based on the consensus sequence for hFX (hFX-Asp$_{103}$). (a) View from a crystal structure of the overall fold of the AspH oxygenase and TPR domains. (b) View of metal coordination at the AspH active site. H679 and H725 directly coordinate Mn. Asp$_{103}$ of hFX-Asp$_{103}$ is modelled as two conformations labelled A and B. (c) View of active site residues involved in 2OG binding (pink dashes) (d) View of active site substrate binding residues (orange dashes).

In comparison to most 2OG-dependent hydroxylases which possess a conserved triad of Fe(II) binding ligands (HXD/E…H), the active site of AspH employs a dyad of histidine residues, HXG…H, to ligate Fe(II) (Figure 1-10b). The HXG…H motif is also found in 2OG-dependent halogenases; however, no evidence for AspH catalysed halogenation has been reported. In human AspH, H679 ligates Fe(II) opposite the 2OG C1 keto oxygen group, while H725 ligates Fe(II) trans to the proposed oxygen binding site. Sequence alignments show that G681 is located in the position where a third acidic residue may normally be found in 2OG-hydroxylases. In the AspH crystal structure a water molecule, w_1, occupies the vacant coordination site created by G681. W_1 is positioned to interact with D721.

2OG is held at the AspH active site by interactions with secondary sphere residue side chains; H690 and R688 interact with the C1 carboxylate, while the 2OG C5 carboxylate is positioned to form a salt bridge to R735 and an additional hydrogen bond to S668 (Figure 1-10c). Substrate binding in $AspH_{315\text{-}758}$ has been studied using a synthetic peptide based on the consensus sequence for hFX (termed hFX-Asp_{103}). This allowed for identification of key residues involved in substrate carboxylate stabilisation at the active site; Q627 and K666 are positioned to interact with the D103 carboxylate side chain while R686 is positioned to interact with the K102 main chain (Figure 1-10d). Interestingly, some crystal structures of AspH revealed that the substrate Asp_{103} can be observed in two conformations; conformation A in which the Asp_{103} carboxylate is directed away from the metal centre, and conformation B in which the Asp_{103} carboxylate is coordinating to Mn at the active site (labelled in Figure 1-10a).

The active site of AspH has proved to be remarkably resilient to substitutions; substitution of key residues for alanine revealed that the loss of only two residues was completely inactivating (Table 1-3).[213] The AspH variants possessing either the R735A or D721A substitutions were inactive, with only background levels of substrate oxidation recorded. R735 coordinates to the 2OG C5

carboxylate at the active site via a salt bridge and D721 is involved in substrate coordination and secondary sphere interactions to other stabilising residues, including R686.

Initial characterisation of bovine AspH suggested that both of the histidine Fe(II)-binding residues are essential for catalysis.[143,214] Substitution of one active site histidine to produce the H675A variant (equivalent to H679A in human AspH) rendered AspH inactive.[214,215] The inactivity was attributed to the fact that the H675A variant could no longer bind either Fe(II) or 2OG, although the methodology used meant it was not possible to distinguish exactly the extent to which Fe(II) and 2OG binding cofactor had been most affected by the mutation.[143,214]

Table 1-3. Reported activity of AspH active site and secondary sphere variants. The experiments were performed by Dr Inga Pfeffer (unpublished results), and incubated with AspH variant for 60 minutes with L-ascorbic acid (LAA), 2OG and ferrous ammonium iron sulphate (FAS) with the hFX-Asp_{103} substrate. Data are the mean and standard deviation of triplicate repeats (n = 3, ± SD).

Mutation	Targeted residue function	Turnover % [213]
Wild-type	-	85
Control	-	4
H679A	Fe(II) binding	84
H725A	Fe(II) binding	81
H679W	Fe(II) binding	1
R735A	2OG-binding	4
R688A	2OG-binding	84
H690A	2OG-binding	60
R686A	Substrate coordination	56
D721A	Substrate coordination and R686 stabilisation	5
K666A	Substrate coordination	87
Q627A	Substrate coordination	85

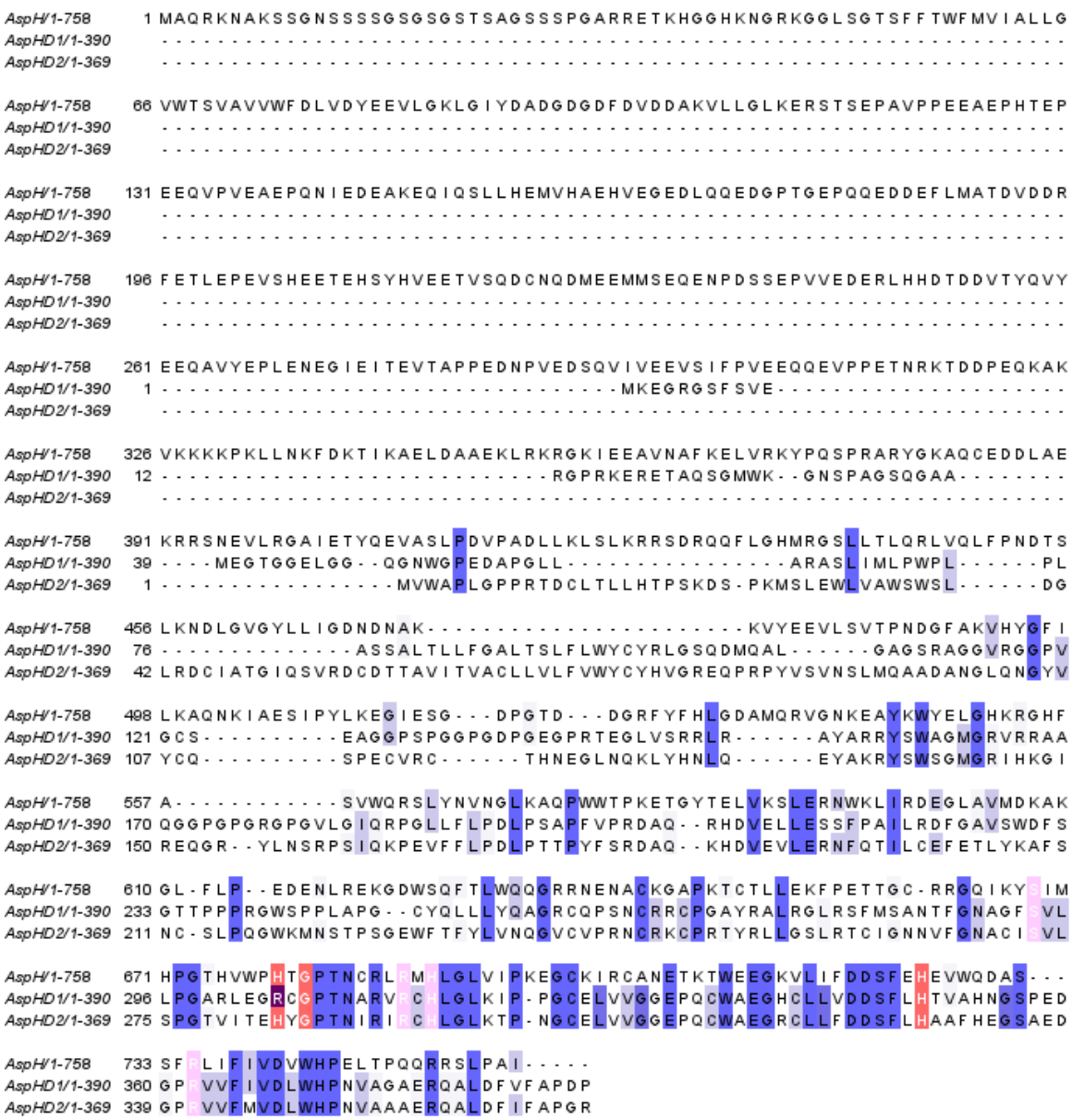
Somewhat surprisingly, more recent evidence from in vitro studies has disputed the inactivating nature of the active site His to Ala mutations in AspH. There is now evidence that AspH is capable

of catalysing hydroxylation with only a single Fe(II) binding ligand at the active site.[213,216] A 2018 study of the role of human AspH in cancer used the H679A mutation (equivalent to the H675A mutation in bovine AspH) to investigate the effect of impaired hydroxylation on tumour growth.[216] The H679A mutation reduced the activity of AspH by 76 %; however, the His to Ala substitution was not observed to be completely inactivating as observed by in cell assays.[217]

Assays using the $AspH_{315-758}$ construct and the synthetic peptide substrate mimic hFX-Asp_{103} further supported these observations.[213] Initial activity assays revealed that the H679A and H725A AspH variants could achieve comparable levels of turnover compared to wild-type AspH, albeit at a reduced rate of reaction.[213] This finding suggests that AspH is capable of catalysing hydroxylation with a single Fe(II)-binding ligand; however, no further characterisation of the AspH active site variants was performed. Needless to say, this discovery warrants further investigation and may have implications for the assignment of the 2OG-dependent oxygenase superfamily.

1.3.4.4 AspH homologues: Aspartate-β-hydroxylase domain-containing proteins

The human genes ASPHD1 and ASPHD2 encode for two putative proteins, aspartate β-hydroxylase domain-containing proteins 1 and 2 (AspHD1 and AspHD2), which are predicted to be 2OG-dependent oxygenases. AspHD1 and AspHD2 share ~36% of their sequence with AspH and ~56% sequence similarity to one another (Figure 1-11). Analysis of the AspHD1 and AspHD2 sequences predicts both proteins to be composed of a small cytoplasmic domain connected to a larger lumenal domain by a single-pass type II transmembrane domain (Uniprot Q5U4P2 and Q6ICH7).


AspH/1-758 1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMVIALLG
AspHD1/1-390 - - -

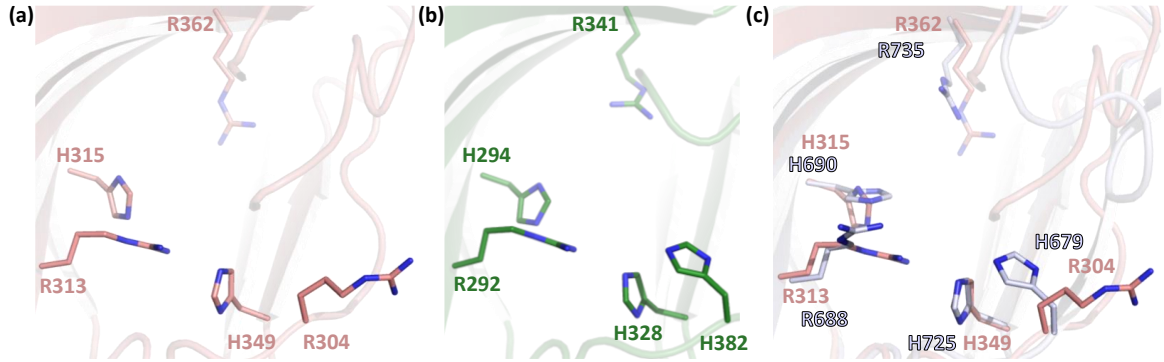


Figure 1-12. Computationally predicted active site structures of AspHD1 (a), AspHD2 (b) and an overlay of the active site of AspH (PDB 6YYU) with the predicted AspHD1 active site. Positioning of predicted metal and 2OG binding residues shown. The structures were predicted using AlphaFold.

Interestingly, comparison of the amino acid sequences of AspH and AspHD1 predicts a RXG...H Fe(II) binding motif in AspHD1. R304 in AspHD1 is predicted to be positioned where H679 is located at the AspH active site and R304 may therefore be expected to have a similar role to H679 in AspH, that is in Fe(II) ligation at the AspHD1 active site (Figure 1-12a). To the best of my knowledge, this type of histidine to arginine substitution in the metal binding triad has not been observed in other members of the 2OG-dependent oxygenase superfamily. The computationally generated AspHD1 structure suggests that the R304 side chain faces outside of the active site (Figure 1-12a), away from the most likely position for a metal ion to occupy. By contrast to AspHD1, AspHD2 is a predicted to have the same histidine dyad binding motif as AspH, that is the HXG…H dyad (Figure 1-12b). The exact function and substrates of AspHD1 and AspHD2 are not known; however, AspHD2 could be an example of another 2OG-dependent hydroxylase with an HXG…H dyad.

Pathogenic mutations in the ASPHD1 gene have been linked to schizophrenia 3 and chromosome 16P11.2 Deletion Syndrome.[219] Recent work has connected expression of *ASPHD1* to patient prognosis in skin cutaneous melanoma (SKCM) and suggested that AspHD1 is a potential biomarker for SKCM and could be future target of therapies.[220] *ASPHD2* overexpression has been reported in glioblastoma tumour stem cell microenvironments.[221] Further research into this area may lead to the assignment of AspHD2 as a tumour biomarker or therapeutic target.

At present, biochemical characterisation of either AspHD1 or AspHD2 has not been reported. It is likely that biomedical research will continue to inform on the significance of AspHD1 and AspHD2 and therefore biochemical characterisation of the proteins will also be required. Additionally, characterisation of the unusual active site Fe(II)-binding motifs predicted for AspHD1 and AspHD2 may have implications for the assignment of 2OG-dependent oxygenases.

1.4 Aims of the work described in this thesis

The work described in this thesis aimed to address outstanding questions surrounding the mechanism of catalysis by AspH and the significance of unconvential Fe(II)-ligating dyad at the AspH active site. **Chapter 2** describes work which aimed to study the Fe(II)-binding HXG…H dyad through the use of site-directed mutagenesis studies. The AspH variants H679A and H725A in which one of the dyad histidine residues is substituted for an alanine were produced recombinantly and subjected to detailed kinetic and structural studies. **Chapter 3** describes investigations that were performed to investigate clinically relevant AspH variants that affect key 2OG-binding residues. The work described in **Chapter 4** was designed with the long-term goal of developing a system for studying the AspH mechanism by means of time-resolved femtosecond serial crystallography experiments. In **Chapter 5,** the work described aimed to investigate the nature of Fe(II)-binding in 2OG-dependent oxygenases by establishing a system for the recombinant production and purification of the AspH homologue, AspHD1. The work also aimed to investigate the role of arginine as a metal-ligating residue in a model system.

Chapter 2 | Investigation into the mechanism of AspH catalysis using Fe(II)-binding variants

2.1 Introduction

It is widely perceived that a characteristic structural feature of the 2-oxoglutarate (2OG)-dependent oxygenase family is the triad of Fe(II)-ligating residues at the active site.[55] Crystallographic studies and analysis of sequence alignments of 2OG-dependent oxygenases have revealed a near universally conserved HXD/E…H motif located in the oxygenase domain.[222] During catalysis, a six coordinate environment around the Fe(II) cofactor is therefore constructed from three protein residues, bidentate coordination from 2OG and coordination from one oxygen of a dioxygen molecule.[223] By contrast, the 2OG-dependent hydroxylase AspH has an unusual active site Fe(II)-binding motif, that is an HXG…H dyad (Figure 2-1).[214]

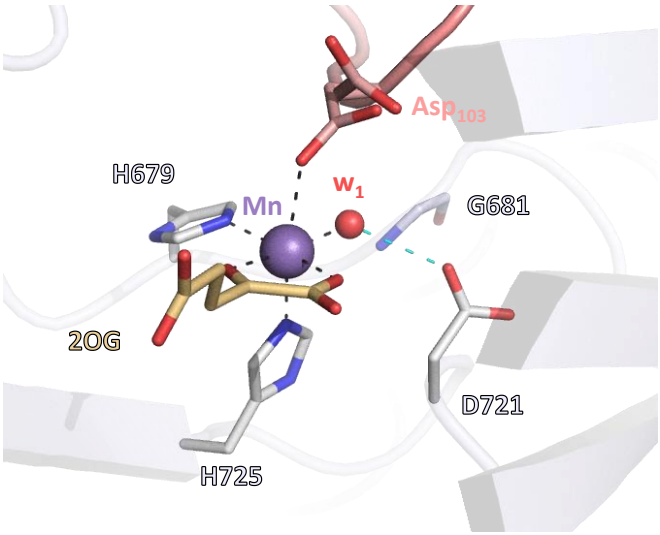


Figure 2-1. View from the active site of AspH in complex with Mn, 2OG and a synthetic substrate mimic (hFX-Asp$_{103}$). H725 and H679 are directly involved in ligating Mn. D721 interacts with w_1 (cyan dashes), which is coordinated to Mn. G681 is not directly involved in metal coordination. Asp$_{103}$ of the substrate is shown in two conformations; one in which the carboxylate coordinates to Mn (B), and one in which the carboxylate is directed away from Mn (A).

Sequence alignments show a glycine, G681, is located where a third Fe(II) binding residue would normally be predicted to be found.[143] Structural studies confirmed that only two histidine residues, H679 and H725, are involved in Fe(II)-ligation at the AspH active site.[124] The significance of the unusual metal-binding pattern for a 2OG-dependent hydroxylase, and, any effect this may have on the catalytic mechanism in AspH, is currently unknown. The HXG…H dyad is employed by the 2OG-dependent halogenases, in which the vacant coordination site created by the E/D to G substitution

is occupied by a halogen ion during catalysis.[111] However, no evidence for AspH-catalysed halogenation has been reported.

Although initial activity assays using bovine AspH indicated that both Fe(II)-ligating active site histidine residues were essential for AspH activity, more recent results have disputed these findings.[213,214] *In vitro* assays showed that substitution of either histidine residue for alanine to afford the H679A and H725A AspH variants resulted in two AspH variants that surprisingly retained substantial levels of catalytic activity (Table 2-1). Additionally, restoration of the more typical 2OG-dependent hydroxylase binding triad by substitution of G681 to either E or D produced catalytically active AspH variants.[213] In these mutagenesis studies, the only active site residue in close proximity to the Fe(II) centre that was determined to be essential for catalysis was D721, as substitution of D721 for alanine eliminated the catalytic activity of AspH.

Table 2-1. Reported activity of AspH active site and secondary sphere variants. The experiments were performed by Dr Inga Pfeffer (unpublished results), and incubated with AspH variant for 60 minutes with L-ascorbic acid (LAA), 2OG and ferrous ammonium iron sulphate (FAS) with the hFX-Asp_{103} substrate.[213] Data are the mean and standard deviation of triplicate repeats (n = 3, ± SD).

Mutation	Targeted residue function	Turnover % [213]
Wild-type	-	85.3 ± 0.5
Control	-	2.8 ± 0.5
H679A	Fe(II) binding	84.1 ± 1.2
H725A	Fe(II) binding	81.1 ± 0.4
G681E	HX**G**..H of triad	85.1 ± 0.6
G681D	HX**G**..H of triad	85.1 ± 0.9
D721A	Coordinates w_1	5.2 ± 0.3

2.2 Aims of the work outlined in this chapter

The finding that AspH can catalyse hydroxylation with a single active site Fe(II)-ligating residue warrants further investigation. Thus, the work described in this chapter aimed to examine the nature of Fe(II)-binding at the AspH active site in detail by subjecting the H679A and H725A variants to mechanistic and structural investigations. The work described herein aimed to (i) develop a method for the production and purification of AspH variants, (ii) investigate the activity of the H679A and H725A AspH variants by the determination of kinetic parameters and inhibition studies, (iii) explore the structural basis for the activity of the H679A and H725A variants using crystallographic studies, (iv) investigate the mechanism of oxygen incorporation in the AspH mechanism by isotopic labelling experiments and (v) evaluate the significance of the conventional catalytic triad using the G681E and G681D AspH variants.

2.3 Results and Discussion

2.3.1 Production and purification of recombinant AspH variants

A method for the production and purification of an AspH construct including the oxygenase and TPR domains, that is $AspH_{315\text{-}758}$, was previously developed in the Schofield group.[124] The $AspH_{315\text{-}758}$ construct will henceforth be referred to as "wildtype AspH" (wt AspH). The plasmids for the H679A and H725A AspH variants were subsequently generated by Dr Inga Pfeffer by site-directed mutagenesis of the pet28a(+) His_6-$AspH_{315\text{-}758}$ plasmid and therefore comprise the same domains as wt AspH. A plasmid encoding an AspH variant in which both H679 and H725 were substituted for alanine was also later produced. This variant will henceforth be referred to as the double histidine variant.

For the production of AspH variants, the plasmids encoding the H679A, H725A and double histidine variants were transformed into BL21 DE3 competent cells and a starter culture was grown from a single colony at 37°C overnight. The starter cultures were diluted 1:100 in fresh 2xYT media and 600 mL cultures were grown at 37°C with shaking until OD_{600} 1-1.0 was reached. Expression of the

ASPH gene was induced by addition of IPTG to a final concentration of 0.1 mM and the cultures were incubated for 16-18 hours at 18°C. The cultures were pelleted by centrifugation (11867 x *g*, 8 min, 4°C).

The composition of buffers used for the purification of AspH variants is outlined in Table 2-2. Cell pellets were resuspended in buffer A (100 mL per 30 g of wet cell pellet) with protease inhibitor tablets (Roche, 1 tablet per 50 g of cell pellet) and DNAse I (Roche Diagnostics). Cells were lysed by sonication (8x30 s bursts) or high-pressure cell disruption at 25 kPsi. The lysates were centrifuged (30 min, 58387 x *g*, 4°C) and the supernatant was filtered through 0.45 µm filters. The clarified lysates were loaded onto a 5 mL His-Trap column (Cytivia), which had been pre-equilibrated with buffer A. The column was washed with buffer A, a mixture of 5%$_{v/v}$ buffer B and 95%$_{v/v}$ buffer A, and then His$_6$-tagged AspH was eluted by a linear gradient from 5%$_{v/v}$ B to 100%$_{v/v}$ B. Eluted fractions were pooled and then either immediately further purified by size-exclusion chromatography (SEC) using a Superdex 75 300 mL size exclusion column, or the His$_6$-tag was removed by the addition of thrombin protease. His$_6$-cleaved protein was then purified by a reverse HisTrap and subsequently purified by size exclusion chromatography (Superdex 75, 300 mL). The masses of the purified AspH variants were confirmed by mass-spectrometry (Figure 2-2).

Table 2-2. The composition of buffers used for the purification of wt AspH and AspH variants

Buffer	Composition	pH
Buffer A	50 mM HEPES 500 mM NaCl 5 mM imidazole	7.5
Buffer B	50 mM HEPES 500 mM NaCl 500 mM imidazole	7.5
Size exclusion buffer	50 mM HEPES 150 mM NaCl	7.5

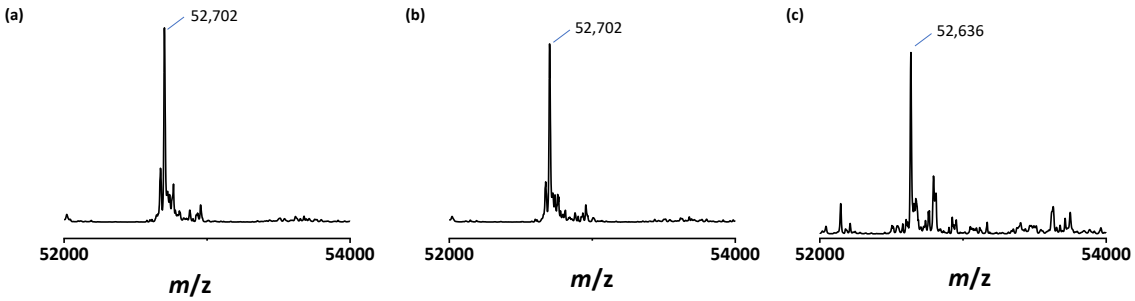


Figure 2-2. Mass spectra of the purified H679A and H725A AspH variants. (a) The deconvoluted mass spectrum of the H679A variant. (b) The deconvoluted mass spectrum of the H725A variant. Calculated masses: 52,702 Da for both variants. (c) The deconvoluted mass spectrum of the double histidine (H679A + H725A) variant. Calculated mass: 52,636 Da.

2.3.2 Development of AspH activity assays

In order to assess AspH activity *in vitro*, a synthetic substrate based on the consensus AspH hydroxylation sequence was previously developed by the Schofield group.[124] Initial characterisation of AspH activity determined that AspH only catalyses the hydroxylation of epidermal growth factor-like domains (EGFDs) with a non-canonical pattern of disulphide bonds.[124] To fulfil this requirement, a peptide based on the sequence of human coagulation factor X (hFX) was designed, with cysteine residues 1, 2, 5 and 6 substituted for serine residues (Figure 2-3). The substrates, originally termed "NC-4Ser-EGFD$_{86\text{-}124}$" based on the numbering of residues in the hFX sequence, will henceforth be referred to as hFX-Asp$_{103}$.[124] A D103N variant of hFX-Asp$_{103}$, that is hFX-Asn$_{103}$, was also employed in the studies. The hFX-Asp$_{103}$ and hFX-Asn$_{103}$ substrates were used in all activity assays and crystallography experiments described in this thesis. The hFX-Asn$_{103}$ cyclic peptide serves as a model asparagine-based AspH substrate as the asparagine residue is not part of the wt hFX sequence. The majority of work was performed on the hFX-Asp$_{103}$ substrate due to the cost of purchasing the substrates and difficulties with supply during the COVID-19 pandemic.

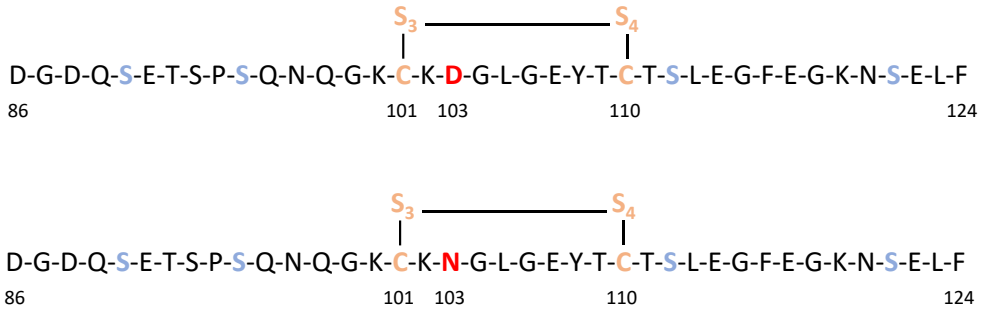


Figure 2-3. The synthetic AspH substrates hFX-Asp$_{103}$ and hFX-Asn$_{103}$. The sites of hydroxylation, Asp$_{103}$ (a) or Asn$_{103}$ (b), are shown in red. A disulfide bond between Cys$_{101}$ and Cys$_{110}$ (orange) forms a 10-membered macrocycle that is part of the recognition sequence for AspH catalyzed hydroxylation (Cys$_3$-Cys$_4$). Sites in the sequence where Cys$_1$-Cys$_2$ and Cys$_5$-Cys$_6$ disulfide bonds are normally located have been substituted for serine residues (blue) to avoid shuffling of the disulfide pattern in solution.[124]

Using a model substrate, the activity of wt AspH and all AspH variants can be monitored by mass-spectrometric analysis of the product; the hydroxylated product has mass of +16 Da relative to the substrate. Thus, the turnover can be calculated by integrating the substrate and product mass peaks, and calculating the proportion of product relative to the total amount of substrate and product present in the assays.

Activity assays for 2OG-dependent oxygenases typically contain 2OG, Fe(II), enzyme, substrate and L-ascorbic acid (LAA). The role of LAA in 2OG-dependent oxygenase catalysis is not fully understood; however, with some 2OG oxygenases LAA is reported to improve the robustness of assays.[224–227] The activity of wt AspH was previously reported to be independent of LAA; however, LAA was added to activity assays to improve assay stability, and maintain consistency with previously reported results.[228] Activity assays described herein are therefore are composed of LAA, 2OG, ferrous ammonium iron sulphate (FAS), hFX-Asp$_{103}$ or hFX-Asn$_{103}$ substrate and AspH, unless stated otherwise.

2.3.3 Initial assessment of the activity of the H679A and H725A AspH variants

H679A, H725A variant-catalysed and wt AspH-catalysed substrate hydroxylation was monitored in real-time via injecting assay aliquots into a solid phase extraction mass-spectrometry (SPE-MS) system coupled to a RapidFire sample handler as described previously.[229] The initial results

confirmed the previously reported findings that the H679A and H725A variants catalysed the hydroxylation of the hFX-Asp_{103} substrate.[124]

The assay buffer conditions were subsequently optimised for the H679A and H725A variants (Figure 2-4). The activity of the H679A and H725A variants decreased with increasing NaCl concentration present in the buffer. By contrast, wt AspH was substantially more active in the presence of NaCl. The same result was observed when using either the hFX-Asp_{103} or hFX-Asn_{103} substrate (Figure 2-4a and b compared to Figure 2-4c and d respectively). The optimised reaction conditions and concentrations of cofactors, cosubstrates and enzymes are given in Table 2-3.

The double histidine to alanine AspH variant was catalytically inactive under all conditions that were tested. These data provide confirmation that a single protein Fe(II)-ligand is sufficient for productive catalysis by AspH. The double histidine to alanine variant can be used as a control system in cellular studies, and the single histidine to alanine variants are not in fact appropriate controls, as previously proposed.[143]

Table 2-3. The optimised buffer conditions for activity assays and determination of kinetic parameters of the wt AspH and the H679A and H725A variants.

Enzyme	Buffer composition	Concentration of reactant (μM)				
		LAA	2OG	Substrate	FAS	Enzyme
wt AspH	25 mM HEPES pH 7.5, 50 mM NaCl	100	20	4	20	0.1
H679A H725A	25 mM HEPES pH 7.5	100	2000	4	20	0.2

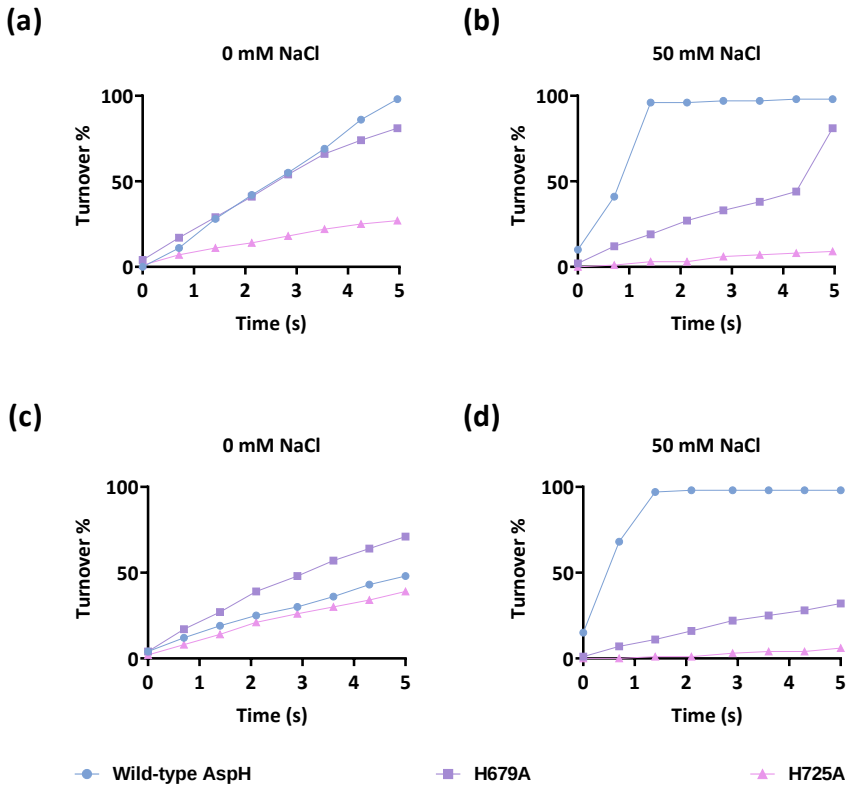


Figure 2-4. The effect of NaCl on the activity of wt AspH and the H679A and H725A AspH variants. The effect of buffer salt concentration on AspH activity was assessed in 25 mM HEPES pH 7.5 buffer. With no NaCl (hFX-Asp_{103} substrate (a) and hFX-Asn_{103} substrate (c)) the activity of the H679A variant is faster than wt AspH. The addition of 50 mM NaCl (hFX-Asp_{103} substrate (b) and hFX-Asn_{103} substrate (d)) dramatically improves activity of wt AspH. Assay conditions are summarized in Table 2-3.

2.3.3.1 Uncoupled turnover

It has previously been reported that some, but not all, 2OG-dependent oxygenases are capable of catalysing the conversion of 2OG to succinate independently from substrate oxidation.[230] Prior investigations determined that wt AspH does not catalyse the conversion of 2OG without substrate present.[228] In order to assess whether this AspH characteristic was affected by the active site His to Ala substitutions, the H679A and H725A AspH variants were subjected to NMR studies. Assays containing 2OG, LAA, FAS and either the H679A or H725A variant were monitored using a Bruker AVIII 700 NMR machine over 60 min (Figure 2-5).

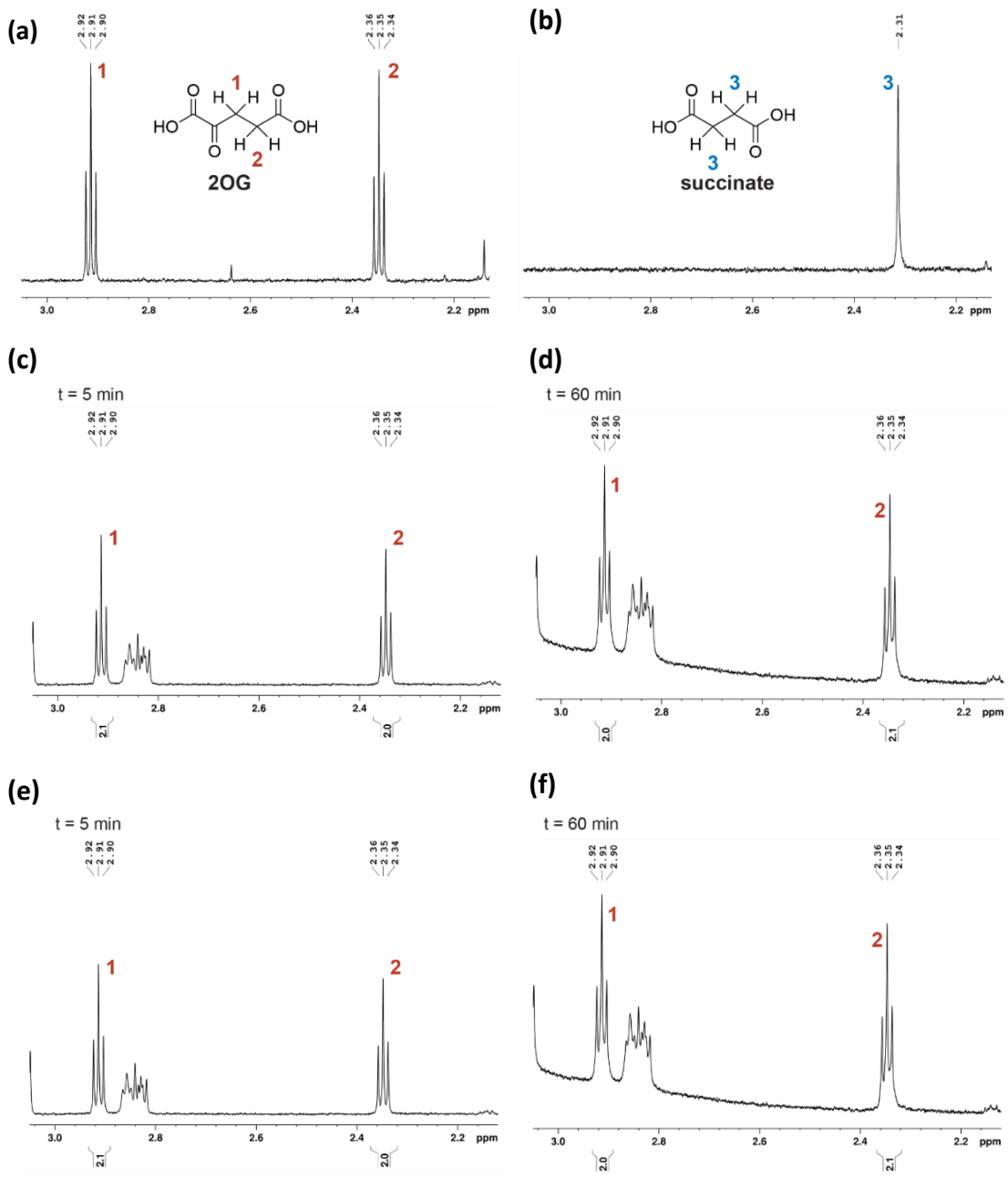


Figure 2-5. The H679A and H725A variants do not catalyse the oxidative decarboxylation of 2OG to succinate in the absence of substrate. NMR spectrum of (a) 2OG and (b) succinate standards. Reaction assays containing LAA, 2OG, FAS, substrate and enzyme were set up and monitored over the course of an hour. Panels (c) and (d) show a timecourse NMR experiment in which the H679A variant does not perform uncoupled turnover of 2OG. Panels (e) and (f) show a timecourse NMR experiment in which the H725A variant does not perform uncoupled turnover of 2OG. Figure adapted from Brasnett et al 2021.

The results showed that the H679A and H725A variants did not catalyse the oxidative decarboxylation of 2OG to succinate and CO_2 in the absence of substrate. This observation is consistent with the wt AspH results and therefore the active site His to Ala mutation has not affected the uncoupled 2OG turnover.

2.3.4 Determination of kinetic parameters of the H679A and H725A AspH variants

Determination of kinetic parameters in a multifaceted enzyme system, such as that of 2OG-dependent oxygenases, can be challenging. In order to apply Michaelis Menten kinetics, the system must be in a pseudo-first order state. For the study of 2OG-dependent oxygenases, this means all cofactors except the one to be determined must be in excess so as not to limit the reaction rate. Having optimised the buffer conditions for the H679A and H725A variants, the kinetic parameters with respect to Fe(II) and 2OG were determined for both of the variants and wt AspH. The data presented are an average of independent triplicate repeats and the kinetic parameters for wt AspH and the H679A and H725A variants are summarized in Table 2-4 and Michelis Menten plots are shown in Figure 2-6.

The K_m and k_{cat} values determined for wt AspH for Fe(II) and 2OG are comparable to reported data, which was obtained using a different cyclic peptide substrate which was also based on hFX, but contained a thioether-linked cyclic sequence.[228] The k_{cat} values for the H679A variant are approximately two thirds of the k_{cat} value for wt AspH. By contrast, the k_{cat} value for the H725A variant are considerably lower than the k_{cat} values for wt AspH, and about half the magnitude of the k_{cat} values for the H679A variant. This trend is in agreement with the levels of substrate hydroxylation observed in initial activity assays and during the buffer optimisation process (Figure 2-4).

It is interesting to note that the K_m values for H679A with respect to Fe(II), either in the presence or absence of LAA, are within the range of those of wt AspH and other reported 2OG oxygenases.[231–233] The addition of LAA to the reaction decreases the K_m with respect to Fe(II) for both the H679A variant and wt AspH, however, LAA is not required for effective catalysis (Table 2-4). By contrast to the kinetic parameters determined for Fe(II), the Km value with respect to 2OG for the H679A variant is two orders of magnitude higher than the wt AspH 2OG K_m value.

Table 2-4. Kinetic parameters determined for wt AspH and the H679A and H725A variants. Fe(II) was measured in the presence and absence of L-ascorbic acid. a) Measured in the presence of LAA. b) Measured in the absence of LAA. Reported wt AspH values measured using an alternative cyclic peptide substrate are shown for comparison.[228] Data are reported as the average of independent triplicate repeats ± standard deviation.

Cofactor	AspH variant	k_{cat}^{app} [s^{-1}]	K_m^{app} [μM]	k_{cat}/K_m [$mM^{-1}\,s^{-1}$]
Fe(II)[a]	wt	0.27 ± 0.04	4.3 ± 1.6	62.8 ± 25.2
	wt[228]	0.19 ± 0.03	1.42 ± 0.16	130 ± 0.03
	H679A	0.19 ± 0.01	2.7 ± 0.5	70.4 ± 13.6
	H725A	0.10 ± 0.01	8.1 ± 1.8	12.3 ± 3.0
Fe(II)[b]	wt	0.24 ± 0.04	7.2 ± 2.6	33.3 ± 13.2
	wt[228]	0.19 ± 0.03	4.76 ± 0.48	39 ± 0.01
	H679A	0.18 ± 0.01	4.4 ± 0.8	40.9 ± 7.8
	H725A	0.09 ± 0.01	12.3 ± 3.0	7.3 ± 2.0
2OG	wt	0.31 ± 0.03	1.1 ± 0.4	282 ± 106
	wt[228]	0.19 ±0.03	0.60 ± 0.09	320 ± 0.07
	H679A	0.18 ± 0.01	109 ± 20	1.7 ± 0.3
	H725A	0.08 ± 0.01	211 ± 65	0.38 ± 0.13

The H725A substitution had a more significant effect on the AspH kinetic profile than the H679A substitution. The K_m value for the H725A variant with respect to Fe(II) is around twofold and threefold greater than those for wt and the H679A AspH variant respectively (Table 2-4). The greatest difference is observed with respect to 2OG, as the H725A variant 2OG K_m is about 200 times larger than that of wt AspH and twice that of H679A variant.

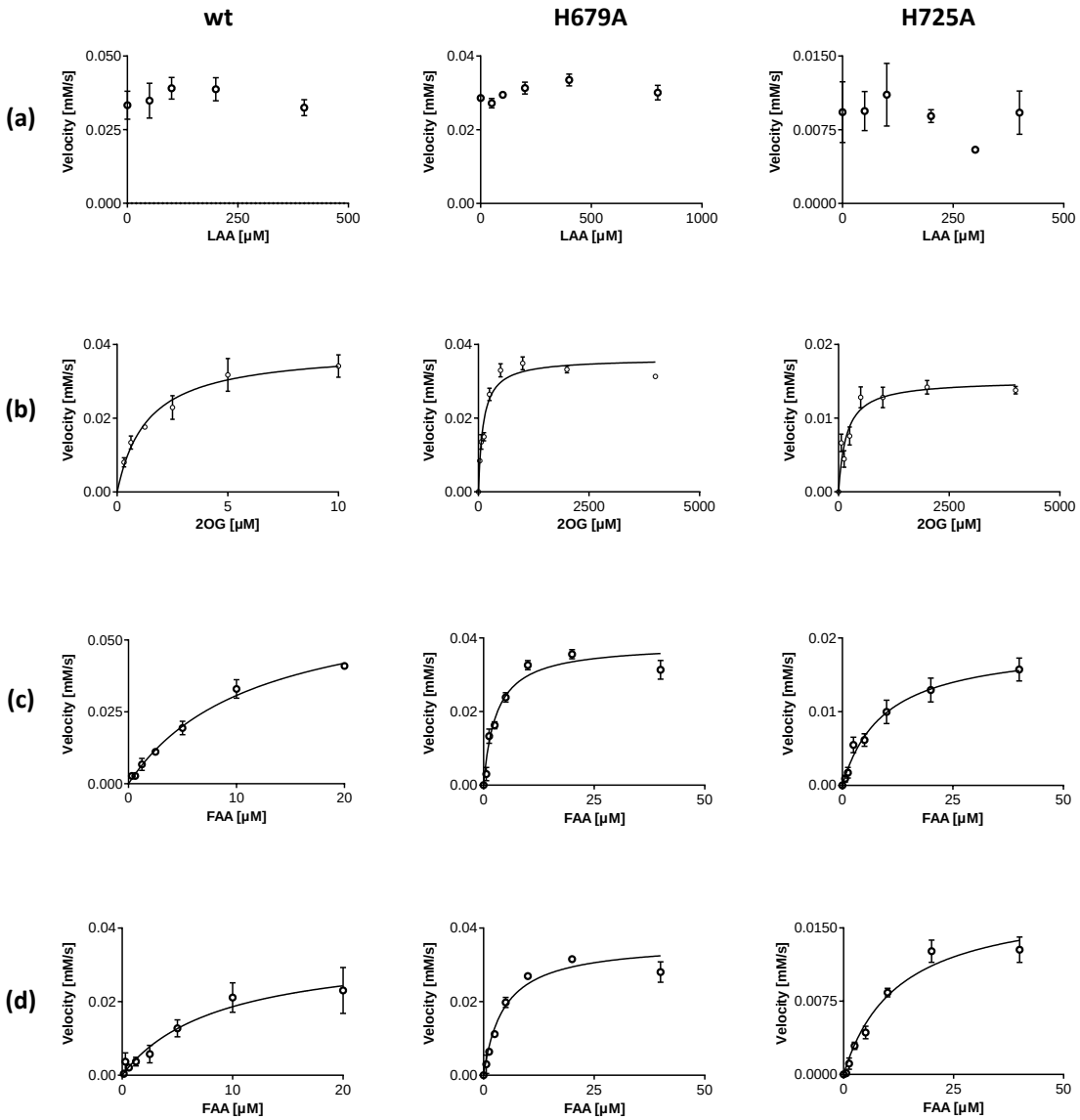


Figure 2-6. Michaelis Menten plots for the determination of the kinetic parameters for wt AspH and the H679A and H725A variants. Data are the average of three independent triplicate repeats ± the standard deviation. Assay conditions and cofactor concentrations are outlined in Table 2-3, except for the concentration of the given cofactor which was varied to determine the kinetic parameters with respect to that cofactor. (a) Data with respect to LAA, (b) Data with respect to 2OG, (c) data with respect to Fe(II) and (d) data with respect to Fe(II) in the presence of LAA.

Although the K_m values with respect to 2OG for the H679A and H725A variants are substantially higher than the 2OG K_m value for wt AspH and most reported human 2OG oxygenases, the K_m values are in the range of 2OG concentrations detected in cells and lower than 2OG K_m value for γ-butyrobetaine hydroxylase.[234] Additionally, the substantial differences in 2OG K_m values is not due to 2OG oxidative decarboxylation occurring in the H679A and H726A variants, as demonstrated by the NMR studies performed previously (Figure 2-5).

As evidenced by the high 2OG K_m values compared to wt AspH, the His to Ala substitutions in the AspH active site have most significantly affected 2OG binding. By contrast, the minimally affected Fe(II) K_m values for the H679A and H725A variants suggest that a single histidine Fe(II) ligating residue at the AspH active site is sufficient for catalysis, at least under the tested assay conditions.

Measurement of the H679A and H725A variant kinetic parameters with respect to hFX-Asp_{103} was attempted using the same SPE-MS method as used for the determination of kinetic parameters with respect to Fe(II) and 2OG; however, this proved to be challenging. At low hFX-Asp_{103} concentrations, substrate and product masses were not detected. High hFX-Asp_{103} concentrations blocked the equipment tubing and could therefore not be measured.

Optimisation of a method to measure the kinetics parameters of the H679A and H725A variants with respect to O_2 concentration was also attempted, as reported for wt AspH.[228] This method relied on manual quenching of assays which were subsequently analysed by SPE-MS, instead of real-time monitoring. Unfortunately, the reproducibility of the manual quenching method proved to be problematic, particularly at high O_2 concentrations where catalysis proceeded rapidly. Therefore, these data have been omitted and the methodology requires further optimisation.

2.3.5 Inhibition studies on the H679A and H725A variants

The effect of the His to Ala substitution at the AspH active site was further investigated by use of wt AspH inhibitors. The half-maximum inhibitory concentrations (IC_{50}) of reported 2OG oxygenase inhibitors including metal ions and small molecules, were determined for the H679A and H725A variants using SPE-MS assays (Table 2-5).

Table 2-5. IC_{50} values of metal ions and 2OG analogues for wt AspH and the H679A and H725A AspH variants. Data are the mean of two independent runs (*n* = 2; mean ± SD). Conditions for the assays are outlined in Table 2-3.

Inhibitor	AspH variant	IC_{50} [μM]
$MnCl_2$	wt	5.1±1.1
	H679A	23.1±8.6
	H725A	>40
$CoCl_2$	wt	0.4±0.2
	H679A	1.5±0.2
	H725A	3.0±0.5
$NiSO_4$	wt	0.17±0.08
	H679A	2.7±0.4
	H725A	1.8±0.4
$Zn(OAc)_2$	wt	0.05±0.01
	H679A	0.16±0.01
	H725A	0.16±0.01
N-oxalyl-glycine (NOG)	wt	1.1±0.3
	H679A	inactive
	H725A	inactive
Pyridine-2,4-dicarboxylic acid (2,4-PDCA)	wt	0.03±0.01
	H679A	0.13±0.01
	H725A	11.2±1.4

Mn, Co, Ni and Zn are reported inhibitors of wt AspH, with Mn (IC_{50} = 5.1 μM) the least potent inhibitor while Zn (IC_{50} = 0.05 μM) was the most effective inhibitor.[229] The results from the H679A and H725A variants are consistent with this trend; however, with the variants all metal ions are overall less effective at inhibition than for wt AspH. The IC_{50} value for Zn for the H679A and H725A variants (0.16 μM) is about three-fold greater than wt AspH (0.05 μM). The IC_{50} values for Mn reveal an even greater effect of the H to A substitutions; the value for the H679A variant is over four times greater than wt AspH, and the value for the H725A variant is nearly 8 times that of wt AspH.

Differences between the wt and variant inhibition behaviour were also clearly observed with 2OG analogue inhibitors. *N*-Oxalyl-glycine (NOG) is a broad-spectrum 2OG oxygenase inhibitor which potently inhibits wt AspH (IC_{50} = 1.1 μM).[235] Interestingly, whilst inhibition of wt AspH is apparent, NOG does not appear to inhibit either the H679A or the H725A variant. Similarly, pyridine-2,4-

dicarboxylic acid (2,4-PDCA) is a potent inhibitor of wt AspH but is far less effective at inhibition of variant activity. The IC_{50} value for 2,4-PDCA for the H725A variant (11.2 μM) is over three hundred times larger than that for wt AspH (0.03 μM). In comparison the IC_{50} for the H679A variant is just over 4 times that of wt AspH for 2,4-PDCA. The reduced inhibition potency of 2OG analogues for the H679A and H725A variants is consistent with the kinetic data as the high K_m values for 2OG for the variants indicate a significant reduction in the stability of the AspH:Fe(II):2OG/2OG analogue complex. The values for 2,4-PDCA exhibit a similar trend to that of the K_m values for 2OG; the H725A substitution thus appears to have a significant impact on the ability of the enzyme to bind 2OG at the active site.

2.3.6 Crystallographic studies of the H679A and H725A AspH variants

The structural basis for the difference in kinetic parameters between wt AspH and the H679A and H725A variants was investigated by crystallographic studies. Broad crystallisation screens (Pact Premier, Hampton) were set up in an attempt to obtain crystals of the double histidine H679A/H725A variant, and the single H679A and H725A AspH variants in complex with Mn(II) in place of Fe(II) to avoid turnover occurring in solution, 2OG and either the hFX-Asp_{103} or hFX-Asn_{103} substrate.

Initial co-crystallisation solutions were based on previously reported wt AspH crystallisation conditions, that is final concentrations of 2 mM 2OG, 1mM Mn(II), 800 μM substrate and 374 μM AspH.[124] Crystals of the H679A and H725A variants from multiple conditions of the broad screen were harvested and data were collected using beamline I03 at Diamond Light Source. Both variants crystallised in the form of the $P2_12_12_1$ space group, consistent with wt AspH crystallisation behaviour.[124] The structures were solved by molecular replacement using reported wt AspH structures (PDB 5JTC or 5JQY) as search models. No crystals of the double histidine variant could be obtained from any of the broad crystallisation screens that were tested (Pact Premier, PEG/ION, BCS and LMB).

Comparison of the overall structures of the H679A and H725A variants with that of wt AspH revealed that the overall fold of AspH was generally unaffected by the active site substitutions (RMSD = 0.380 Å for wt AspH and H679A variant structures, RMSD = 0.322 Å for wt AspH and H725A variant structures). Examination of the electron density at the active sites confirmed the presence of the His to Ala substitution in both variants (Figure 2-7a-c).

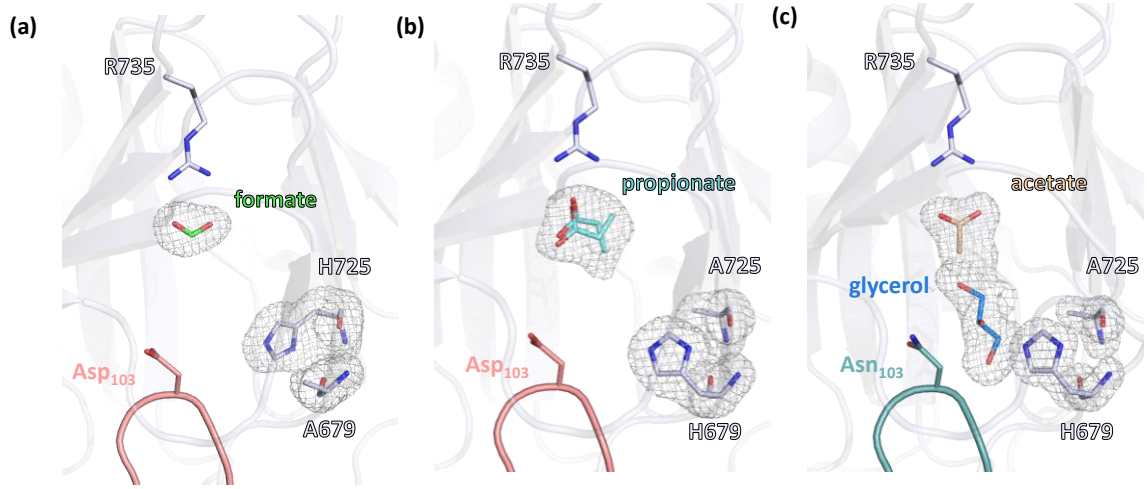


Figure 2-7. Views from crystal structures of the active sites of the H679A and H725A AspH variants. (a) The H679A active site with formate and hFX-Asp$_{103}$ (PDB 69QI).[236] (b) The H725A active site with hFX-Asp$_{103}$ and propionate (PDB 7E6J).[236] (c) The H725A active site with hFX-Asn$_{103}$ and acetate and glycerol. Polder omit density is shown carved to 3σ.[237]

Although the co-crystallisation solutions contained Mn(II) and 2OG, the active sites of the H679A and H725A variants did not contain either Mn or 2OG. In the H679A active site a formate ion originating from the buffer is positioned to interact with R735 and S668 (Figure 2-7a). R735 and S668 are usually involved in 2OG binding in the wt AspH active site (Figure 1-10). Similarly, at the H725A variant active site, electron density that could be modelled as a small molecule was observed; a propanoic acid molecule was positioned nearby R735 and S668 in the active site of the hFX-Asp$_{103}$ substrate structure (Figure 2-7b), and a glycerol molecule and acetate could be modelled in to the electron density at the active site in the hFX-Asn$_{103}$ substrate complex (Figure 2-7c). The difficulties with obtaining a metal-cofactor complex at the active site of the histidine binding variants is consistent with the high 2OG K_m values of the variants compared to wt AspH

(Table 2-4). It seems likely that 2OG is outcompeted for binding by carboxylic acids present in the crystallisation buffer.

2.3.6.1 Optimisation of crystallisation procedures

In pursuit of 2OG-bound crystal structures, the crystallisation conditions were then optimised to reflect the modified activity of the H679A and H725A variants compared to wt AspH. Firstly, the final concentration of 2OG in the crystallisation solution was increased from 2 mM to 10 mM to reflect the high 2OG K_m values of the variants. Secondly, 96-well broad screens were set up with the addition of wt AspH crystal seeds to improve crystal quality and promote the growth of crystals in conditions that did not contain an excess of small molecules in the precipitant. The cross-seeding method was found to be highly effective in obtaining crystallis in new conditions; the final concentration of the wt AspH seeds is sufficiently dilute such that it should not interfere with the structure of the variants. Finally, crystals were soaked in solutions of 10 mM 2OG (made up in 30% (w/v) PEG 400 as a cryoprotectant and 70% precipitant well solution) for 10 min prior to freezing in liquid N_2. The combination of seeding, using a higher concentration of 2OG in co-crystallisation solutions and soaking crystals with 2OG successfully gave 2OG-bound structures of both the H679A and H75A variants (Figure 2-8).

The overall fold of AspH was again unaffected by the active site His to Ala substitution in both the H679A and H725A variants when complexed with Mn, 2OG and hFX-Asn_{103} (wt AspH and H679A variant RMSD = 0.157 Å, wt AspH and H725A variant RMSD = 0.143 Å). Examination of the active sites showed similarities between the metal complexes formed in both variants. In both structures, Mn is coordinated by a single histidine residue (H725 in H679A and H679 in H725A), bidentate coordination from 2OG and two water molecules (Figure 2-8). One of these water molecules (w_2 in H679A and *w_2 in H725A) resides in the space normally occupied by the alanine substituted histidine residue.

The occupancy of the sixth Mn coordination site is different in the H725A and H679A active sites. In the H725A active site, the sixth coordination site remains unoccupied and the overall structure is virtually identical to that of wt AspH (Figure 2-8b). Interestingly, in the H679A active site, the six-coordinate geometry around the Mn centre is completed by a thiocyanate ion from the buffer (Figure 2-8a). Electron density around the SCN^- ion indicates that it is coordinating to Mn through the terminal nitrogen atom. H679A crystals were formed in the presence of KSCN which is a common precipitant in crystallisation screens. The binding of SCN^- at the AspH active site is also observed under equivalent conditions with wt AspH. This observation is discussed in further detail in Chapter 4; however, it is likely that the SCN^- ion coordinates to Fe(II) in the oxygen binding site.

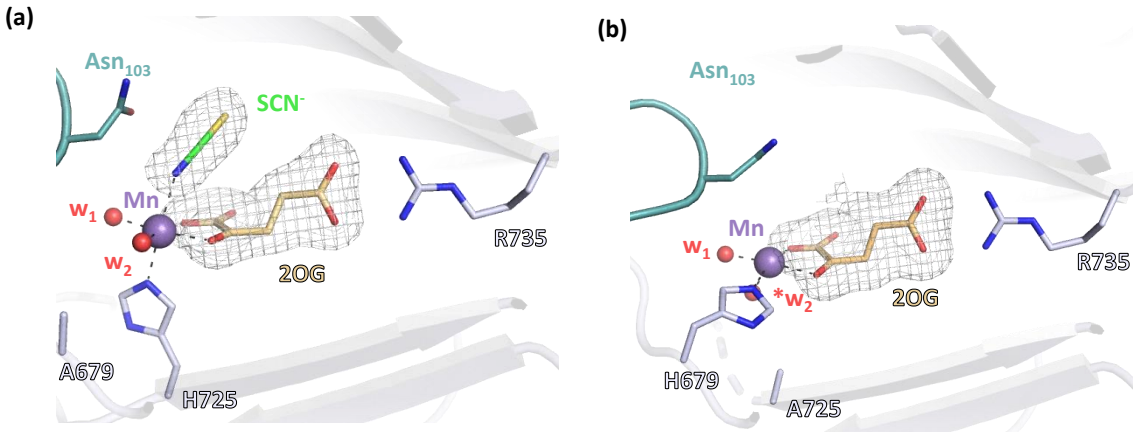


Figure 2-8. Views from the crystal structures of the active sites of the H679A and H725A AspH variants. (a) The active site of the H679A variant in complex with Mn, 2OG, hFX-Asn_{103} and SCN^-. Polder omit electron density is carved to 3σ around 2OG and SCN^-. (b) The active site of the H725A variant in complex with Mn, hFX-Asn_{103} and 2OG. Polder omit electron density is carved to 3σ around 2OG and SCN^-.

The vacant spaces created by the His to Ala substitutions are occupied by water molecules or hydroxide ions in the active site of both the H679A and H725A structures (Figure 2-9). These additional water molecules, termed w_2/*w_2, are positioned within hydrogen bonding distance to Mn-coordinating waters or carbonyls of the AspH backbone. At the H679A active site, w_3 is positioned to interact with w_2 and the main chain amide oxygen atom of W677 (Figure 2-9a). In the H725A active site, *w_3 is positioned to interact with *w_2 (Figure 2-9b).

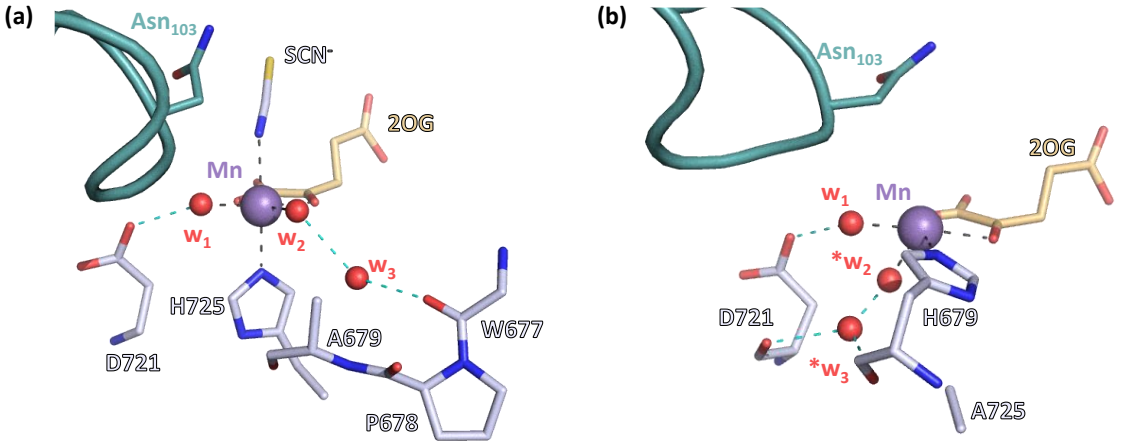


Figure 2-9. Alternative views of the active sites of the H679A and H725A variants showing the network of water molecules occupying the vacant space created by the H679A substitution (a) or H725A substitution (b). Cyan dashes show likely interactions between protein backbone or side chains and water molecules.

Superimposition of the equivalent wt AspH structures with either the H679A variant or H725A variant structures shows that the H to A substitutions have not altered the conformation of other active site residues (Figure 2-10). The 2OG-binding residues R688, H690, S668 and R735 are unperturbed by the His to Ala substitution at the metal centre. Similarly, the positioning of the hFX-Asn_{103}, Mn and 2OG in the wt AspH active site is virtually unchanged compared to that in the H679A or H725A AspH structures.

The ability of both the H679A and H725A variants to bind 2OG in a manner similar to that of the wt AspH indicates that both variants are positioned for productive catalytic activity. The additional optimisation steps required in order to detect 2OG at the active site is consistent with the kinetic data, in the particular the higher 2OG K_m values (Table 2-4), and supports the finding that 2OG complex formation is significantly impaired as a result of both of the H to A mutations.

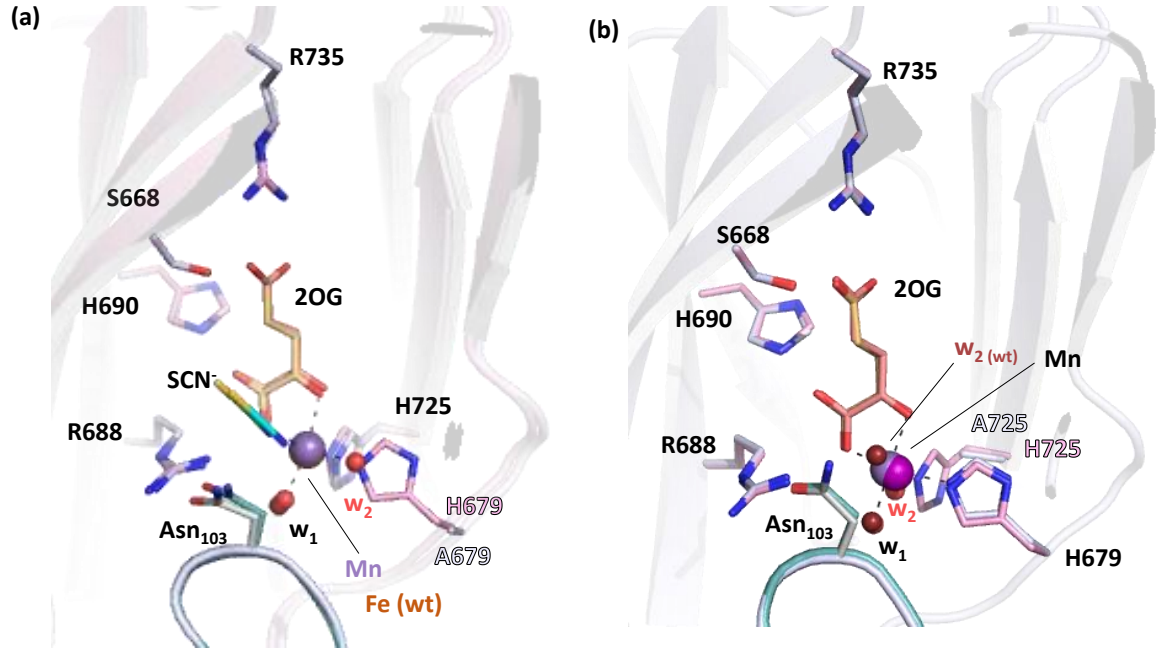


Figure 2-10. Superimposition of the active site views of the H679A (a) and H725A (b) variants with the equivalent wt AspH structure. (a) The structure of the H679A variant in complex with Mn, 2OG and SCN^- superimposed with an equivalent wt AspH structure (complexed with Fe, 2OG and SCN^-). (b) The structure of the H725A variant active site in complex with Mn, 2OG and hFX-Asp_{103} superimposed with wt AspH in complex with Mn, 2OG and hFX-Asp_{103}. H679A and H275A variant structure colouring as in Figure 2-8. Wt AspH protein residues: pink, wt AspH hFX-Asn_{103} : white, wt AspH Mn: bright purple, wt AspH 2OG: red, wt AspH water: dark red. Labels are black for features that are the same in both structures and coloured-coded to the corresponding structure for features that are different.

2.3.7 The mechanism of oxygen incorporation into the product

The consensus mechanism for 2OG-dependent oxygenases is initiated by 2OG chelating to Fe(II) at the oxygenase active site, displacing two solvent water molecules (Figure 2-11).[223] This is followed by substrate binding in close proximity to the Fe(II) centre and, at least in some cases, subsequent displacement of a third Fe(II)-bound water molecule.[55] Coordination of dioxygen to Fe(II) then affords an Fe(III)-superoxo intermediate.[223] The uncoordinated dioxygen oxygen atom attacks C2 of 2OG, and then oxidative decarboxylation of 2OG to form CO_2 and succinate occurs.[75] The resulting ferryl intermediate abstracts a hydrogen from the substrate and then in the conventional mechanism the hydroxylated product is yielded by radical rebound of the hydroxyl radical (intermediates (vii) to (viii) in Figure 2-11).[223]

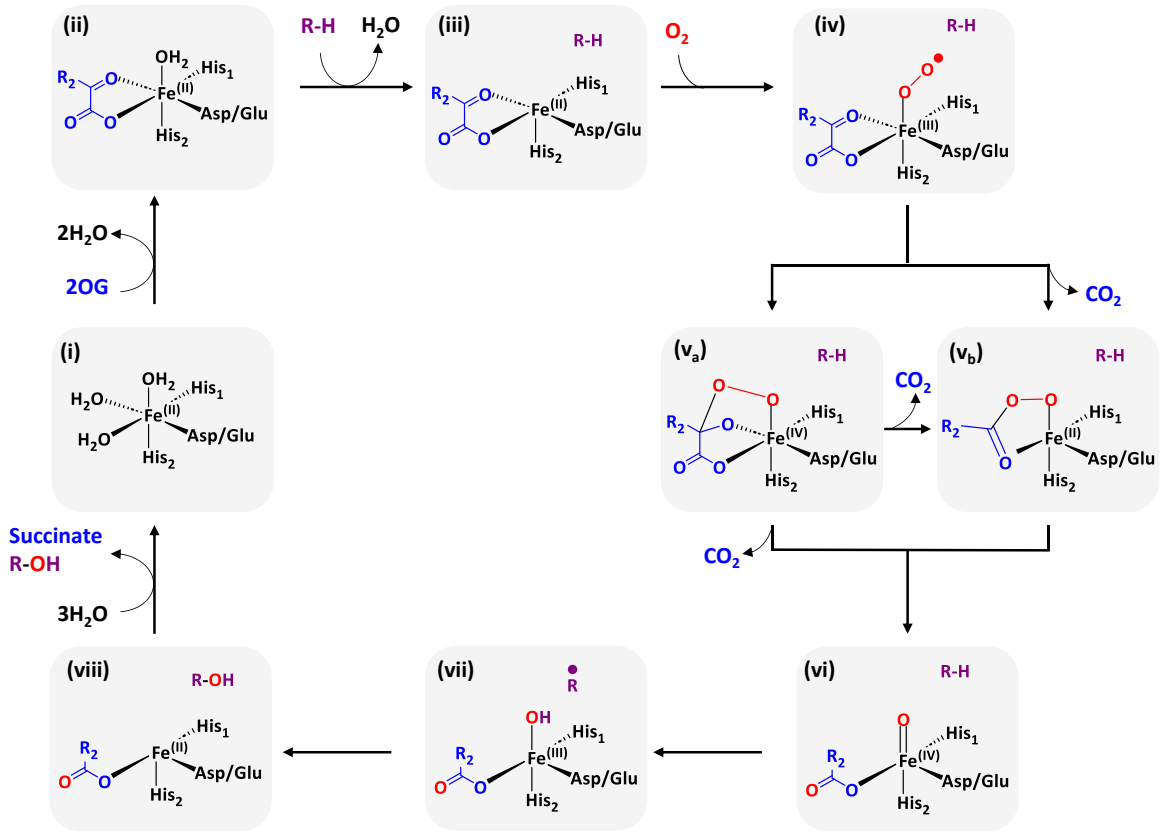


Figure 2-11. The 2OG-dependent hydroxylase catalytic mechanism. The consensus mechanism implies that oxygen in the hydroxylated product is derived from O_2 only and one oxygen from O_2 is also incorporated into succinate. His_n = histidine residues of the HXD/E…H triad. R = substrate. $R_2 = CH_2CH_2CO_2H$

Oxygen labelling experiments have provided evidence for the mechanism of oxygen incorporation during the 2OG-oxygenase catalytic cycle.[238,239] Characterisation of the products formed by 2OG-oxygenase catalysed reactions under an $^{18}O_2$ atmosphere, including the reactions catalysed by PHD2 and HIF, provided evidence that oxygen incorporated into the hydroxylated product originates from O_2.[240–243] Complementary analysis of the succinate formed during the same reaction showed that succinate generated from AlkB, PHD2, FIH, anthocyanidin synthase and flavonol synthase catalysed reactions incorporate one oxygen atom from O_2.[93] A less than stoichiometric amount of oxygen is observed in products formed by some 2OG-dependent hydroxylase catalysed reactions where solvent water exchange is proposed to occur with the reactive Fe(IV) intermediate, or subsequent intermediates.[244,245] This may be possible where such intermediates are particularly long-lived, such as in the reactions of clavaminic acid synthase, deacetoxy/deacetyl cephalosphorin C synthase, or eukaryotic lysyl hydroxylase.[242]

To date, investigations into the mechanism of oxygen incorporation in AspH catalysed reactions have not been reported. Oxygen isotope labelling experiments were therefore executed to investigate whether the AspH catalytic mechanism of oxygen incorporation is consistent with previous findings of 2OG-dependent oxygenases. The H679A and H725A variants were also subjected to the same investigations in order to examine the role of the Fe(II) coordination environment in directing the reaction pathway. Reaction assays containing LAA, 2OG, FAS, hFX-Asp_{103} and either wt AspH, the H679A variant or the H725A variant were set up under an N_2 atmosphere and then exposed to $^{18}O_2$ or $^{16}O_2$ gas using a custom-made air-tight container (P. Rabe, unpublished results). The assays were incubated overnight and then quenched under an N_2 atmosphere by addition of $10\%_{v/v}$ formic acid.

MS analysis of the product from the wt AspH catalysed reaction showed evidence for near complete incorporation of an oxygen atom originating from gaseous O_2 into the product (Figure 2-12). The $[M+4H]^{4+}$ mass of the product formed under $^{18}O_2$ conditions was +0.5 Da greater than the $[M+4H]^{4+}$ mass of the product formed under $^{16}O_2$ conditions. This corresponds to an overall mass increase of +2 Da in the product, that is the difference between ^{18}O and ^{16}O.

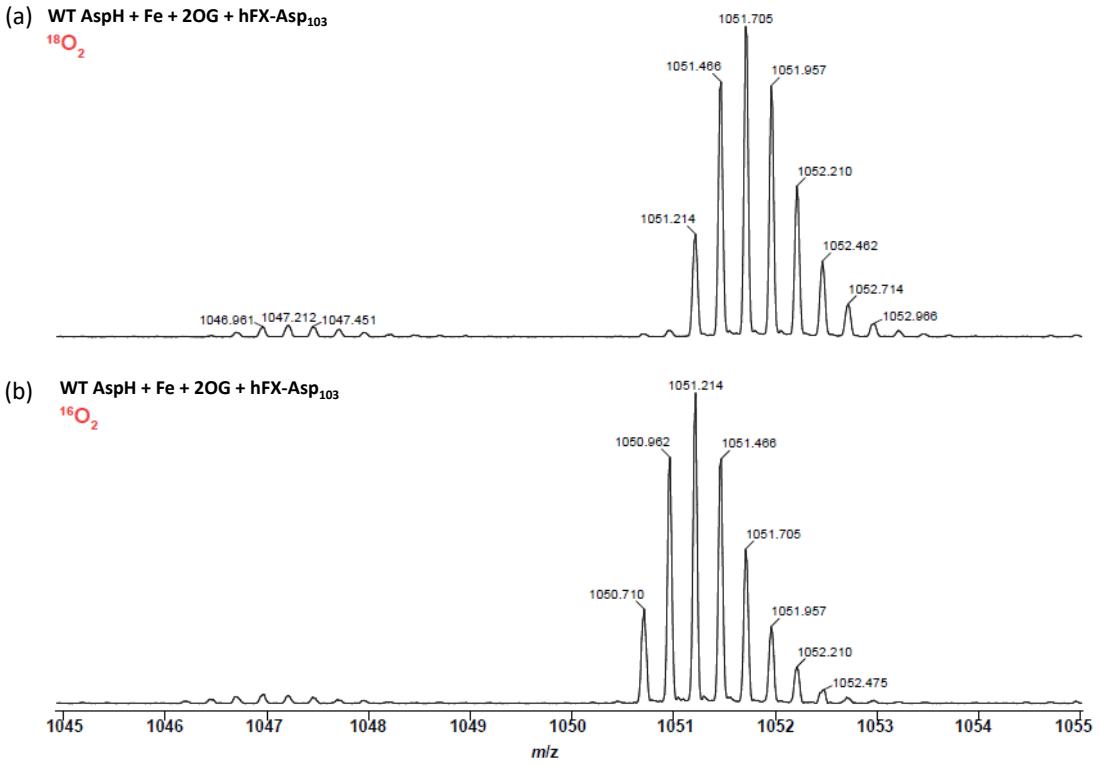


Figure 2-12. LC-MS trace of product formed in reactions performed under $^{18}O_2$ atmosphere. The spectra show the regions containing $[M+4H]^{4+}$ substrate and product masses. Calculated substrate mass: 1047.2 Da. Calculated ^{16}O product mass: 1051.2 Da. Calculated ^{18}O product mass: 1051.7 Da. Buffer conditions: 50 mM HEPES pH 7.5 with 50 mM NaCl. Assay conditions: 100 µM LAA, 200 µM 2OG, 20 µM FAS, 5 µM hFX-Asp_{103}, 0.1 µM enzyme.

The results from the reactions catalysed by the H725A variant showed the same behaviour, with oxygen in the product originating from dioxygen only (Figure 2-13a and b). By contrast, evidence from the product resulting from H679A catalysed reactions under an $^{18}O_2$ atmosphere showed incorporation of a mixture of both ^{16}O and ^{18}O. The mass of the product formed under an $^{18}O_2$ atmosphere was around +0.25 Da greater than the mass of the product formed under $^{16}O_2$, which corresponds to a mix of ^{16}O and ^{18}O atoms (Figure 2-13c and d). It has not yet been possible to directly quantify the proportion of isotopes in the product due to the overlap of the isotope patterns in the spectra.

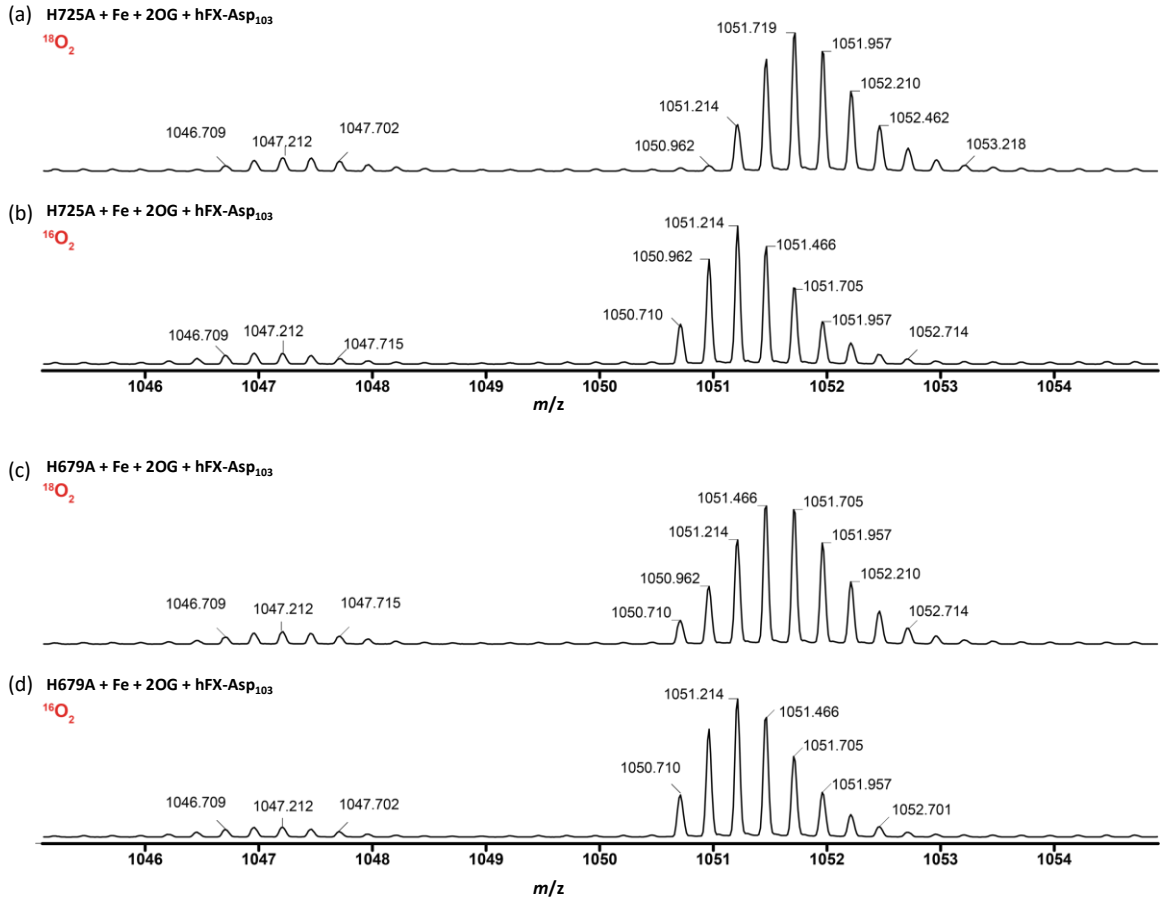


Figure 2-13. LC-MS trace of product formed in reactions performed under $^{18}O_2$ and $^{16}O_2$ atmospheres, catalysed by the H679A and H725A AspH variants. The spectra shows the region containing the $[M+4H]^{4+}$ substrate and product masses. Calculated substrate mass: 1047.2 Da. Calculated mass of product containing 100% ^{16}O: 1051.2 Da. Calculated mass of product containing 100% ^{18}O: 1051.7 Da. Spectra show product of H725A AspH catalysed reactions performed under $^{18}O_2$ (a) and under $^{16}O_2$ (b), and product of H679A AspH catalysed reaction performed under $^{18}O_2$ (c) and $^{16}O_2$ (d). Buffer conditions: 50 mM HEPES pH 7.5. Assay conditions: 100 µM LAA, 2000 µM 2OG, 20 µM FAS, 5 µM hFX-Asp_{103}, 0.2 µM enzyme.

Complementary reactions were then performed using solutions composed of $^{18}OH_2$ under atmospheric conditions (that is, exposure to $^{16}O_2$ only). The results from these assays (Figure 2-14) were consistent with previous observations. Products formed from reactions catalysed by wt AspH or the H725A variant showed incorporation of ^{16}O from gaseous atmospheric O_2 (Figure 2-14a and c). By contrast, the product of the H679A-catalysed reaction again showed evidence for incorporation of a mix of ^{16}O originating from atmospheric O_2 and ^{18}O from the $^{18}OH_2$ solutions (Figure 2-14b).

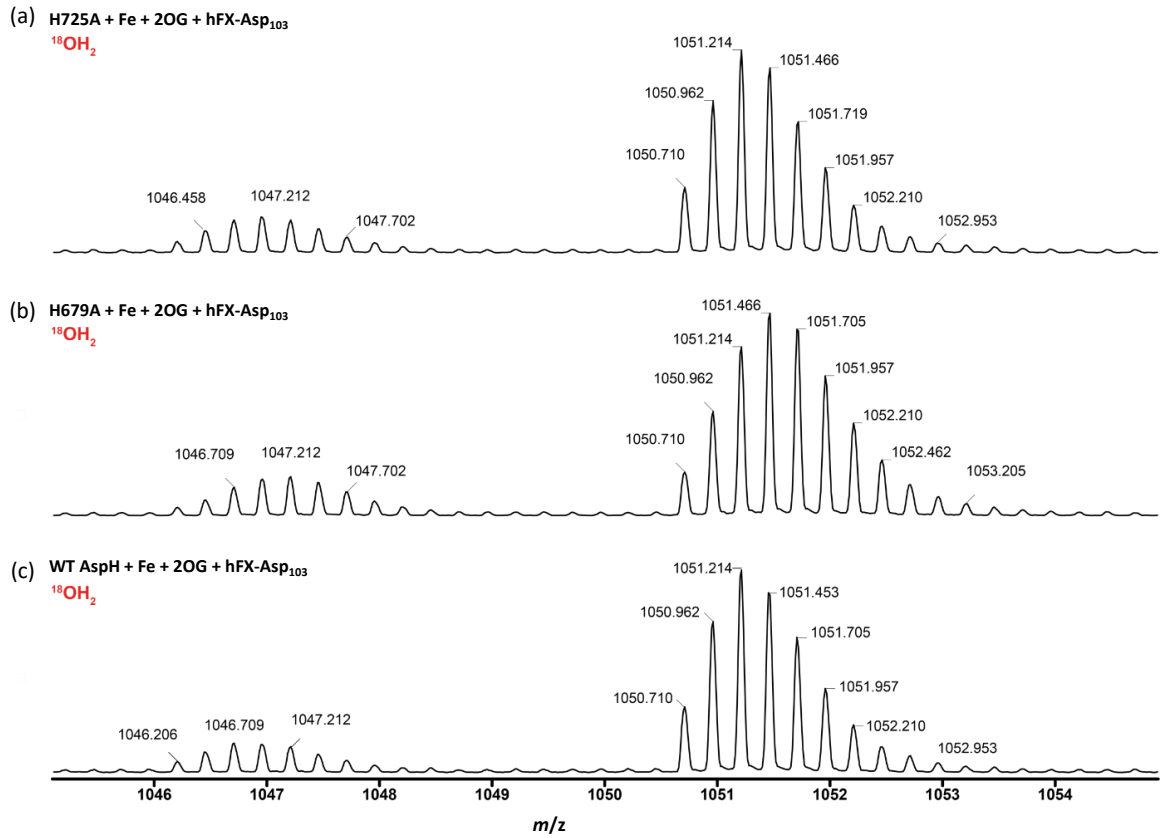


Figure 2-14. LC-MS trace of product formed in ^{18}O labelled reactions. Spectrum shows region containing $[M+4H]^{4+}$ substrate and product masses. wt AspH assay conditions: 50 mM HEPES pH 7.5 with 50 mM NaCl, 100 µM LAA, 200 µM 2OG, 20 µM FAS, 4 µM hFX-Asp_{103}, 0.1 µM enzyme. H679A and H725A variant assay conditions: 50 mM HEPES pH 7.5, 100 µM LAA, 2000 µM 2OG, 20 µM FAS, 5 µM hFX-Asp_{103}, 0.2 µM enzyme.

2.3.7.1 The pH dependence of the reaction pathway

The initial oxygen labelling experiments were carried out at pH 7.5, close to physiological pH. Next, the pH dependence of the AspH reaction pathway was investigated. It was proposed that at high pH values a higher concentration of hydroxide ions in the solvent may lead to a great likelihood of solvent exchange occurring with the Fe(II) intermediate and/or the mechanism may be affected by the active site protonation state. The $^{18}O_2$ experiments were therefore repeated for wt AspH and both the H679A and H725A variants at pH values 6, 6.5, 7, 7.5, 8, 8.5.

The products were analysed by LC-MS and the proportion of oxygen isotopes in the products was assessed visually. Raw data are presented in the appendix (Figure 9-1). Unfortunately, the H679A and H725A variants were not active at pH 8 or 8.5 and therefore the full range of pH values could not be assessed.

The products formed by wt AspH and H725A variant catalysed reactions consistently showed >99% incorporation of ^{18}O into the product across all tested pH values. The products formed by H679A variant catalysed reactions were derived from a mixture ^{18}O and ^{16}O at all pH values. Therefore, it is possible to conclude that the AspH reaction pathway, and the diverging mechanism of the H679A variant, are not pH dependent.

2.3.7.2 The mechanism of oxygen incorporation in AspH

A summary of the hydroxylation reactions catalyzed by wt AspH and the H679A and H725A variants is illustrated in Figure 2-15.

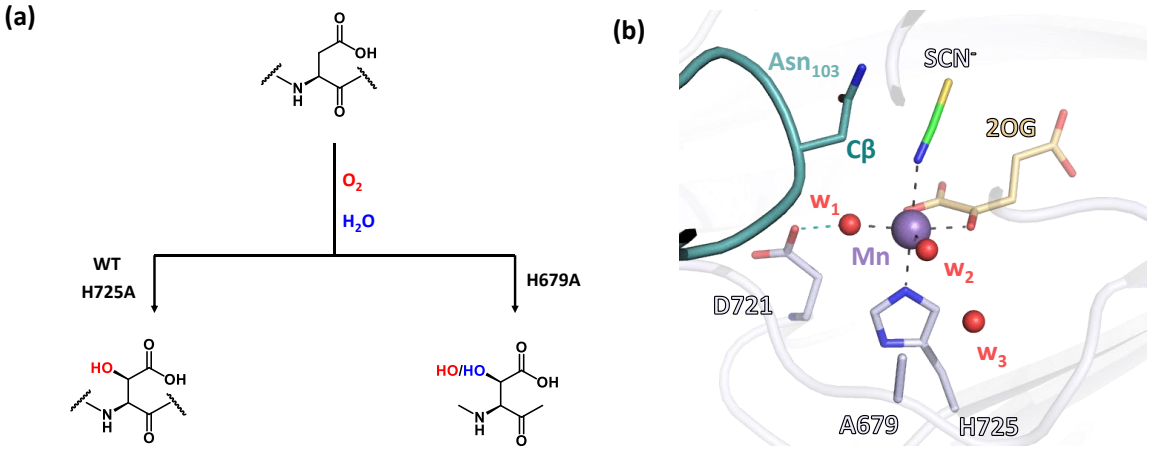


Figure 2-15. (a) Illustrative summary of the reaction outcomes in the hydroxylation reactions catalysed by wt AspH and the H725A variant, compared to the H679A variant. (b) View from the crystal structure active site of the H679A variant.

There are two possible options for the origin of the oxygen incorporated into the hydroxylated product in the H679A variant catalysed reactions. Examination of the H679A variant crystal structure in complex with Mn, 2OG and hFX-Asp_{103} revealed that a water molecule (termed w_2) occupies the vacant coordination space created by the H679A substitution (Figure 2-15b). Additionally, another water molecule (termed w_1), is present in both the wt AspH active site and the H679A and H725A variant active sites. W_1 is positioned to interact with the side chain oxygen of D721. The D721A substitution reportedly renders AspH inactive and likely removes the stabilisation of w_1 at the metal centre. It can therefore be proposed that the loss of w_1 is

inactivating. Consequently, it may be more likely that w_2, which is unique to the H679A variant active site, is the source for water-derived oxygen incorporation into the product.

The proposed AspH mechanism, based on the consensus mechanism for 2OG-dependent hydroxylases is shown in Figure 2-11. This mechanism suggests that oxygen in the wt AspH product will be derived from O_2 only. While the $^{18}O_2/^{18}OH_2$ labelling studies show to be true for wt AspH, the same reaction scheme cannot be accurate for the H679A variant, when the product manifests partial incorporation of oxygen from water. Note, it is assumed that the C3-stereo selectivity of the hydroxylated products is the same in as in the consensus mechanism; however, this is not necessarily the case and this is being investigated in ongoing work.

There are at least three possible pathways by which a hydroxide from a solvent water molecule could be incorporated into the product (Figure 2-16). These are: (a) Direct recombination of oxygen derived from w_2 from the A679 position, (b) incorporation of w_2 derived oxygen following isomerisation of ferryl intermediate (vi), or (c) incorporation of w_2 following isomerisation of intermediate (vii).

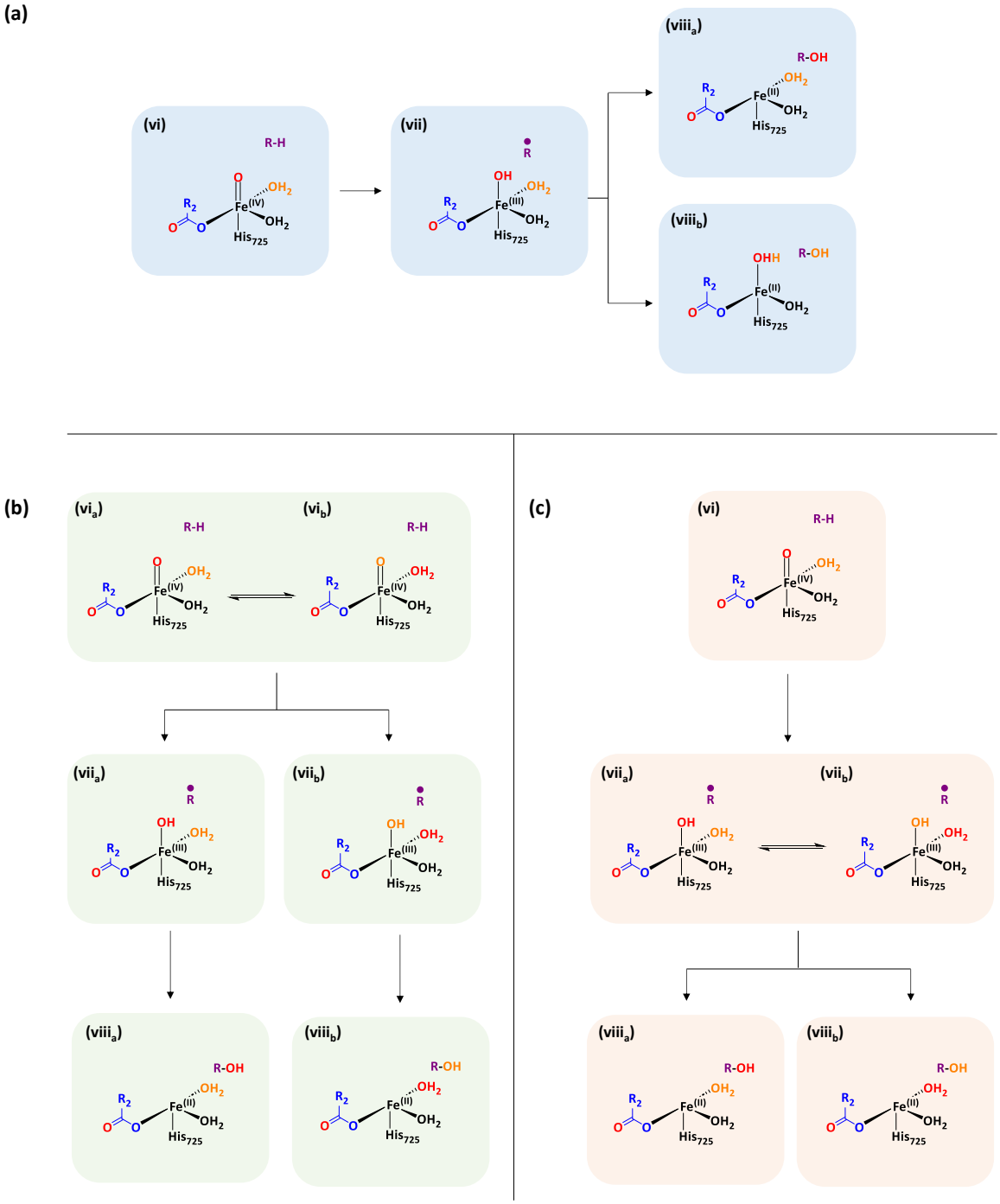


Figure 2-16. Possible reaction pathways that lead to the incorporation of OH originating from water in the H679A AspH variant mechanism. Orange OH_2 – w_2 occupying H679 space, red OH = from O_2, black OH_2 – w_1, observed in wt AspH structures and coordinates to D721. (a) Direct incorporation of either w_2 (orange) from original coordination site (b) Isomerisation of intermediate (vi) followed by protonation, (c) isomerization of ferric intermediate (vii) after protonation of the ferryl species. R = substrate, $R_2 = CH_2CH_2CO_2H$

Direct recombination of the substrate radical with a water molecule without rearrangement of the intermediate is one possible reaction mechanism (Figure 2-16(a)). However, in the H679A crystal structure w_2 is positioned ~4.7 Å from Cβ of Asn_{103}. This is considerably further away from the likely

position of the hydroxyl formed from the cleavage of O_2. Therefore, this or closely related pathways seem unlikely to be the most favourable.

The proposed alternative pathways, involving isomerisation of intermediates (vi) or (vii) have been proposed in part based on studies of the halogenase SyrB2 mechanism and a 2OG oxygenase study on cephalosporin biosynthesis.[96,245,246] Mossbauer spectra revealed that the equivalent ferryl intermediate (vi) in the SyrB2 mechanism (with a Cl^- coordinating in the w_1 site), is a mixture of two species that exist in fast equilibrium.[68,247] QM/MM calculations have reported that it is possible, in theory, for the rearrangement of coordinating intermediates to occur.[104,246] Such a rearrangement was originally proposed to involve isomerisation of the ferryl intermediate (intermediate (vi)), however, later calculations reported it is more likely to involve the ferric intermediate (intermediate (vii)).[246] The isomerisation process is proposed to be facilitated by a vacant coordination site as a result of monodentate succinate coordination.[96,246] It is possible that an analogous mechanism occurs in the H679A variant catalysed reaction. Isomerisation of ferric intermediate (vii) would position the hydroxyl group derived from water close to Cβ of the substrate, and would be more favourable for recombination. Pathways (b) or (c) are both feasible for the H679A variant and, at this time, it is not possible to determine which is more likely based on current data. Future detailed mechanistic work could study the characterisation of intermediates using Mössbauer, EPR or UV-vis spectroscopy, including using Cβ ^{2}H-labelled Asp/Asn residues with the aim of identifying the rearrangement step.

In order for a rearrangement to occur, the preceding intermediate must be relatively long-lived. Loss of H679 may therefore destabilise the transition state between intermediates (vi) and (vii) (pathway (b)), or (vii) and (viii) (pathway (c)), thus leading to a longer-lived intermediate than that which occurs in the wt AspH mechanism. H679 therefore likely has a key role in stabilising transition states in the latter part of the reaction mechanism, and modification of H679 can therefore alter the reaction pathway taken.

2.3.8 Investigation into the incorporation of anions in AspH reaction

Although it has been established that the Fe(II)-binding motif is not solely responsible for controlling the reaction outcome in 2OG-dependent hydroxylases and halogenases, some cases of 2OG-dependent hydroxylase active site engineering have resulted in a gain of halogenase activity. For example, the *N*-acyl amino acid hydroxylase SadA was reported to perform halogenation of an *N*-succinyl-L-Leu substrate following substitution of the aspartic acid in the HXD…H Fe(II)-binding triad with glycine, to give the AspH and halogenase HXG…H dyad.[114]

No evidence for wt AspH catalyzed halogenation has yet been reported. Evidence from ^{18}O labelling studies suggests that radical recombination of a hydroxy group derived from O_2 is favoured during the hydroxylation step, and therefore other anions present at the active site are unlikely to be incorporated into the product. By contrast, evidence from ^{18}O labelling studies indicated that the AspH H679A variant can catalyse two possible reaction pathways, one of which facilitates the incorporation of an Fe(II)-bound water molecule that is not involved in the wt AspH mechanism. Therefore it was proposed that if other anions coordinated to Fe(II) during catalysis by the H679A variant, these could recombine with the substrate

To test this hypothesis, the reaction outcomes of the H679A variant catalysed reactions in the presence of a range of anions were assessed. 100 µL volume standard H679A activity assays (conditions as in Table 2-3) were set up with the addition of Cl^-, Br^-, SCN^- and N_3^- in ratios of 1:1, and 1:1000 substrate:anion. The reactions were allowed to proceed for 2 h at room temperature before quenching with 10 $\%_{v/v}$ formic acid. The samples were analysed by LC-MS.

Table 2-6. Results of the reactions catalysed by the H679A variant in the presence of Cl^-, Br^-, SCN^- and N_3.

Anion	Product from H679A variant catalysed reaction
Cl^-	Hydroxylated
Br^-	Hydroxylated
SCN^-	Hydroxylated
N_3^-	Hydroxylated

In the presence of Cl^-, Br^-, SCN^- or N_3^- ions, the only product observed from the H679A variant catalyzed reaction is the normal hydroxylated product (Table 2-6). It is possible that these anions do not coordinate to Fe(II) at the active site, or if coordination occurs, the anion rebound path is less favorable. The crystal structure of the H679A variant in complex with Mn, 2OG and hFX-Asn_{103} revealed an SCN^- ion coordinating to the Mn centre (Figure 2-8a). However, the position of the SCN^- coordination site is likely to be the position of the O_2 binding site, and not the vacant site occupied by w_2. Therefore it is possible that any anions that can coordinate to Fe(II) may subsequently be displaced by O_2 during catalysis.

2.3.9 Study of an AspH catalytic triad

Sequence alignments of 2OG-dependent hydroxylases and AspH show that a glycine residue, G681, is located in the position of the AspH sequence where a third acidic Fe(II)-ligating residue would normally be found in most hydroxylases.[124] Having shown that AspH variants with a single protein-bound Fe(II) ligand remain catalytically active, it was of interest to investigate the effect of introducing a third protein-bound Fe(II) ligand at the AspH active site to restore the typical hydroxylase binding triad; thus, the G681 wt AspH residue was targeted by mutagenesis studies.

The G681D and G681E substitutions were introduced in the wt AspH gene by site-directed mutagenesis (performed by Dr Inga Pfeffer), affording AspH variants with HXD…H and HXE…H Fe(II)-binding triads. The G681D and G681E AspH variants were produced and purified as described previously for the H679A and H725A variants (Section 2.3.1). The activity of the G681D and G681E

variants was initially assessed by monitoring of hFX-Asp_{103} hydroxylation by real-time SPE-MS analysis in 50 mM HEPES pH 7.5 in the absence and presence of NaCl (Figure 2-17).

Interestingly, both of the G681D and G681E variants catalysed the hydroxylation of hFX-Asp_{103} (Figure 2-17). The G681D variant catalysed hydroxylation of the substrate faster than the G681E variant in the absence of NaCl in the buffer (Figure 2-17a). In the presence of 50 mM NaCl both of the G681D and G681E variants catalysed hydroxylation at virtually the same rate. In the absence of NaCl (Figure 2-17b) or presence of NaCl (Figure 2-17c) in the buffer, wt AspH was significantly faster at catalysing hydroxylation than either of the G681D or G681E variants. The presence of a third iron-binding residue, that is D or E, at the active site did not improve the activity of AspH, and in fact hindered catalysis in comparison with wt AspH. Compared to the single Fe(II)-binding variants, that is the H679A and H725A variants, catalysis was impaired by the D or E substitutions (Figure 2-17d). Time limitations of the project meant that a full characterisation of the kinetics of the G681D and G681E variants was not subsequently performed.

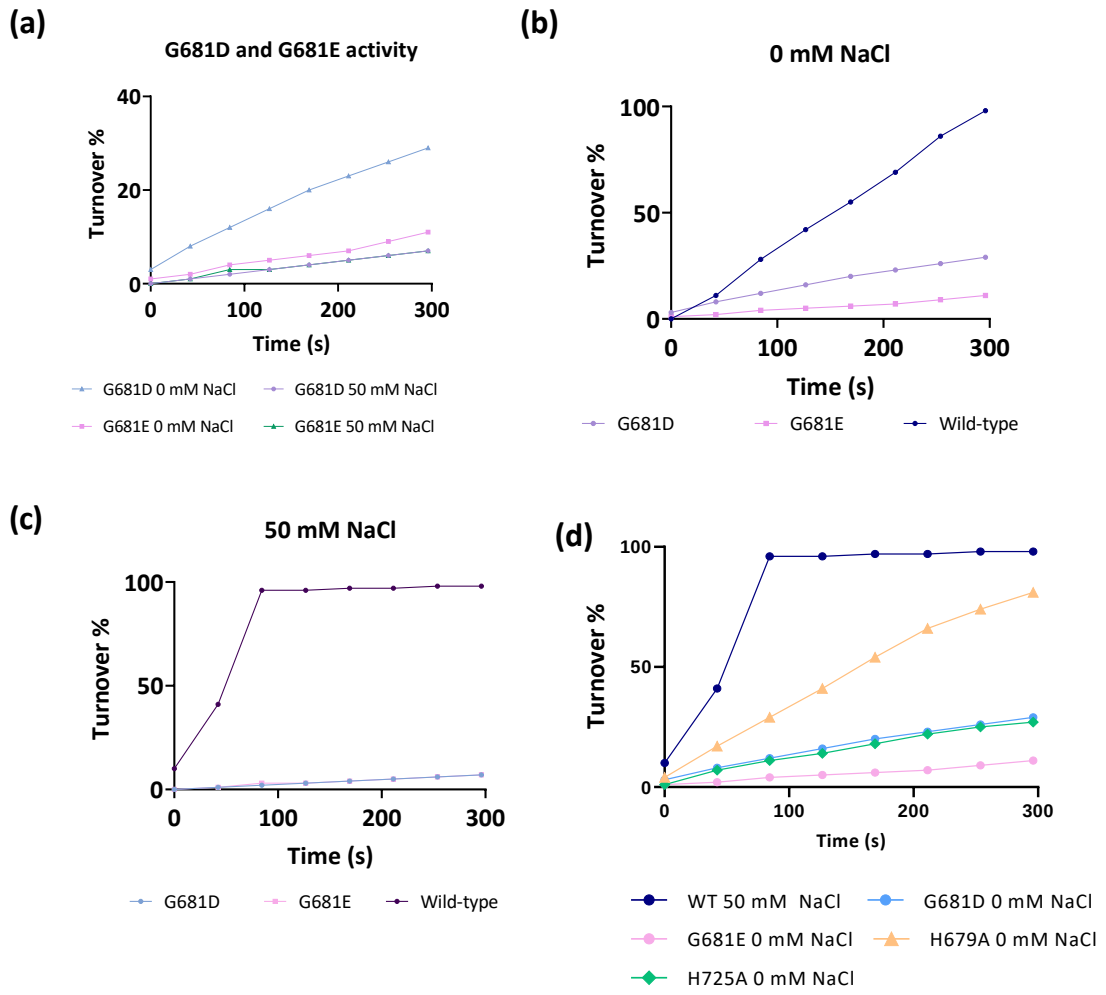


Figure 2-17. The activity of the G681D and G681E variants. (a) Overview of activity of both variants in pH 7.5 buffer with and without NaCl. (b) Activity of variants compared to wt in pH 7.5 buffer without salt and in presence of 50 mM NaCl (c). (d) Comparison of activity of wt AspH, G681E, G681D, H679A and H725A variants

Despite repeated attempts to crystallise the G681E and the G681D variants, no crystals could be obtained. Broad screens including Pact Premier, BCS and LMB were used with the addition of wt AspH crystal seeds, a method which had previously proved to be very effective. Therefore, the structural basis of the reduced activity of the G681E and G681D variants could not be investigated. It is unknown as to whether the substituted E and D residues could be positioned for Fe(II) ligation, or whether the presence of the acidic side chains are not positioned in sterically productive conformations. Alternatively, the impaired hydroxylation activities of the G681E and G681D variants could be attributed to the increased steric bulk of side chains around the active site that may interfere with substrate binding. Overall, it is possible to conclude that the unusual AspH dyad,

HXG…H, is highly specialised for AspH; however, given that the G681D and G681E variants are active it is unclear why AspH has evolved to use a dyad rather than a triad of Fe(II) binding residues.

2.4 Conclusions and future work

The worked described in this chapter aimed to investigate the nature of Fe(II) cofactor binding at the AspH active site by use of two Fe(II)-binding variants, H679A and H725A, and to assess the importance of the catalytic triad in the AspH mechanism. Previous studies indicated that both H679 and H725 residues were essential for AspH catalysis, however more recent results disputed these conclusions.[213] Work performed in this chapter confirmed that AspH variants employing a single Fe(II)-binding ligand, that is the H679A and H725A variants, are catalytically active, but not the H679A/H725A double mutant which lacks protein-bound Fe(II) ligands. The discrepancies between these findings can perhaps be attributed to differences in the reaction conditions, as initial activity tests of the H679A and H725A variants in lab-based assays showed the variants are highly sensitive to changes in NaCl concentration in the buffer.

From a study of the kinetic parameters of the H679A and H725A variants, it is apparent that the His to Ala substitution most significantly impairs the ability of AspH to bind 2OG at the active site, as judged by 2OG K_m values (Table 2-4). Surprisingly, in comparison to 2OG binding, Fe(II) kinetic parameters are minimally affected by the His to Ala substitutions. Crystal structures of the H679A and H725A variants revealed that 2OG can coordinate at the active site in the same conformation as observed in wt AspH, although a much higher concentration of 2OG is required in co-crystallisation solutions in order to observe electron density that can be modelled as 2OG. In future, comparison of thermodynamic parameters could be attempted to obtain binding constants for cofactors.

Investigations into oxygen incorporation into the product also revealed the significance of H679 in directing the reaction pathway. The H679A variant was able to catalyse the final radical rebound step of the AspH mechanism via two pathways, leading to incorporation of a hydroxide group

derived from the solvent into the product. It is therefore probable that H679 has a role in stabilising reaction Fe(IV) or Fe(III) intermediates.

A summary of the roles of Fe(II)-binding residues in AspH, and key residues in the vicinity of the Mn centre is shown in Figure 2-18.

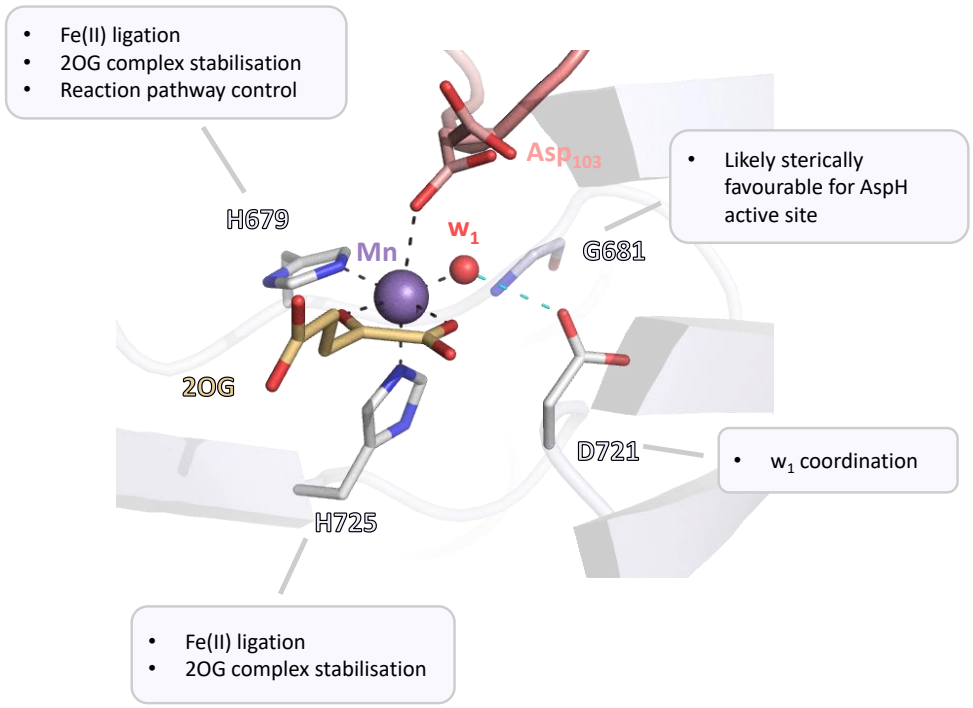


Figure 2-18. A summary of the roles of the Fe(II)-binding ligands and associated residues at the AspH active site.

The discovery that AspH variants are catalytically active with only one directly enzyme provided ligand has potential general implications for the metal-enzyme and non-enzyme metal dependent catalysis. The widespread perception that a facial triad of residues is required for non-heme oxygenase catalysis is incorrect as evidenced by the wt AspH and the 2OG-dependent halogenases, which employ a 2His motif to coordinate the active site Fe(II). Sequence alignments indicate the potential that at least one human oxygenase, AspHD1, may have only a single enzyme ligating Fe-ligand. This proposal needs to be experimentally tested; attempts to characterise AspHD1 are outlined in Chapter 5.

2.4.1.1 Future work: engineering of AspH to catalyse halogenation reactions

The functionalization of sp^3 carbon centres is generally challenging to perform via conventional synthetic methods. Thus, there is interest in understanding the mechanism of 2OG-dependent halogenases, which are capable of catalyzing the halogenation of substrates under ambient conditions. However, the halogenases often act on specific protein-tethered substrates and this limits the scope of potential substrates. In comparison, the closely-related 2OG-dependent hydroxylases act on a range of substrates, from small molecules, e.g. lysine, to larger proteins, e.g. EGFDs. Therefore, there is interest in expanding the potential reactions catalyzed by 2OG-dependent hydroxylases.

Engineering of 2OG-dependent hydroxylases to perform halogenation has proved to be challenging but possible. Neugebauer et al. successfully modified a lysine hydroxylase to catalyse halogenation of lysine by identifying key differences in the secondary sphere residues at the active site of the lysine hydroxylase and the closely related halogenase, BesD.[103] Mutation of a single hydrophobic valine residue to a polar asparagine resulted in a lysine hydroxylase variant that catalyzed the formation of the chlorinated product.[103] Further modification of secondary sphere residues produced a hydroxylase variant that performed halogenation with comparable activity to and better selectivity than another closely related halogenase.[103]

The hydroxylation reaction catalysed by the H679A AspH variant showed incorporation of oxygen derived from water into the product. The mechanism for this reaction is likely to be similar to that which occurs in the 2OG-dependent halogenases. It may therefore be possible to further engineer the H679A AspH variant to catalyse halogenation reactions.

An overlay of the structures of the halogenases WelO5, SyrB2, Bes and HalB with the H679A crystal structures centred around AspH the first H residues of the HXG…H dyad enabled the for identification of key differences in the properties of secondary sphere residues around the AspH active site compared to the halogenase active site (Figure 2-19). The valine residue identified by

Neugebauer et al. as being key to controlling halogenation in a lysine hydroxylase is positioned in close proximity to two valine residues in AspH, that is V676 and V727. In the halogenases SyrB2, BesD and HalB these valine residues are equivalent to polar residues, either threonine, histidine or serine (Table 2-7 and Figure 2-19).[46,70,103]

Table 2-7. Non-polar hydrophobic residues at the AspH active site and comparable residues in 2OG-dependent halogenases based on superimposition of the crystal structures.

AspH residue	Residue in equivalent position in halogenase		
	SyrB2	BesD	HalB
Val 727	Ser 237	Thr 206	Thr 215
Val 676	Thr 113	His 134	His 143

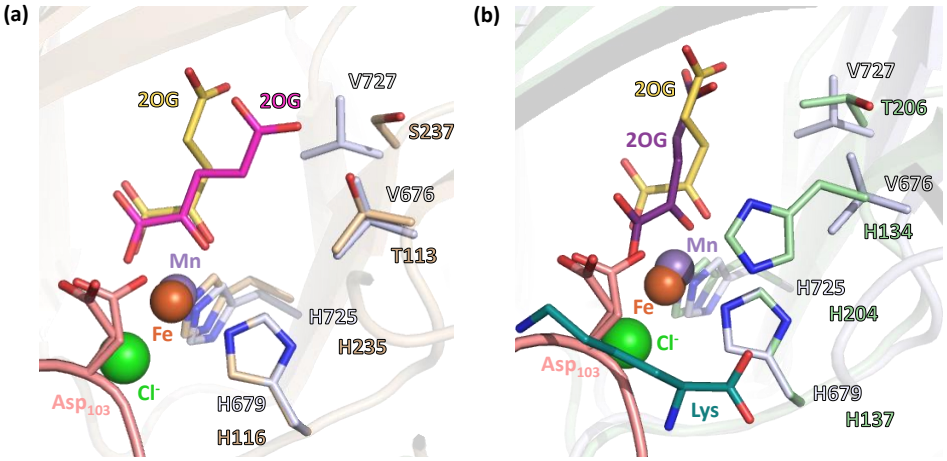


Figure 2-19. Superimposition of the structures of the active sites of the halogenases SyrB2 (a) and BesD (b) with the AspH active site. Superimpositions are centred around first histidine reside of the HXG…H dyad. The hydrophobic pocket in the AspH active created by V676 and V727 overlays with polar residues in both active sites of the halogenases. AspH colouring: protein – blue-white, 2OG – yellow, hFX-Asp_{103} – salmon. SyrB2 colouring: protein – beige, 2OG – pink, Cl^- - green, Fe – orange. BesD colouring: protein – green, 2OG – purple, Cl^- - green, Fe – orange, lysine substrate – teal.

Future work could aim to engineer an AspH variant that possesses halogenase activity. A starting point for this project would be to introduce mutations in V727 and V676 residues of wt AspH and the H679A variant, for example substitution of either or both valine residues (V727 and V676) with a polar amino acid such as threonine or histidine.

2.5 Acknowledgements

Plasmids encoding the H679A, H725A, G681D and G681E variants of AspH were produced by site-directed mutagenesis by Dr Inga Pfeffer. Inhibition assays and 2OG turnover NMR experiments were performed by Dr Lennart Brewitz. Crystals for structures 7E6J were harvested by and the data were refined by Dr Yu Nakashima. Crystal data for structure 6Q9I was refined by Dr Rashed Chowdhury, and crystallisations were performed by Dr Inga Pfeffer.

2.6 Supplementary information

Table S2-1. Data collection and refinement statistics for crystal structures of the H679A and H725A AspH variants (Figure 2-7a and Figure 2-7b). The crystal for 7E6J was harvested by and data were refined by Dr Yu Nakashima. The crystallisation for 6Q9I was performed by Dr Inga Pfeffer and data were refined by Dr Rashed Chowdhury. Table adapted from Brasnett et. al.[236]

	H679A:formate:hFX-Asp$_{103}$ (PDB 6Q9I)[236]	**H725A:propionate:hFX-Asp$_{103}$ (PDB 7E6J)[236]**
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Symmetry	orthorhombic	orthorhombic
Cell dimensions:		
a, *b*, *c* (Å)	50.13, 91.36, 123.54	50.24, 86.84, 123.54
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
X-Ray source	Synchrotron (DLS I04)	Synchrotron (DLS I03)
Resolution (Å)	46.45-1.85 (1.92-1.85)	46.54-1.90 (1.93-1.90)
R_{merge}	0.077 (1.000)	0.153 (1.792)
I / σI	27.1 (1.6)	10.5 (0.8)
CC (1/2)	0.998 (0.705)	0.998 (0.699)
Total number unique reflections	49046 (2451)	43495 (2159)
Completeness (%)	100.0 (100.0)	100.0 (99.3)
Multiplicity	9.8 (9.5)	13.3 (13.7)
Refinement		
R_{work} / R_{free}	0.202 / 0.223	0.170 / 0.213
No. atoms:	3923	7532
B-factors:	59.0	43.0
R.m.s. deviations:		
Bond lengths (Å)	0.007	0.002
Bond angles (°)	0.973	0.571

Table S2-2. Data collection and refinement statistics for the crystals structures of the H679A and H725A AspH variants (Figure 2-7c and Figure 2-8a)

	H725A:glycerol:acetate:hFX-Asp_{103}	H679A:2OG:Mn:SCN^-:hFX-Asn_{103}
Wavelength (Å)	0.9763	0.9762
Resolution range (Å)	50.12 - 1.8 (1.864 - 1.8)	46.34 - 2.0 (2.071 - 2.0)
Space group	P 21 21 21	P 21 21 21
Unit cell		
a, b, c (Å)	49.983 86.1559 123.234	49.97 86.61 123.86
α, β, γ (°)	90 90 90	90 90 90
Total reflections	670304 (57862)	481216 (48691)
Unique reflections	50086 (4856)	37129 (3661)
Multiplicity	13.4 (11.9)	13.0 (13.3)
Completeness (%)	99.85 (98.76)	99.98 (100.00)
Mean I/sigma(I)	19.22 (1.77)	11.53 (1.35)
Wilson B-factor ($Å^2$)	24.70	41.82
R-merge	0.08184 (0.7114)	0.1265 (1.868)
CC1/2	0.999 (0.909)	0.998 (0.619)
CC*	1 (0.976)	1 (0.874)
Reflections used in refinement	50074 (4854)	37125 (3661)
Reflections used for R-free	2000 (195)	1940 (191)
R-work	0.1655 (0.2289)	0.1841 (0.2778)
R-free	0.1934 (0.2932)	0.2123 (0.3226)
CC(work)	0.964 (0.924)	0.961 (0.783)
CC(free)	0.961 (0.855)	0.952 (0.644)
Number of non-hydrogen atoms	4062	3817
macromolecules	3577	3511
ligands	21	18
solvent	475	292
Protein residues	447	446
RMS(bonds)	0.012	0.012
RMS(angles)	1.16	1.38
Ramachandran favored (%)	98.65	99.55
Ramachandran allowed (%)	1.35	0.45
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.55	0.28
Clashscore	3.68	4.50
Average B-factor	28.86	45.17
Macromolecules ($Å^2$)	27.73	45.14
Ligands ($Å^2$)	28.77	36.75
Solvent ($Å^2$)	37.38	45.99

Table S2-3. Data collection and refinement statistics for the crystals structures of the H679A and H725A AspH variants (Figure 2-8b)

	H725A:2OG:Mn:hFX-Asn$_{103}$
Wavelength (Å)	0.9763
Resolution range (Å)	50.12 - 1.8 (1.864 - 1.8)
Space group	P 21 21 21
Unit cell	
a, b, c (Å)	49.983 86.1559 123.234
α, β, γ (°)	90 90 90
Total reflections	670304 (57862)
Unique reflections	50086 (4856)
Multiplicity	13.4 (11.9)
Completeness (%)	99.85 (98.76)
Mean I/sigma(I)	19.22 (1.77)
Wilson B-factor (Å^2)	24.70
R-merge	0.08184 (0.7114)
CC1/2	0.999 (0.909)
CC*	1 (0.976)
Reflections used in refinement	50074 (4854)
Reflections used for R-free	2000 (195)
R-work	0.1655 (0.2289)
R-free	0.1934 (0.2932)
CC(work)	0.964 (0.924)
CC(free)	0.961 (0.855)
Number of non-hydrogen atoms	4062
macromolecules	3577
ligands	21
solvent	475
Protein residues	447
RMS(bonds)	0.012
RMS(angles)	1.16
Ramachandran favored (%)	98.65
Ramachandran allowed (%)	1.35
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.55
Clashscore	3.68
Average B-factor	28.86
Macromolecules (Å^2)	27.73
Ligands (Å^2)	28.77
Solvent (Å^2)	37.38

Chapter 3 | Characterisation of clinically relevant AspH 2OG-binding variants

3.1 Introduction

In 1995 Shawarf et al. identified six members of a family who presented with distinctive ocular abnormalities and facial dysmorphisms; they termed the disease Facial Dysmorphism Lens Dislocation (FDLAB) syndrome, or Traboulsi syndrome.[206] Affected individuals suffer from ocular abnormalities including partial or complete dislocation of the lens leading to severe loss of vision and cysts in the subconjunctival space (subconjunctival spontaneous filtering blebs).[206,207,209,210] Further characterisation of the facial dysmorphisms identified that the patients display dolichocephaly (elongation of the skull), irregular or crowded teeth, poor cheekbone development and malformed lower jaws.[210,212] Three patients also exhibited cardiac abnormalities that were believed to be associated with the disease.[208] In the following twenty-eight years, over thirty further individuals from a range of ethnicities have been diagnosed with FDLAB syndrome; the disease is rare and estimated to affect less than one in every one million live births.[207]

Whole-exome sequencing of FDLAB patients identified *ASPH* as a disease locus.[210] ASPH-knockout mice exhibit phenotypes which are consistent with FDLAB characteristics, including shortened snouts and mild palate defects; note that effects on the ocular system were not reported.[210,186] Genetic sequencing confirmed that FDLAB is distinct from other genetic disorders which cause similar phenotypes, such as Marfan syndrome.[248]

The human *ASPH* gene has 25 exons and alternative splicing of *ASPH* gives rise to multiple isoforms with four characterised proteins; Aspartate/asparagine-β-hydroxylase (AspH), humbug, junction, and junctate.[149] Only two isoforms contain the C-terminal catalytic domain that is responsible for catalysing the hydroxylation of aspartate and asparagine residues in epidermal growth factor-like domains (EGFDs), including full-length AspH (isoforms A (Uniprot Q12797-1) and F (Uniprot Q12797-10).[143]

Genome sequencing has revealed an array of pathogenic mutations in the *ASPH* gene that are linked to FDLAB syndrome (selection of mutations shown in Table 3-1). Most mutations are located

in exons 22-25, and thus have potential implications for the functioning of the oxygenase domain. One pathogenic mutation in a patient of Peruvian origin identified a mutation in exon 2, in the N-terminus of the cytoplasmic domain.[249] Pathogenic mutations in the AspH oxygenase and TPR domains have been predicted to impair hydroxylation or completely eliminate AspH activity, and therefore lack of AspH-catalysed hydroxylation is postulated to be responsible for FDLAB syndrome phenotypes.[207]

Table 3-1. A selection of reported AspH single nucleotide polymorphisms associated with FDLAB syndrome.

Mutation	Implication
R735W[210]	Mutation of 2OG C5 carboxylate binding residue at active site
R735Q[208]	Mutation of 2OG C5 carboxylate binding residue at active site
G434V[250]	Mutation of residue located in hinge-loop region of TPR domain
R688Q[207]	Mutation of 2OG C1 carboxylate binding residue at active site

Analysis of crystal structures suggests that three FDLAB-linked mutations in *ASPH* will likely affect 2-oxoglutarate (2OG) binding residues at the AspH active site, that is R688 and R735 (Figure 3-1).[124] R688 coordinates to the C1 carboxylate of 2OG, and R735 forms a salt bridge to the C5 carboxylate group.[124] Previous mutagenesis studies performed in the Schofield group on the $AspH_{315-758}$ construct established that mutation of these arginine residues to alanine reduced the activity of the enzyme in the case of the R688A variant and eliminated catalytic activity for the R735A variant.[213] Note, these mutations have not been observed in FDLAB patients yet. Substitution of R688 or R735 to Q or W, as observed for the clinical mutations, may affect the AspH active site in a similar way and therefore conclusively link FDLAB genetic profiling to impaired AspH activity. However, there has been little biochemical evidence that unambiguously supports this proposal.

There are currently no therapeutic approaches for management of FDLAB, and the patients often suffer from a substantially reduced quality of life. It is therefore of interest to investigate the

biochemical basis of FDLAB syndrome, with the long-term aim of informing on possible treatments for the disease.

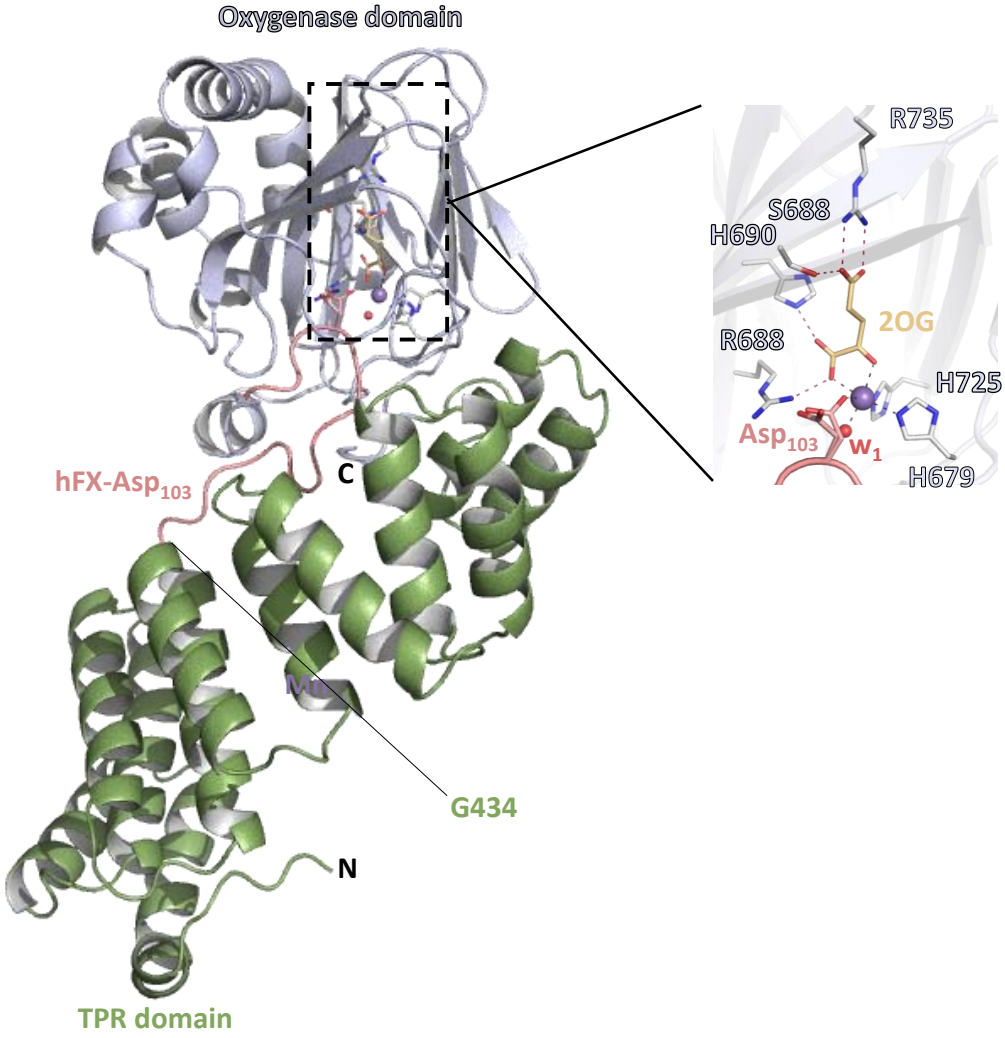


Figure 3-1. View from a crystal structure of wt AspH. G434 is located in the N terminal TPR domain (green). The active site is situated in the oxygenase domain (blue-white), and a close-up view of the active site shows the position of key 2OG-binding residues. 2OG (yellow) shown coordinating at active site (pink dashes) to secondary sphere residues.

3.2 Aims of the work described in this chapter

The work presented in this chapter aimed to (i) generate plasmids encoding for the FDLAB-associated R688Q, R735Q, R735W and G434V AspH variants, (ii) characterise the resultant R688Q, R735Q, R735W and G434V AspH variants by kinetic and crystallographic studies, and (iii) investigate whether the activity of FDLAB associated AspH variants can be modulated by the use of cosubstrate analogues.

3.3 Results and discussion

3.3.1 Purification of clinically-relevant AspH variants

Plasmids encoding for the truncated $AspH_{315-758}$ R688Q, R735Q, R735W and G434V variants were generated by site-directed mutagenesis from the pET28a(+) plasmid encoding for wild-type $AspH_{315-758}$ (henceforth referred to as "wt AspH"). Recombinant protein production and purification was performed as described in Chapter 2; mass spectrometry was used to confirm the masses of the purified AspH variants (Figure 3-2 and Table 3-2). Assays and structural studies on the R688Q, R735W and G434V variants used the His_6-tagged proteins. The determination of kinetic parameters for the R735Q variant were measured using His_6-tagged protein, while all other structural and activity assays used the His_6-tag-cleaved protein.

The pET28a(+) plasmid contains a thrombin protease recognition site (LVPR/GS; cleavage between R and G) located between the sequence encoding for the His_6 tag and the AspH gene sequence, to enable removal of the His_6 tag during purification. Another region of the non-protein coding multiple cloning site sequence contains a sequence encoding for "GRGS". This sequence can potentially also be recognised by thrombin as a secondary cleavage site. Therefore, removal of the His_6 tag by thrombin can lead to the formation of two major products, which differ in length by fifteen residues on the AspH N-terminal. Indeed, two His_6-tag cleaved products were observed by MS analysis following purification of the R735Q variant (Table 3-2). In cases where the His_6 tag was not removed during the purification procedure, that is for the R688Q, R735W and G434V variants, partial gluconylation of the His_6-tag in *E. coli* results in the observation of a second protein species of mass +178 Da higher than the expected AspH variant masses.[251] In future, it would be advisable to optimise the His_6-tag cleavage process and/or purification process to avoid producing protein with a mixture of N-terminal modifications. However, as the N-terminal of AspH is far from the active site, it was assumed that the mixture of protein products would not affect the activity or characterisation of the AspH variants.

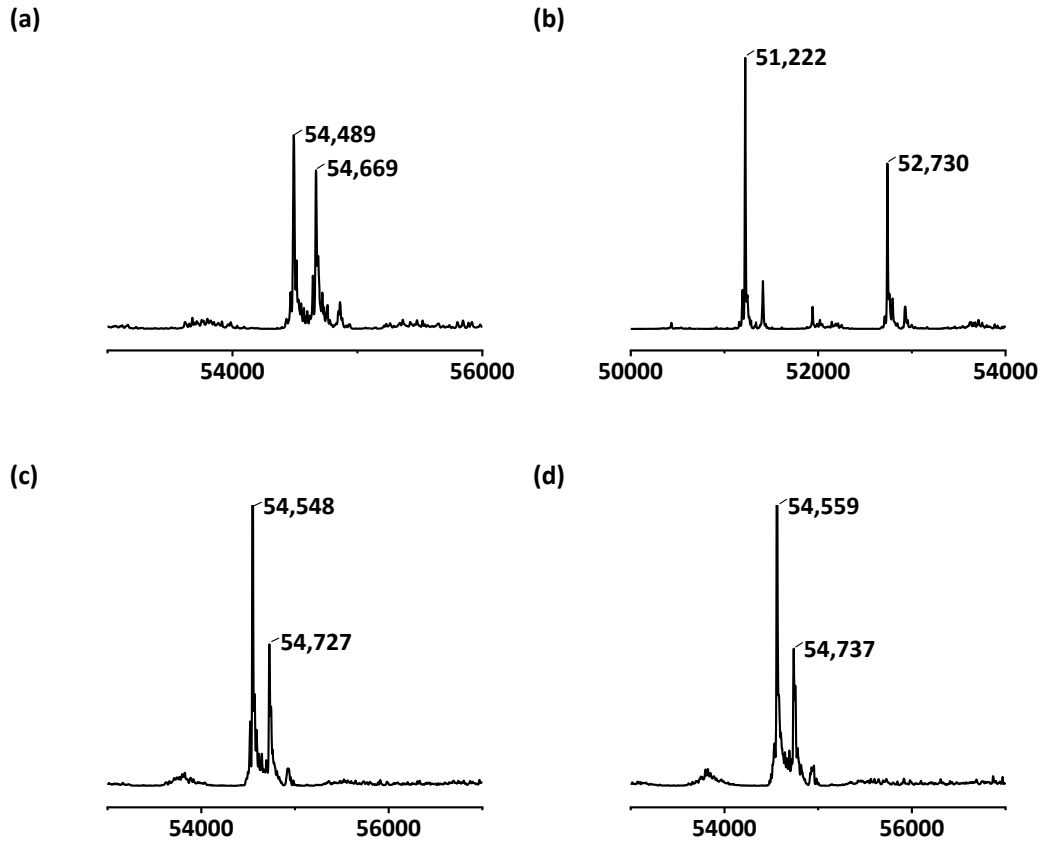


Figure 3-2. MS spectra of the R688Q (a), R735Q (b), R735W (c) and G434V (d) variants. Calculated masses are shown in Table 3-2.

Table 3-2. The calculated molecular masses of AspH variants produced for characterization and the masses observed by MS-analysis.

AspH variant	Calculated mass (Da)	Observed masses (Da)
R735Q	52,740	52,740 51,222
His_6-R688Q	54,491	54,489 54,669 (Gluconylated)
His_6-R735W	54,548	54,549 54,727 (Gluconylated)
His_6-G434V	54,561	54,559 54,737 (Gluconylated)

3.3.2 Characterization of clinically relevant 2OG-binding AspH variants

3.3.2.1 Initial activity assays

Measurement of AspH variant activity was monitored in real-time by SPE-MS assays, as described in Chapter 2. In brief, a mixture containing L-ascorbic acid (LAA), 2OG, the hFX-Asp_{103} substrate, ferrous ammonium iron sulphate (FAS) and an AspH variant was used in the assay. The AspH-catalyzed substrate hydroxylation was monitored via injecting assay aliquots into a SPE-MS system coupled to a RapidFire sample handler.

Initial assays revealed that the R688Q, R735Q and G434V variants retained varying levels of catalytic activity under conditions reported for wt AspH (that is in 25 mM HEPES buffer pH 7.5, with 50 mM NaCl). Thus, the buffer conditions were subsequently optimised for each variant individually. Buffer pH values were incrementally increased in 0.5 pH steps from pH 6.0 to pH 8.0, and each buffer was tested both in the addition and in the absence of 50 mM NaCl.

The buffer optimisation process revealed that the G434V AspH variant catalysed substrate hydroxylation at a rate similar to that of the wt AspH under the originally reported conditions, that is: 25 mM HEPES buffer, pH 7.5 with 50 mM NaCl (Table 3-3; Figure 3-3). By contrast to G434V AspH, both the R688Q and R735Q AspH variants were apparently most active in more acidic buffer (that is, 25 mM MES, pH 6.0 buffer) in the absence of salt (Table 3-3; Figure 3-3). The R735W AspH variant was inactive under all the tested conditions and therefore not subjected to further kinetic analyses.

Table 3-3. Optimised conditions used for activity measurement and kinetic parameter determination of wt AspH and the G434V, R688Q and R735Q variants.

Enzyme	Buffer composition	Concentration of reactant (μM)				
		LAA	2OG	hFX-Asp$_{103}$	FAS	Enzyme
wt AspH	25 mM HEPES pH 7.5, 50 mM NaCl	100	20	4	20	0.1
G434V	25 mM HEPES pH 7.5, 50 mM NaCl	100	20	4	20	0.1
R688Q	25 mM MES pH 6.0	400	30	4	100	0.2
R735Q	25 mM MES pH 6.0	400	2000	4	20	0.2

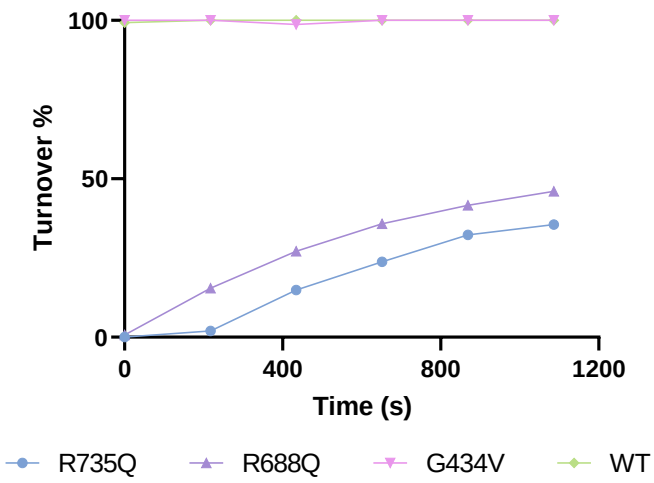


Figure 3-3. Time course of hydroxylation of hFX-Asp$_{103}$ substrate catalysed by wt AspH and the G434V, R735Q and R688Q variants. Turnover is a measure of the percentage of substrate converted to hydroxylated product, as monitored by SPE-MS. The assay conditions are outlined in Table 3-3.

3.3.2.2 Determination of kinetic parameters

Having established the conditions for optimal catalysis for the G434V, R688Q and R735Q AspH variants, subsequent studies focused on determining the kinetic parameters of the variants using SPE-MS. The Michaelis-Menten plots for each cosubstrate/cofactor for each AspH variant that were used to calculate the kinetic parameters are shown in Figure 3-4 and kinetic parameters are summarised in Table 3-4.

Table 3-4. The kinetic parameters determined for wt AspH, and the G434V, R735Q, R688Q clinically relevant variants. The turnover of substrates was measured by SPE-MS. K_m, V_{max} and k_{cat} were calculated using the GraphPad Prism function for Michaelis Menten kinetics. Assay conditions are outlined in Table 3-3, except for the stated cofactor, where the concentration was varied.

Reagent	AspH variant	k_{cat}^{app} (s^{-1})	K_m^{app} (μM)	$\left(\frac{k_{cat}}{K_m}\right)^{app}$ ($mM^{-1}s^{-1}$)
2OG	WT	0.31 ± 0.03	1.1 ± 0.4	282 ± 106
	G434V	0.59 ± 0.037	0.98 ± 0.26	604 ± 95
	R735Q	0.0065 ± 0.00069	69.6 ± 34	0.093 ± 0.03
	R688Q	0.014 ± 0.00089	0.84 ± 0.26	16.2 ± 3.0
Fe (II)	WT	0.24 ± 0.04	7.2 ± 2.6	33.3 ± 13
	G434V	0.54 ± 0.056	2.11 ± 1.0	256 ± 76
	R735Q	0.014 ± 0.0045	63.3 ± 40	0.23 ± 0.10
	R688Q	0.015 ± 0.0013	16.8 ± 4.8	0.91 ± 0.16
Fe (II) with LAA	WT	0.27 ± 0.04	4.3 ± 1.6	62.8 ± 25
	G434V	0.57 ± 0.056	1.5 ± 0.56	378 ± 84
	R735Q	0.0098 ± 0.0018	12.2 ± 7.3	0.80 ± 0.29
	R688Q	0.015 ± 0.0011	22.8 ± 5.2	0.66 ± 0.10

The presence of LAA in the assay mixtures did not affect the rate of reaction for any of the three tested AspH variants, consistent with reported results for wt AspH and the Fe(II)-binding H679A and H725A AspH variants (Chapter 2). The kinetic parameters were determined with respect to 2OG, and Fe(II) in the presence and absence of LAA; LAA is reported to increase assay robustness although it is not required for efficient AspH catalysis.[228]

Surprisingly, the k_{cat} values for 2OG and Fe(II) for the G434V AspH variant were almost twice that of wt AspH k_{cat} value. By contrast to the G434V variant, the R688Q and R735Q variants had k_{cat} values of ~10- and 20-fold less than for wt AspH.

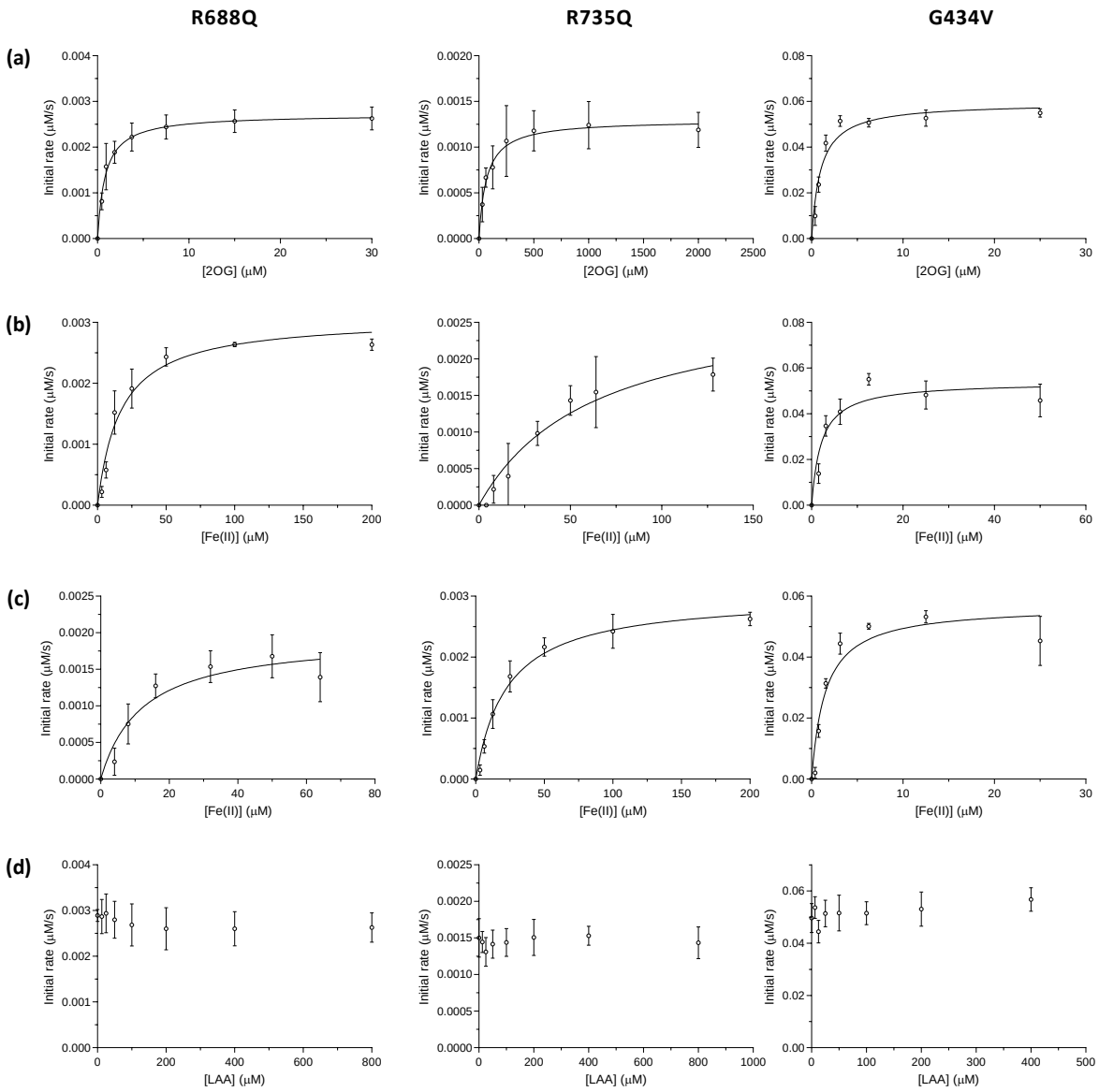


Figure 3-4. Michaelis-Menten plots for the determination of the kinetic parameters of the R688Q, R735Q and G434V AspH variants, as measured by SPE-MS assays. (a) 2OG, (b) FAS, (c) FAS in the presence of LAA, (d) LAA. Assay conditions are outlined in Table 3-3, except for the stated cofactor, were the concentration was varied. Data were analysed using GraphPad prism.

The K_m values for the G434V AspH variant with respect to 2OG and Fe(II) in the presence or absence of LAA were within experimental error to the wt AspH values. The K_m values determined for the R735Q variant with respect to 2OG and Fe(II) were all substantially higher than wt AspH, indicating a significant weakening in the ability of R735Q to bind 2OG or Fe(II). Interestingly, the K_m value for 2OG for the R688Q AspH variant was within one standard error to the wt AspH 2OG K_m value, whereas the K_m value for the R688Q variant with respect to Fe(II) was over two-fold and five-fold

higher than that of wt AspH in the absence and presence of LAA, respectively. This observation suggests that the R688Q substitution has impaired the ability of R688Q AspH to ligate Fe(II), although R688 does not directly coordinate to Fe(II).

For the R735Q variant, the K_m value for Fe(II) in the presence of LAA was lower compared to the K_m value in the absence of LAA, consistent with the results for wt AspH and the G434V variant. By contrast, the K_m values for the R688Q AspH variant with respect to Fe(II) were higher in the presence of LAA than without LAA.

The values of k_{cat}/K_m for the G434V variant are higher than that of wt AspH indicating a surprising level of efficiency of catalysis by the G434V variant. Contrastingly, the k_{cat}/K_m values for the R735Q and R688Q variants with are one or two orders of magnitude lower than for wt AspH. These data reflect the overall inefficient turnover of the R735Q and R735Q variants, likely as a consequence of the aforementioned impaired Fe(II) and 2OG binding.

Attempts to obtain the kinetic parameters with respect to hFX-Asp_{103} and O_2 proved to be unsuccessful, as previously discussed in Chapter 2. The SPE-MS method for analysing peptide hydroxylation was unsuitable for use at very low and very high substrate concentrations. At low concentrations, the peptide was not detected, and at high concentrations, the peptide precipitated in the machine and clogged the tubing. Manual methods of measuring substrate turnover under varying O_2 atmospheres were not optimised for the Fe(II)-binding variants, H679A and H725A, as discussed in Chapter 2, and therefore the trials were not extended to the study of the clinically relevant 2OG-binding variants, R688Q and R735Q, and the G434V variant.

3.3.2.3 Structural studies

Optimisation of the crystallisation methods

The FDLAB-associated AspH variants were subjected to crystallographic studies to investigate the structural basis of the observed variations in the kinetic parameters. Co-crystallisation solutions

containing final concentrations of 2 mM 2OG, 1 mM Mn(II) as a substitute for the natural Fe(II) cofactor, 800 µM hFX-Asp_{103} substrate, and 374 µM of one of the R688Q, R735Q, R735W and G434V variants, were mixed with various crystallisation precipitants in a 96-well broad screen (Pact Premier, Hampton). Needle-like crystals of the G434V, R688Q, and R735Q variants grew in numerous conditions following incubation at 4°C. No crystals of the R735W AspH variant were observed when using commercial broad crystallisation screens. Crystals of the R688Q, R735Q and G434V variants were harvested into liquid N_2 and diffraction data were collected using MX beamline I03 at Diamond Light Source.

In these initial crystallisation trials, only R735Q crystals diffracted to a high-resolution (1.9 Å), which is the similar to the resolutions typically achieved with wt AspH crystals. Structural studies described in Chapter 2 performed on AspH Fe(II)-binding variants, H679A and H725A, discovered that a four-fold excess of 2OG relative to the enzyme concentration was insufficient to give a 2OG-bound complex at the active site. This was likely as a consequence of the impaired 2OG-binding capacities of the H679A and H725A variants. Similarly, the active site of the first solved structure of the R735Q variant did not contain 2OG. Crystals of the G434V and R688Q variants diffracted to a resolution greater than 3 Å and these data could not be used to resolve detailed information about the G434V and R688Q variants.

Crystallisation experiments with the R688Q, R735Q and R735W AspH variants were then repeated with the optimisation procedure described in Chapter 2 to improve the crystal quality and in an attempt to detect a metal ion and 2OG in the variant active sites. Thus, the concentration of 2OG in the co-crystallisation solution was increased to 10 mM, crystal growth was assisted by cross-seeding with wt AspH crystal seeds, and variant crystals were soaked with 2OG solutions (comprising 10 mM 2OG in 30%$_{v/v}$ PEG 400 as a cryoprotectant and 70%$_{v/v}$ precipitant solution) immediately before harvesting. Additionally, in pursuit of growing larger crystals, the crystallisation method was changed from using 300 nL drops set up with a Mosquito liquid handling robot to 3 µL

sitting-drop or hanging-drop wells. Crystals for the R688Q, R735Q and R735W AspH variants that grew under these optimised conditions were larger and diffracted to a higher resolution.

The G434V variant was crystallised with the addition of wt AspH crystal seeds in an attempt to improve diffraction quality; however, the co-crystallisation conditions were not modified and the additional soaking step was not performed to reflect that the kinetic profile of G434V AspH was determined to be similar to that of wt AspH (Table 3-4). Nevertheless, the quality of diffraction from G434V variant crystals remained low (>2.4 Å) in comparison to other successful wt AspH or variant crystals. The G434V substitution likely increases flexibility in the hinge-loop region of AspH and the overall structure is less rigid in crystal form, resulting in lower resolution diffraction.

Crystal structure of the R735Q variant

As outlined previously, initial broad-screen crystallisation trials only resulted in high-quality crystals of the R735Q variant. The R735Q variant crystallised in the $P2_12_12_1$ space group and structures were solved by molecular replacement using a reported wt AspH structure (PDB 6YYW) as a search model (Figure 3-5). The overall fold of the R735Q variant is very similar to that of wt AspH (RMSD = 0.329 Å). At the active site, electron density for Q735 was observed, confirming the presence of the substituted residue (Figure 3-5a).

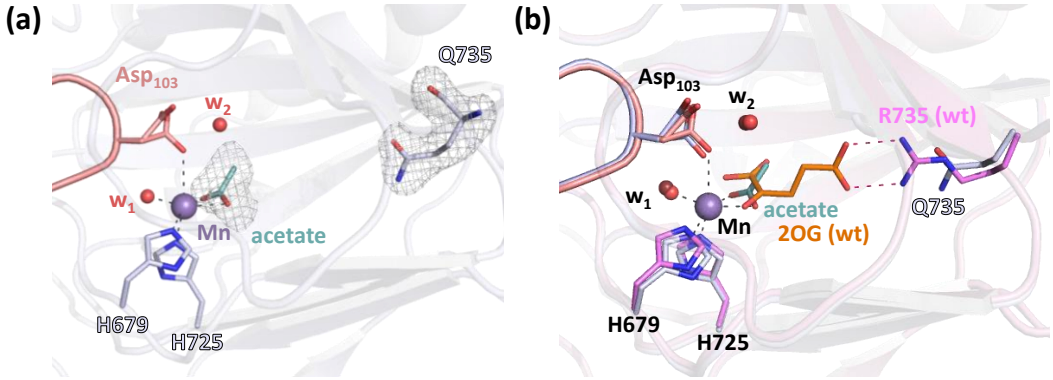


Figure 3-5. View from a crystal structure of the R735Q variant in complex with Mn, acetate and hFX-Asp$_{103}$. (a) View of the R735Q active site with the polder omit electron density carved to 3σ around Q735 and the acetate ion. (b) Superimposition of the R735Q active site in complex with acetate with the wt AspH active site in complex with 2OG. R735Q residues coloured as in (a). Wt AspH residues: pink, wt 2OG: orange, wt water: dark red, wt hFX-Asp$_{103}$: lilac.

No density corresponding to 2OG was observed at the active site. Instead, the electron density near the Mn was refined with an acetate ion, coordinating in a bidentate manner to the Mn. The source of the acetate is unknown; it could originate from the commercially-sourced substrate peptide which was synthesized by solid phase peptide synthesis. This absence of 2OG at the active site in crystals derived from low concentrations of 2OG in co-crystallisation solutions is consistent with the different kinetic parameters for the R735Q variant compared to that of wt AspH (Table 3-4). The elevated K_m indicates of the R735Q AspH variant compared to wt AspH indicate that its ability to bind 2OG has been impaired.

Comparison of the active site of wt AspH in complex with 2OG and the R735Q AspH active site also shows that in the R735Q active site the distance between the side chain of residue 735 to the metal centre has increased by 1.4 Å relative to wt AspH as a result of the R to Q substitution (Figure 3-5b). Therefore, it is perhaps unsurprising that density for 2OG was not observed at the active site of the R735Q variant as 2OG is likely unable to form stable interactions with Q735.

Crystal structures of the R735Q and R735W variants in complex with 2OG

Optimisation of the crystallisation process proved to be successful in the pursuit of 2OG-bound structures of the R735Q and R735W variants (Figure 3-6). Interestingly, the active site structures of the R735Q and R735W variants in complex with Mn, 2OG and hFX were found to be remarkably similar (Figure 3-6a and b). Electron density around either Q735 or W735 confirmed the presence of the substitutions at position 735 and pleasingly, electron density that could be modelled as 2OG was observed. (Figure 3-6a and b). In the structures of both the R735Q and R735W variants, Mn is coordinated by H679, H725, two water molecules (w_1 and w_2), and bidentate coordination from 2OG. The side chain carboxylate group on Asp_{103} of the substrate was modelled in a single conformation with Cβ of the Asp_{103} side chain angled towards the Mn centre. This is in contrast to the R735Q:acetateMn:hFX-Asp_{103} structure (Figure 3-5a), in which the substrate Asp_{103} carboxylate group was observed in two conformations, one of which was involved in Mn coordination.

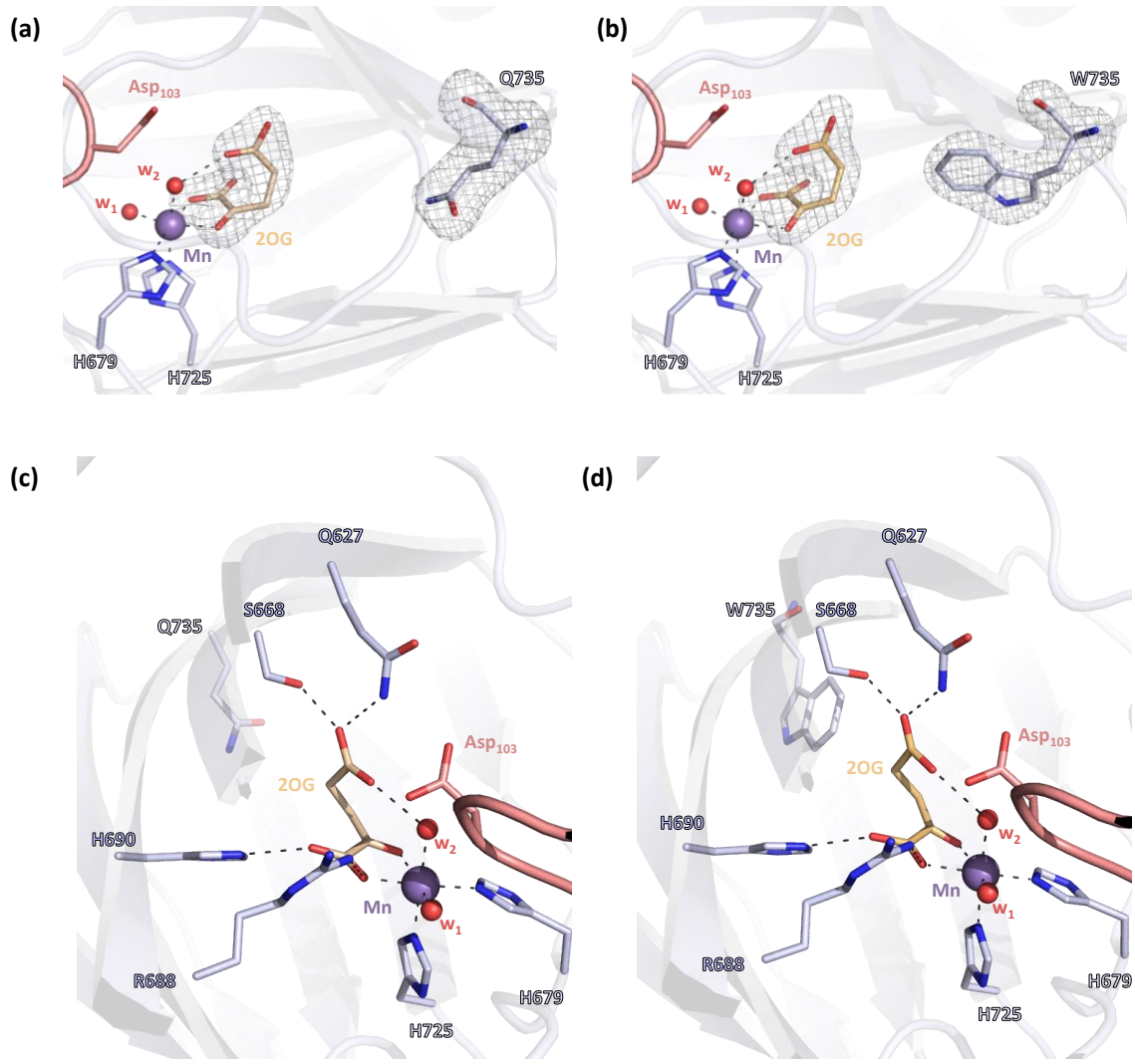


Figure 3-6. Views from crystal structures of the R735Q and R735W AspH variant active sites in complex with Mn, 2OG and hFX-Asp$_{103}$. (a) The active site of the R735Q variant and (b) the active site of the R735W variant. Polder omit electron density carved to 3σ in (a) and (b) confirms the R to Q/W mutation at residue 735, and the unusual conformation of 2OG. Alternative views of the R735Q active site (c) and R735W active site (d) to show possible secondary sphere residues that are positioned to interact with 2OG.

The positions and conformations of the 2OG C1 carboxylate and C2 keto in the R735Q and R735W variant complex structures are similar to those observed in the wt AspH complex structure (Figure 3-1). One oxygen of the 2OG C1 carboxylate is positioned to interact with H690, the other C1 carboxylate oxygen forms electrostatic interactions with R688 and Mn, while the 2OG C2 keto oxygen also coordinates to Mn (Figure 3-6c and d). By contrast to wt AspH, the 2OG C5 carboxylate group adopts an unprecedented conformation in the R735Q and R735W variant active sites (Figure 3-6). This is likely due to the loss of a stabilising salt-bridge interaction from R735. One of the C5 carboxylate oxygens of 2OG is within hydrogen bonding distance to the side chain hydroxy group

of S688 (2.6 Å in R735Q, 2.6 Å in R735W) and to the side chain of Q627 (3.3 Å in R735Q, 2.8 Å in R735W). The other 2OG C5 carboxylate oxygen is positioned 2.7 Å from w_2, one of the Mn-coordinating water molecules.

Superimposing the structures of the R735Q or R735W variants with the structure of wt AspH reveals that neither of the mutations significantly affected the overall fold of the enzyme (RMSD = 0.344 Å, RMSD = 0.366 Å respectively). However, comparison of the active sites shows some movement of residues around to accommodate the unusual 2OG conformation (Figure 3-7a and b). For both of the R735Q and R735W variants, Q627 has shifted further away from 2OG and S668 has moved closer to residue 735. The position of the side chain of the substrate Asp_{103} has also moved slightly away from the metal centre. A non-metal bound water molecule, w_3, that is normally observed in the wt AspH active site, is absent in the R735Q and R735W structures. One of the oxygens of the 2OG C5 carboxylate instead occupies this space (Figure 3-7)

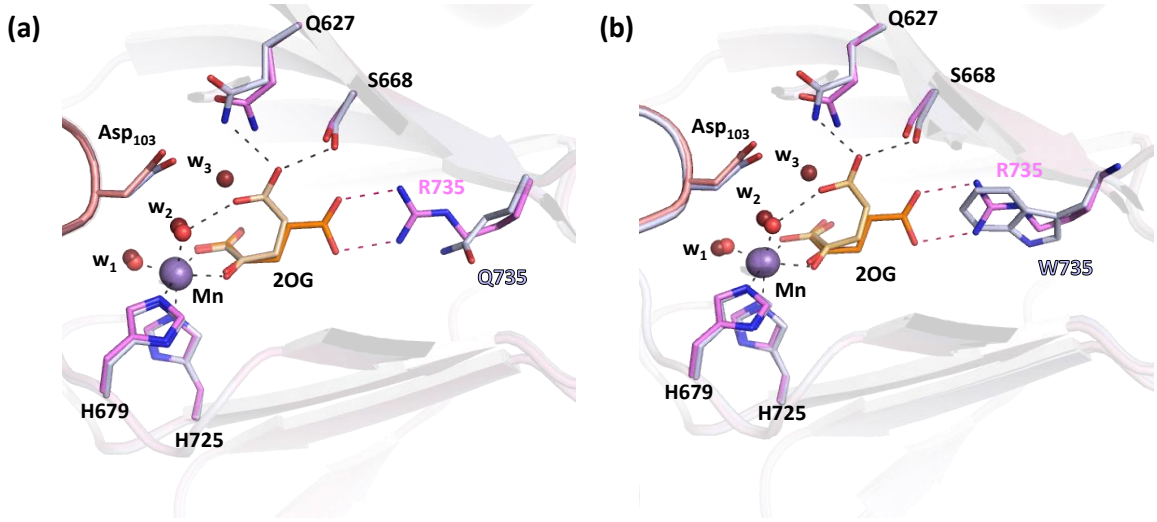


Figure 3-7. Superimposition of the active site structure of wt AspH with the R735Q (a) and R735W (b) variant active sites. (a) R735Q protein residues: blue-white, R735Q 2OG: yellow, R735Q hFX-Asp_{103}: salmon. Wt AspH protein residues: pink, wt AspH 2OG: orange, wt AspH hFX-Asp_{103}: lilac. (b) R735W protein residues: blue-white, R735W 2OG: yellow, R735W hFX-Asp_{103}: salmon. Wt AspH protein residues: pink, wt AspH 2OG: orange, wt AspH hFX-Asp_{103}: lilac.

It is interesting to note that that both the R735W and R735Q variants crystallised with 2OG in the same unusual conformation, despite the difference in the nature of the W and Q side chains, and the fact that the R735W variant is completely catalytically inactive whereas the R735Q variant

conserves a substantial level of activity. It should be noted that the conformation of 2OG observed *in crystallo* is likely to be the most energetically favourable conformation for crystallisation and not necessarily an accurate representation of how 2OG is complexed in a catalytically active AspH in solution. Hydrogen bonding between one of the 2OG C5 carboxylate oxygens to w_2, which ligates the Mn ion, could result in a stable, but inactive, enzyme complex. The interaction with the 2OG C5 carboxylate may hinder w_2 leaving the active site, thus blocking O_2 binding to the metal centre. In the case of the R735Q variant (and wt AspH), it is possible that 2OG can move freely between this (likely) inactive *in crystallo* conformation and other conformations which allow for movement of w_2 out of the active site, thus creating a vacant coordination space at the metal centre and allowing O_2 to bind. This is consistent with the elevated 2OG K_m value for R735Q compared to the 2OG K_m value of wt AspH (Table 3-4). In the case of the R735W variant, the increased steric bulk of the tryptophan group may not allow for 2OG to adopt a catalytically productive conformation, so 2OG at the active site may be trapped in an inactive conformation.

Crystal structure of the R688Q variant

Although the soaking of additional 2OG into R735Q and R735W AspH crystals proved to be effective in obtaining 2OG bound-structures of these variants, the same method was unsuccessful when applied to the R688Q variant. The addition of wt AspH crystal seeds to the crystallisation mixture and using hanging drop trays to increase crystal size significantly improved the diffraction quality, with crystals diffracting to 1.7 Å in the highest resolution shell. However, subsequent molecular replacement and analysis of the active site electron density indicated that no 2OG was present in the R688Q variant structure (Figure 3-8a).

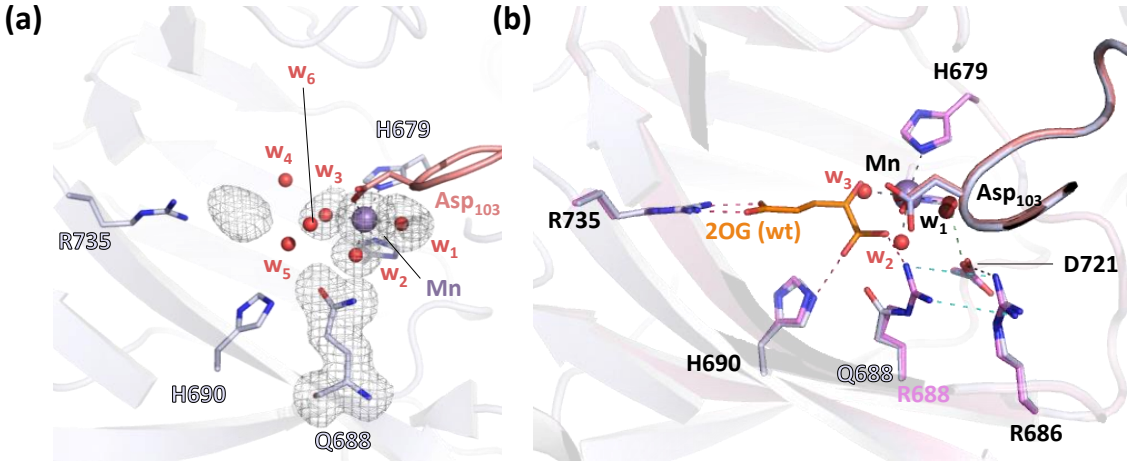


Figure 3-8. View from the active site of the R688Q variant structure. (a) Polder omit electron density carved to 3σ confirms the R to Q mutation and the coordination of three water molecules around Mn. An ambiguous region of electron density in close proximity to R735 has been left unmodelled. (b) Superimposition of R688Q active site with wt AspH active site with 2OG. Waters 4-6 of R688Q have been omitted for clarity. Pink dashes shown interactions only present to 2OG in wt AspH structures. Cyan dashes show π-π stacking interaction between R688 and R686 in wt AspH. R688Q structure coloured as previously in (a). wt AspH protein residues: pink, wt 2OG: orange, wt water: brown-red, wt hFX-Asp_{103}: lilac.

The active site of the R688Q variant revealed Mn coordinated by H679, H725, a water molecule (w_1) and the carboxylate from the substrate Asp_{103}, as typically observed in wt AspH and the R735Q and R735W structures (Figure 3-8a). However, electron density around the 2OG binding site of wt AspH was difficult to resolve. The electron density suggests that the six coordinate environment around the Mn(II) ion is completed by two water molecules (w_2 and w_3 in Figure 3-8a). The location of w_2 positions it within hydrogen bonding distance to Q688 (2.79 Å) and D721 (3.08 Å). Other electron density in the space normally occupied by 2OG was modelled as water molecules, numbered w_4-w_6, or left unmodelled (Figure 3-8a). Electron density next to R735 was left unmodelled as it is too large to represent a single water molecule or other anions in the precipitant mixture, and the shape of this density is unlikely to correspond to other small organic molecules that could be present, e.g. acetate or formate.

Superimposing the structure of the R688Q active site without 2OG and wt AspH with 2OG gives further insight into the effect of the substitution on the active site (Figure 3-8b). In the wt AspH structure, 2OG is positioned to interact with R688 and H690 through its C1 carboxylate and S688

and R735 through its C5 carboxylate. R688 normally forms a hydrogen-bonding interaction with the 2OG C1 carboxylate oxygen closest to the metal centre (2.74 Å in wt AspH).

The position of H690 has not changed in the R688Q active site structure compared to wt AspH, and H690 would remain within hydrogen bonding distance to the other C1 carboxylate oxygen atom of 2OG, if 2OG were to bind in R688Q in a comparable orientation. Additionally, the position of R735 has not changed in the R688Q structure compared to the wt AspH structure and would therefore be expected to bind 2OG in a productive conformation. In solution, 2OG may be able to bind at the active site using these interactions only. However, the combined effect of these interactions is apparently insufficient to compensate for the loss of stabilisation from R688 *in crystallo*.

The R688Q substitution also results in loss of a stabilising π-π interaction between R688 and R686 (Figure 3-8b, cyan dashes). In wt AspH, the side chain of R686 is positioned to interact with D721, which forms a hydrogen bond to w_1 at the active site metal centre; D721 was shown to be essential for AspH catalysis and the D721A variant is catalytically inactive.[213,252,] The destabilisation of this secondary sphere interaction would likely have consequences on the stability of Fe(II) binding at the R688Q active site. This instability is likely reflected in the Fe(II) K_m value for the R688Q variant (16.8 μM) which is over twice that of wt AspH (7.2 μM) (Table 3-4).

Another interesting point to note about the R688Q active site is that the substrate Asp_{103} is only observed in a single conformation, by contrast with wt AspH in which two Asp_{103} conformations are frequently observed (Figure 1-10). The Asp_{103} conformations are denoted A, where the Asp_{103} carboxylate faces away from the metal and is likely to be a productive conformation, and B, in which the Asp_{103} carboxylate interacts with the active site metal. In many other structures of wt AspH and AspH variants, the substrate is either observed as a mix of conformers or in conformation A only (e.g. R735Q structure shown in Figure 3-6), whereas in the R688Q structure only conformation B where the carboxylate group coordinates with Mn is observed. This observation is perhaps a consequence of the aforementioned loss in metal binding stability in the R688Q variant,

and consequently the substrate Asp_{103} adopts a conformation to assist in metal binding. This conformation is likely not catalytically productive, which may be reflected the lower turnover rates observed for the R688Q variant.

Given that the 2OG K_m for the R735Q variant is significantly higher than the 2OG K_m for the R688Q variant (69.6 μM and 0.840 μM respectively), it is somewhat surprising that it was not possible to obtain a 2OG-bound R688Q variant structure, whereas 2OG-bound R735Q and R735W AspH variant complex structures could be obtained. However, there are two points to note when assessing the crystal structures of R6688Q and R735Q. Firstly, 2OG at the R735Q active site appears in an apparently distorted conformation that is able to exploit interactions from residues around the active site (S688 and G627), that are not normally involved in 2OG binding. Secondly, the 2OG conformation in the R735Q structure is likely to represent an unproductive arrangement at the active site given that the inactive R735W variant exhibited the same active site complex. Therefore, the observation of an AspH:2OG:Mn:hFX-Asp_{103} complex does not necessarily reflect the catalytic capability of an AspH variant, and so conversely, the absence of this complex *in crystallo* does not mean the R688Q variant is unable to coordinate 2OG in solution.

Crystal structure of the G434V variant

Crystals of the G434V variant diffracted to a resolution of 2.4 Å or greater and therefore it is difficult to resolve significant details about the effect of the G to V substitution on AspH. Superimposing the structure of wt AspH with the G434V structure (Figure 3-9a) suggests that the G to V substitution has not altered the overall conformation of the enzyme observed *in crystallo* (RMSD = 0.241 Å). The hinge loop region, in which G434 is located, is positioned in the same conformation as observed in wt AspH structures. Due to the relatively low resolution of the G434V variant structure, it is not possible to clearly observe the electron density around V434 to confirm the presence of the mutation (Figure 3-9b). However, the mass of the protein (Figure 3-2d) is as

expected for the G434V variant and it is not unusual for side chains to be absent in electron density of lower resolution crystal structures.

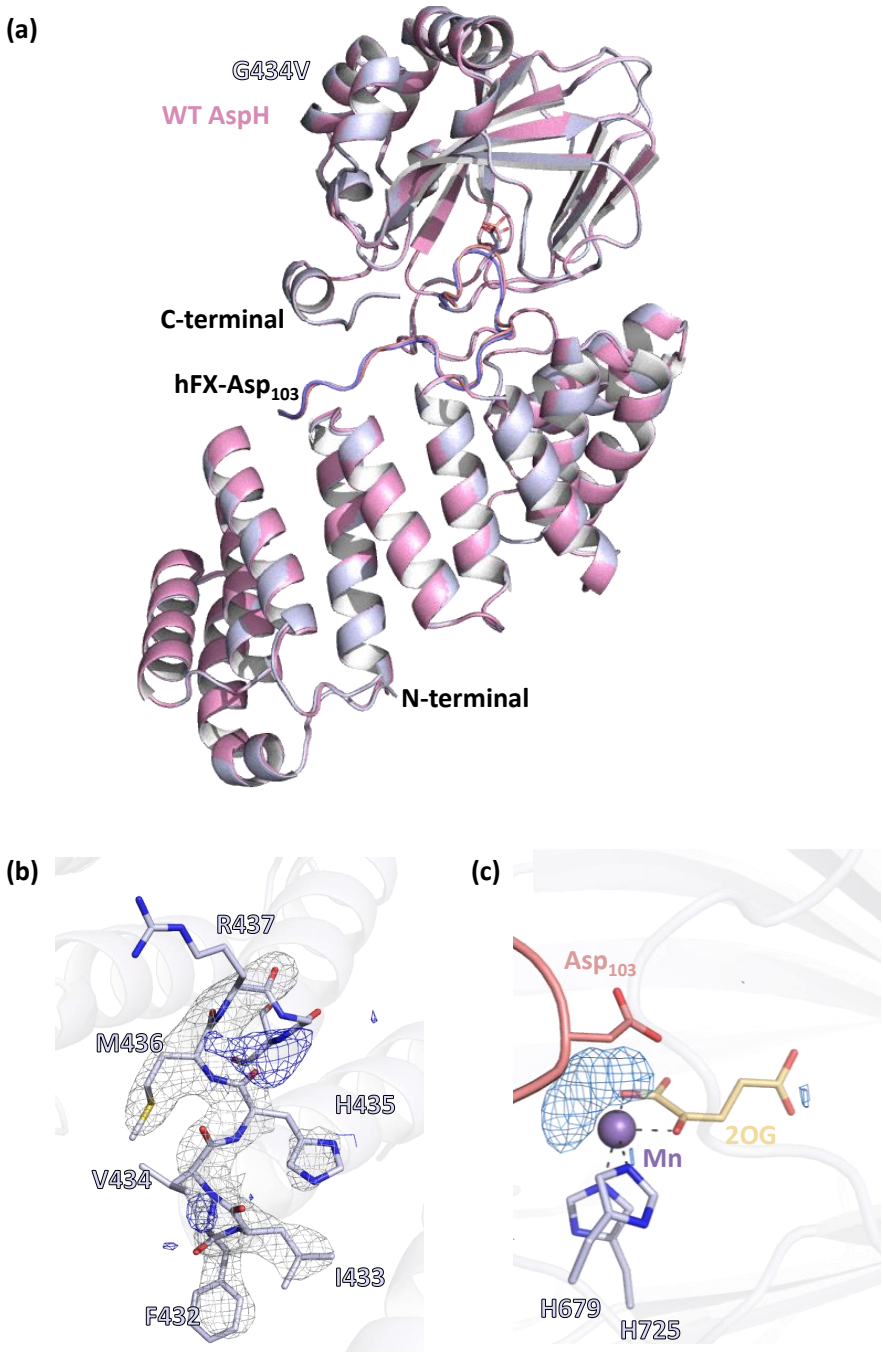


Figure 3-9. View from a crystal structure of the G434V variant. (a) Superimposition of the overall structure of G434V and wt AspH. (b) View from the hinge-loop region of G434V. The grey polder omit electron density is carved to 3σ around the residues and the blue mesh is carved to 3σ around the mF_o-DF_c map. (c) View from the active site of the G434V variant. The mesh is carved to 3σ around the mF_o-DF_c map showing the unmodelled density at the active site. This is likely to be occupied by water molecules or an alternative conformer of Asp_{103}.

Electron density in the hinge-loop region of the G434V variant near H435 has been left unmodelled (Figure 3-9b, blue mF_o-Df_c mesh) as it is unclear as to what exactly this corresponds to. However, this could indicate additionally movement in the hinge-loop region, or alternative conformations of the H435 side chain, both of which suggest the G to V mutation has increase the flexibility of the hinge-loop region of the G434V variant. At the G434V active site, Mn is coordinated by 2OG, H679 and H725 (Figure 3-9c). The substrate Asp_{103} was modelled in a single conformation (conformation A), as the low resolution of the structure meant that electron density surrounding Mn was ambiguous and has been left unmodelled. It is likely that another water molecule (denoted w_1 in previous structures) is present, and the electron density between Mn and Asp_{103} likely represents a second water molecule or conformation B of Asp_{103}.

The presence of 2OG bound at the active site in the G434V variant structure is consistent with the near identical 2OG K_m value (0.983 μM) to that of wt AspH (1.1 μM). Higher 2OG concentrations or an additional soaking step were not required to observe 2OG at the active site likely because the substitution has not affected the ability of the variant to bind 2OG. The elevated k_{cat} of the variant compared to wt AspH is difficult to rationalise from a single crystallographic structure; it is possible that in solution the G to V mutation affects movements in the hinge-loop region of the enzyme, and this causes G434V to be a more efficient catalyst than wt AspH, however these dynamic changes are not evident from a single static structure.

3.3.2.4 The effect of clinically relevant 2OG-binding mutations on AspH

The results described above unambiguously link the clinically observed mutations in the AspH oxygenase domain to reduced AspH hydroxylation activity. The FDLAB-related AspH variants R688Q and R735Q exhibit impaired hydroxylation activity compared to wt AspH and the R735W variant is catalytically inactive under the conditions tested *in vitro*. The impaired activity of the R688Q variant is likely to predominantly originate from a loss of Fe(II) binding stability through a network of active site secondary sphere interactions, as evidenced by the elevated K_m with respect to Fe(II) and observed in the crystal structure of R688Q. The reduced activity of the R735Q variant

is likely to be due to impaired 2OG binding at the active site, as shown in the crystal structure of the R735Q variant in which 2OG adopts an unproductive conformation, and this is reflected in the high K_m value with respect to 2OG.

The missense mutation G434V has also been reported as a pathogenic mutation linked to FDLAB syndrome.[250] In vitro studies showed that the AspH G434V variant exhibited a similar kinetic profile to wt AspH and did not possess impaired hydroxylation activity (Figure 3-3). Therefore, the correlation between FDLAB syndrome and the AspH G434V variant is less conclusive. G434 is located in the hinge-loop region of AspH and therefore G434 has a role in AspH substrate-recognition. It is possible that *in vivo* the G434V mutation could affect AspH activity in a substrate-dependent way. Alternatively, as the patient in which the G434V mutation was identified displayed a broader range of symptoms compared to most FDLAB patients, it is possible that this patient suffered from a different genetic disorder, with similar pathophysiology to FDLAB syndrome.

The activities of the R688Q, R735Q and G434V variants were measured under optimised conditions; however, it is important to note that these conditions do not reflect physiologically relevant cell environments. Therefore *in vivo* the R688Q, R735Q and G434V variants may have undetectable levels of activity, and thus the pathophysiology of FDLAB patients is similar regardless of which AspH mutation is identified in their genome. That is to say, the disease manifests the same in a patient carrying, for example, the R688Q mutation, as a patient carrying the R735W mutation.

As the role of the AspH-catalysed hydroxylation of Asp/Asn residues in EGFDs is still undefined, it is difficult to conclusively link the lack of AspH hydroxylation activity to FDLAB pathophysiology. AspH EGFD substrates include notch, fibrillins, fibulins coagulation factors and latent-transforming growth factor B-binding proteins, which have imperative roles in tissue and organ development. Hydroxylation of EGFD Asp/Asn residues is linked to Ca(II)-binding, which is postulated to stabilise productive conformations of EGFDs.[159] AspH was found to selectively catalyse the hydroxylation of non-canonical EGFDs.[124] Consequently, it is possible that AspH has a role in sensing misfolded

proteins.[124] The absence of AspH-catalysed hydroxylation may therefore lead to an accumulation of misfolded EGFDs, which ultimately manifests in developmental defects.

3.3.3 Characterization of AspH 2OG variant activity with 2OG derivatives

Characterisation of the clinically-relevant 2OG-binding AspH variants R688Q and R735Q revealed that the Arg to Gln substitution impairs the ability of AspH to coordinate 2OG at the active site. Consequently, the catalytic activity of the R688Q and R735Q variants is significantly diminished compared to wt AspH (Table 3-4). Examination of the R735Q variant crystal structure showed that the substitution of R735 to Q735 has significantly increased the distance from the 2OG C5 carboxylate to the Q735 side chain, and consequently 2OG was no longer able to bind at the active site in a productive conformation. A crystal structure of 2OG-bound R688Q has not been attained, consistent with the proposal that the 2OG binding ability of the variant is impaired.

An approach for use in modifying enzyme variant activity, known as the bump-hole method, has been developed.[253] This approach involves modifying the target enzyme (making a hole) to accommodate a cosubstrate with an additional group (the bump) that complements the hole.[253,254] The cosubstrate and enzyme can then be considered a hole-enzyme/bumped-cosubstrate pair. The bump-hole method has been successfully applied in a number of other systems, including the 2OG-dependent histone demethylases (KDM4A) and phytanoyl-CoA 2-hydroxylase (PAHX).[255,253] Pathogenic mutations in the gene encoding PAHX are reported to cause 45% of Adult Refsum Disease, a neurological genetic disorder.[256] Two clinically observed mutations in Adult Refsum Disease, R275Q and R275W, were postulated to affect 2OG C5 carboxylate binding. In *in vitro* assays, modified 2-oxoacids could be used as co-substrates and partially restored the activity of these PAHX mutants.[257]

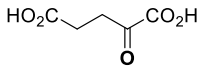
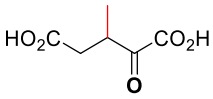
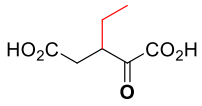
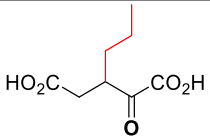
Recent studies have discovered that both naturally occurring and synthetic 2OG derivatives can also be used as cosubstrates for catalysis by 2OG-dependent oxygenases.[258,259] A library of 2OG derivatives bearing substituents at the C3- and/or C4-position were tested against the

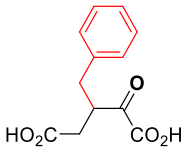
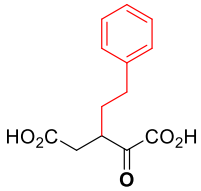
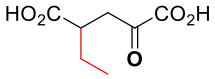
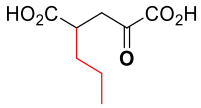
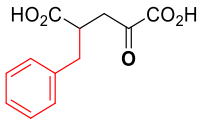
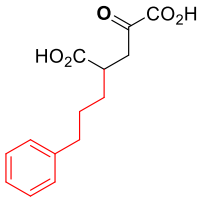
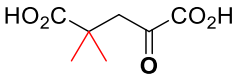
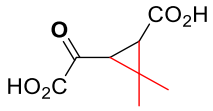
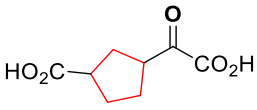
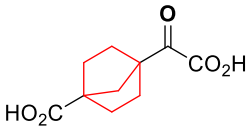
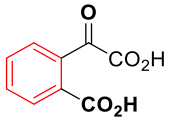
2OG-oxygenases factor inhibiting hypoxia-inducible factor (FIH) and AspH, and some derivatives were shown to selectively enhance enzyme activity.[258,259] The binding of these derivatives can likely exploit the properties of residues surrounding the 2OG-binding site that are not typically thought to be involved in 2OG binding.[259] Thus, it was hypothesised that some of these derivatives may display different activities with the R688Q, R735Q and R735W AspH variants, as the additional functional groups on the derivatives could interact with either the substituted Q residues or other residues in the active site, and stabilise cosubstrate binding in the AspH variants.

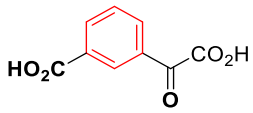
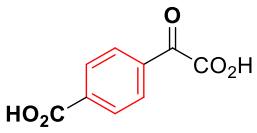
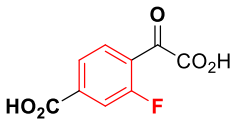
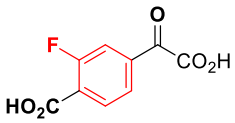
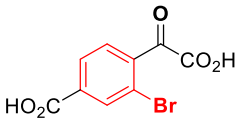
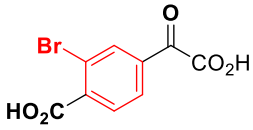
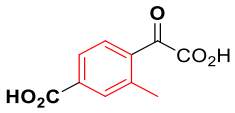
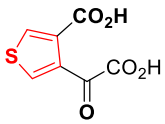
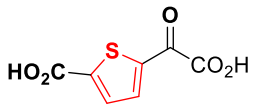
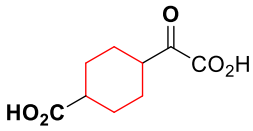
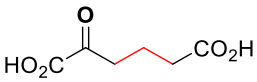
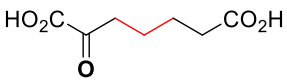
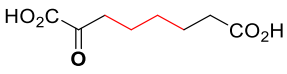
3.3.3.1 2OG derivatives can improve the activity of AspH 2OG binding variants

To assess the hypothesis, assays containing 2OG derivatives in place of 2OG were carried out and substrate turnover catalysed by the AspH R688Q, R735Q and R735W variants was measured by SPE-MS after 30 min. The results are shown in Table 3-5. No activity was restored in the R735W variant; however, the R688Q and R735Q variant activity was selectively enhanced by some, but not all of the tested 2OG derivatives.

Table 3-5. The activity of wt AspH and the R688Q and R735Q variants with synthetic C3/C4-modified 2OG derivatives as cosubstrates in place of 2OG. Assay conditions were as outlined in Table 3-3**.** Data for wt AspH were recorded using an alternative cyclic peptide substrate mimic.[258]

	2OG derivative	Activity		
		WT[259]	R688Q	R735Q
1		**>95%**	**81%**	**7%**
2		**>95%**	**>91%**	0%
3		<1%	0%	0%
4		<1%	0%	0%

5		<1%	**72%**	0%
6		<1%	0%	0%
7		<1%	0%	0%
8		<1%	0%	0%
9		<1%	0%	0%
10		<1%	0%	0%
11		<1%	0%	0%
12		<1%	0%	0%
13		**15%**	0%	**7%**
14		**8%**	0%	**12%**
15		<1%	0%	0%

16		<1%	0%	**10%**
17		**80%**	0%	**30%**
18		**45%**	0%	**37%**
19		**8%**	0%	0%
20		**20%**	0%	**55%**
21		**8%**	0%	0%
22		**5%**	0%	**24%**
23		<1%	0%	0%
24		**15%**	0%	0%
25		0%	0%	**13%**
26		n.d	0%	**21%**
27		n.d	0%	**95%**
28		n.d	0%	**>98%**

The activity of the R688Q variant was modified by 2OG derivatives with substituents on C3 (Table 3-5, compounds **2** and **5**). These derivatives rely on forming alternative binding interactions at the active site, rather than attempting to exploit direct interactions with Q688. Interestingly, the natural product 3-methly 2OG (**2**) significantly improved the level of turnover relative to 2OG, with >95% of substrate turned over compared to 62% turned over with 2OG in the same time.[260] Increasing the length of the substituent carbon chain to give compound **3**, 3-ethyl-2OG, and **4,** 3-propyl 2OG, did not result in any detectable turnover after 30 min; therefore the length of the C3 substituent carbon chain is significant (Table 3-5). Note that lack of turnover does not necessarily reflect lack of binding, as it is possible inactive complexes are formed.

Compound **5** sustained the activity of the R688Q variant to similar levels as 2OG (Table 3-5). This observation is in contrast to wt AspH which is unable to use **5** as a cosubstrate. The aromatic phenyl ring of **5** could be involved in π-π stacking interactions with an aromatic substituent of a side chain in proximity of the active site, such as those of H690 or W711. This type of interaction could stabilise the binding of the 2OG derivative **5** and so promote efficient catalysis. The R688Q substitution may also allow for additional flexibility in cosubstrate binding around the active site due to the presence of a smaller functional group side chain, whereas wt AspH may be unable to accommodate the additional steric bulk of the phenyl ring at the C3 position of **5**.

Derivatives with a longer main carbon chain between the metal binding C1 carboxylate and the terminal carboxylate promoted turnover catalysed by the R735Q variant (Table 3-5 **14**, **16**, **17**, **18**, **20**, **22**, **25**, **26**, **27**, **28**). The most notable effect was observed with compounds **27** and **28**, 2-oxopimelate and 2-oxosuberate, which contain an additional two and three methylene units, respectively, compared to 2OG. Use of **28** resulted in near complete turnover of substrate catalysed by the R735Q variant, in comparison to 7% with 2OG (Table 3-5). By contrast, wt AspH is unable to use **28** as a cosubstrate. Derivative **26** (2-oxoadipate) which contains one more carbon in the main chain compared to 2OG, was also an effective cosubstrate for the R735Q variant,

although less so than **27** or **28** (Table 3-5). This observation suggests that the increased C3 chain length allows the derivative to form alternative interactions at the active site through the C6, C7 or C8 (**26**, **27** or **28**) carboxylate groups, with the optimum interaction distance achieved with **28**, though it could be of interest to test even longer 2OG derivatives. Note that **26** is also a human metabolite and present in cells; it has been previously shown to act as an alternative cosubstrates for some 2OG oxygenases.[257]

The 2OG derivatives bearing an aromatic core, that is **16, 17, 18, 20 a**nd **22,** also manifested improved turnover compared to 2OG, for catalysis by the R735Q variant. The positioning of the carboxylate on the ring is important as turnover was not observed with the regioisomer of **17**, that is **15**. The position of ring substituents is also significant as **21**, a regioisomer of **20**, was not a cosubstrate. Similarly no turnover using **19**, a regioisomer of **18** was detected. Compound **17** was reported to be a good cosubstrate for wt AspH, resulting in conversion of 80% of substrate to product, whereas addition of a bromo-group in **21** gave 20% turnover.[258] The inverse was true for the R735Q variant, with use of **17** resulting in 30% turnover, whereas use of **21** with an additional bromo-substitution resulted in 55% product.

3.3.3.2 Structural studies of 2OG derivatives with the R735Q and R688Q variants

To investigate the structural basis for the activities of the R688Q and R735Q variants with 2OG derivatives, the variants were co-crystallised with the derivative that most improved activity (Table 3-5), in the presence of Mn(II) and hFX-Asp_{103} substrate. Thus, R688Q was co-crystallised with **2** and R735Q was co-crystallised with **28**. To enable detection of the 2OG derivatives in the crystal structures it was necessary to perform the optimised crystallisation procedure with seeding and soaking steps, as described in section 3.3.2.3 for 2OG.

Co-crystallisation of R688Q with **5** and R735Q with **17** was also attempted to further understand the effect of introducing a bulkier, aromatic side chain (**5**) or aromatic core (**17**). However, despite numerous soaking efforts, no density for the derivatives **5** or **17** could be observed at the active

site of the structures. Therefore, it is difficult to attribute specific active site interactions to the modulated activity of the variants with these derivatives.

Structure of the R688Q variant with 3-methyl 2OG (Compound 2)

A crystal structure of the active site R688Q:Mn:**2**:hFX-Asp_{103} is shown in Figure 3-10. The structure reveals Mn ligated by protein residues H679 and H725 at the active site, with further coordination from the substrate Asp_{103}, a water molecule and compound **2.** The C1 carboxylate group of **2** is positioned to interact with H690 and S668, while the C5 carboxylate is held in place by means of a salt bridge to R735. A hydrophobic pocket formed by V727 and V676 houses the 3-methyl group of **2**. A racemic mixture of **2** was used for crystallisation and soaking, however the density appears to suggest that only the (3*R*)-enantiomer, at least predominantly, is bound at the active site.

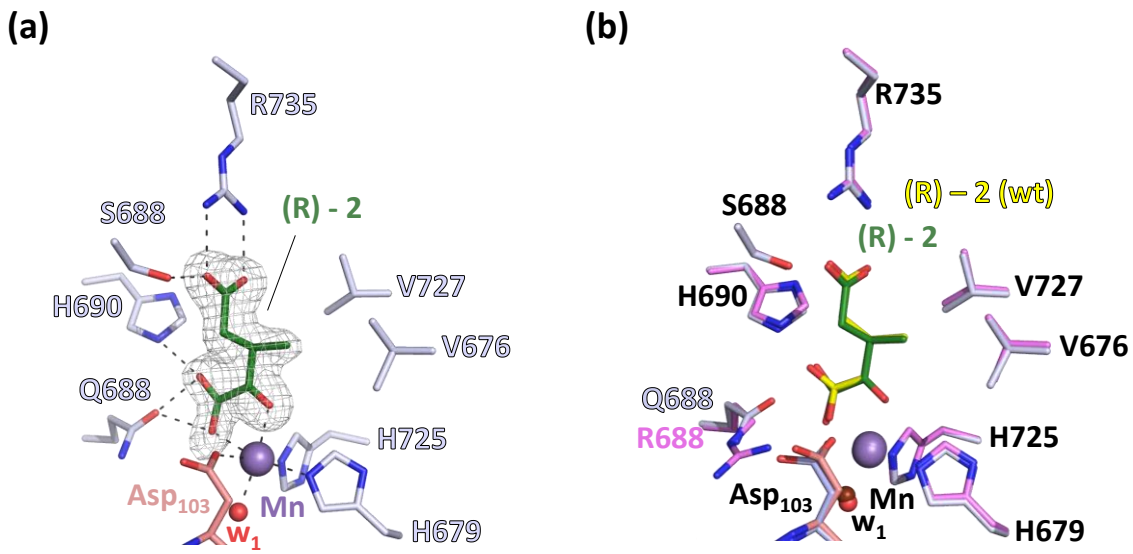


Figure 3-10. View from a crystal structure of the R688Q variant in complex with derivative 2. (a) Polder omit electron density carved to 4σ around **2**. V727 and V676 form hydrophobic pocket that is likely to assist in binding at active site. (b) Overlay with wt AspH (*R*)-**2** structure (PDB 6YYX). R688Q colours as in (a). Wt AspH colours: wt protein: pink, hFX-Asp_{103}: lilac, water: brown, (R)-**2**: yellow.

Overlaying the structure of R688Q:Mn:**2**:hFX-Asp_{103} with the equivalent wt AspH structure shows that the conformation of **2** is identical to that observed in the wt AspH active site, where only the (3*R*)-enantiomer is also present (Figure 3-10b). The additional hydrophobic interactions not present in 2OG from the 3-methyl group of **2** may stabilise binding of **2** to the active site and as

well as subsequently formed intermediates, facilitating a more efficient oxidative decarboxylation step in the mechanism.

Structure of the R735Q variant with 2-oxosuberic acid (Compound 28)

The active site structure of R735Q:Mn(II):**28**:hFX-Asp_{103} is shown in Figure 3-11a. Electron density clearly shows the presence of **28** at the active site. Mn is coordinated to H679, H725, two water molecules and bidentately by **28**. Compound **28** is positioned to interact with R688 and H690 through one of its C1 carboxylate oxygens, in a manner identical to 2OG as observed in wt AspH structures. The positioning of C1, C2 and C3 of **28** overlay directly with the position of C1, C2 and C3 of 2OG in the wild type AspH structure (Figure 3-11b). However, in the R735Q active site, the carbon chain of **28** extends into the vacant pocket created by the R to Q mutation. The C8 carboxylate of **28** is positioned to interact with the S668 side chain (2.53 Å) which is positioned at a shorter distance than from the C5 carboxylate in 2OG in the wt AspH (2.9 Å). The shortened distance from the terminal carboxylate group to the S688 side chain suggests that this is the most influential interaction in stabilising the C8 carboxylate at the active site.

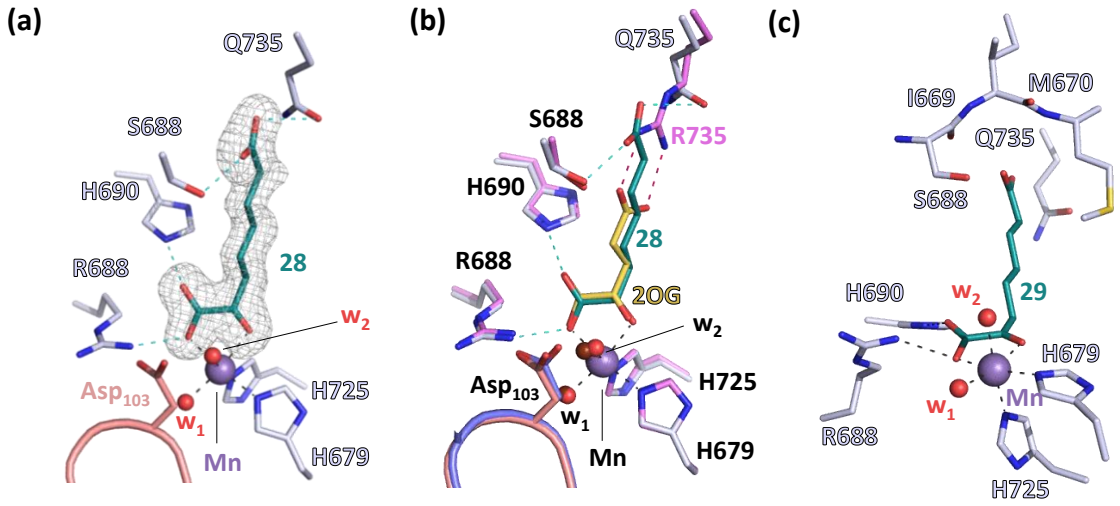


Figure 3-11. View from a crystal structure of the R735Q variant in complex with 2OG derivative 29. (a) The polder omit electron density is carved to 4σ around **28** and shows clear density for **28** bound at active site. (b) Overlay with wt AspH:Mn:2OG:hFX-Asp_{103} active site highlighting the different positioning between at the C5 carboxylate of 2OG and the elongated C8 carboxylate of **28** in the R735Q active site. (c) Alternative view of **28** to show position of C8 carboxylate. The substrate hFX-Asp_{103} is omitted for clarity.

The C8 carboxylate of **28** extends into a pocket in proximity to the backbone of I668 and M670. It is possible that electrostatic stabilising interactions from the backbone of these residues can also assist in positioning **28** at the active site. When considering derivatives **26**, **27** and **28**, it is possible that due to their elongated carbon chains compared to 2OG, these derivatives are all able to form stabilising interactions with S688. This interaction enhances their efficiency as cosubstrates for the R735Q variant compared to 2OG. However, only **28** is long enough to interact to form the most stabilising interaction with the side chain of S668, and therefore **28** improves the activity of R735Q the most.

3.3.3.3 Treatment of FDLAB syndrome with 2OG derivatives

Overall, the activity assays and structural data on these 2OG derivatives showed that the bump-hole theory can be successfully applied to clinically relevant AspH variants R688Q and R735Q. The "bumped" 2OG-derivatives were effectively used as cosubstrates by the R688Q and R735Q variants, in which the substitution of Arg to Gln had created "holes" around the AspH active site. These data, and previous results on Adult Refsum Disease, suggest that there is therapeutic potential in designing "bumped" cosubstrates for used in treatment of genetic diseases where mutations result in impaired enzyme catalysis.

However, the actual application of this theory is likely to be challenging. Firstly, from a pharmacological perspective, ensuring the therapeutic derivative is adequately taken into cells would be difficult. However, some of the tested 2OG derivatives are natural products and present in human nutrition which suggests this may, in theory, be feasible.[259,260] Secondly, AspH is highly expressed during foetal development and the significance of AspH catalysed hydroxylation is likely to be greater at this time than during adulthood; thus, efficient treatment initially requires diagnosis at an early stage as with other metabolic diseases. In this regard, it has been proposed that screening for mutations in the *ASPH* gene should be performed as standard test. Finally, it would be challenging to ensure that 2OG-dependent oxygenases other than AspH and/or other

2OG utilizing enzymes are not inhibited by the therapeutic 2OG-derivative, as has been observed in the case of AspH and FIH.[259]

3.4 Conclusions and future work

FDLAB syndrome associated mutations in the *ASPH* gene are predicted to cause impaired functioning of the AspH enzyme in vivo, leading to pathophysiology in the ocular system and facial dysmorphisms. This chapter describes the characterisation of AspH variants with the clinically-relevant missense mutations R735Q/W and R688Q and has unambiguously shown that the mutations do indeed impair AspH hydroxylation activity in vitro. This finding provides evidence for the causative relationship between FDLAB syndrome and reduced AspH function.

The R688Q, R735Q, R735W and G434V variants presented with different kinetic profiles; this ranged from enhanced activity of the G434V variant compared to wt AspH to complete inactivity of the R735W variant. Further *in vivo* investigations could assess whether these differences manifest in patients, that is to say compare *ASPH* expression or substrate hydroxylation levels between patients carrying different pathogenic mutations.

The activity of the clinically relevant variants R688Q and R735Q can be selectively improved by use of 2OG derivatives. These 2OG-derivatives exploit the physical properties of other residues at the active site and enable the derivatives to likely form more catalytically productive complexes than possible with 2OG. The activity of the R688Q variant was improved by 2OG-derivatives which possess substituents on the C3 carbon and can therefore form stabilising interactions with a hydrophobic active site pocket. The activity of the R735Q variant was improved by increasing the number of carbons in the main chain of the 2OG derivatives. These data indicate the therapeutic potential of using 2OG-derivatives to alleviate symptoms of FDLAB syndrome. The actual application of a bump-hole theory using synthetic 2OG-derivatives as a treatment would require optimisation as the derivatives may also modulate the activity of other human 2OG-dependent oxygenases.

3.5 Acknowledgements

All protein purifications, kinetic assays described in sections 3.3.2.1 and initial crystallisation trials of all variants were performed by Ms Cynthia Hou. 2OG-derivatives were synthesised by Dr Lennart Brewitz.

3.6 Supplementary information

Table S3-1. Data collection and refinement statistics for crystal structures in Figure 3-5 and Figure 3-6.

	R735Q:acetate:Mn:hFX-Asp$_{103}$	R735Q:2OG:Mn:hFX-Asp$_{103}$
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	44.28 - 2.03 (2.103 - 2.03)	46.42 - 2.1 (2.175 - 2.1)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	50.02 88.56 123.42	50.10 90.74 123.37
α, β, γ (°)	90 90 90	90 90 90
Total reflections	335685 (25557)	449379 (44872)
Unique reflections	35755 (3458)	33624 (3315)
Multiplicity	9.4 (7.4)	13.4 (13.5)
Completeness (%)	98.54 (96.91)	99.94 (99.88)
Mean I/sigma(I)	12.93 (4.47)	15.40 (1.49)
Wilson B-factor (Å^2)	30.50	38.46
R-merge	0.09273 (0.4641)	0.09178 (0.6408)
CC1/2	0.998 (0.948)	0.983 (0.938)
CC*	1 (0.986)	0.996 (0.984)
Reflections used in refinement	35715 (3446)	33610 (3311)
Reflections used for R-free	1966 (190)	1997 (196)
R-work	0.2046 (0.3518)	0.1865 (0.2459)
R-free	0.2400 (0.3815)	0.2178 (0.2835)
CC(work)	0.946 (0.853)	0.928 (0.923)
CC(free)	0.940 (0.789)	0.939 (0.932)
Number of non-hydrogen atoms	3959	3817
macromolecules	3592	3490
ligands	14	15
solvent	356	316
Protein residues	447	446
RMS(bonds)	0.002	0.010
RMS(angles)	0.53	0.85
Ramachandran favored (%)	98.65	98.64
Ramachandran allowed (%)	1.35	1.36
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.54	0.85
Clashscore	3.38	2.05
Average B-factor	42.86	44.83
Macromolecules (Å^2)	42.72	44.75
Ligands (Å^2)	49.06	43.84
Solvent (Å^2)	44.05	45.77

Table S3-2. Data collection and refinement statistics for crystal structures in Figure 3-6 and Figure 3-8.

	R735W:2OG:Mn:hFX-Asp$_{103}$	**R688Q:Mn:hFX-Asp$_{103}$**
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	46.5 - 1.948 (2.018 - 1.948)	43.47 - 1.75 (1.813 - 1.75)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	50.12 91.04 124.90	50.23 86.78 123.26
α, β, γ (°)	90 90 90	90 90 90
Total reflections	570332 (56901)	714890 (47748)
Unique reflections	42576 (4164)	55013 (5261)
Multiplicity	13.4 (13.7)	13.0 (9.1)
Completeness (%)	99.87 (98.95)	99.66 (96.90)
Mean I/sigma(I)	21.68 (1.81)	17.65 (0.87)
Wilson B-factor (Å^2)	37.41	27.89
R-merge	0.06868 (0.695)	0.07854 (1.019)
CC1/2	0.918 (0.922)	0.999 (0.764)
CC*	0.978 (0.98)	1 (0.931)
Reflections used in refinement	42568 (4160)	54982 (5244)
Reflections used for R-free	2000 (196)	2755 (295)
R-work	0.2178 (0.3593)	0.1799 (0.3337)
R-free	0.2639 (0.3549)	0.2123 (0.3652)
CC(work)	0.933 (0.633)	0.961 (0.853)
CC(free)	0.926 (0.635)	0.942 (0.773)
Number of non-hydrogen atoms	3796	3985
macromolecules	3514	3439
ligands	15	1
solvent	271	545
Protein residues	447	446
RMS(bonds)	0.012	0.010
RMS(angles)	1.51	1.07
Ramachandran favored (%)	96.61	98.87
Ramachandran allowed (%)	3.39	1.13
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	1.41	0.29
Clashscore	9.00	3.15
Average B-factor	45.97	32.09
Macromolecules (Å^2)	46.07	30.83
Ligands (Å^2)	28.06	26.40
Solvent (Å^2)	45.34	40.05

Table S3-3. Data collection and refinement statistics for crystal structures in Figure 3-9 and Figure 3-10.

	G434V:2OG:Mn:hFX-Asp$_{103}$	R688Q:Mn:3-methyl 2OG:hFX-Asp$_{103}$
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	50.49 - 2.4 (2.486 - 2.4)	43.61 - 1.85 (1.916 - 1.85)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	50.21 86.9 124.07	50.61 85.99 123.96
α, β, γ (°)	90 90 90	90 90 90
Total reflections	294850 (29959)	629100 (63347)
Unique reflections	21922 (2179)	47007 (4616)
Multiplicity	13.4 (13.7)	13.4 (13.7)
Completeness (%)	99.89 (99.95)	99.91 (99.85)
Mean I/sigma(I)	12.93 (1.21)	13.50 (1.31)
Wilson B-factor (Å^2)	51.10	31.79
R-merge	0.1632 (2.462)	0.1301 (1.752)
CC1/2	0.999 (0.61)	0.998 (0.703)
CC*	1 (0.87)	0.999 (0.908)
Reflections used in refinement	21909 (2178)	46967 (4616)
Reflections used for R-free	1051 (106)	1960 (201)
R-work	0.2073 (0.3576)	0.1761 (0.2736)
R-free	0.2524 (0.4119)	0.2124 (0.3412)
CC(work)	0.964 (0.781)	0.968 (0.858)
CC(free)	0.945 (0.728)	0.941 (0.840)
Number of non-hydrogen atoms	3677	3893
macromolecules	3627	3570
ligands	11	44
solvent	39	279
Protein residues	447	447
RMS(bonds)	0.004	0.008
RMS(angles)	0.94	0.92
Ramachandran favored (%)	98.19	98.87
Ramachandran allowed (%)	1.81	1.13
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.78	0.54
Clashscore	4.82	2.81
Average B-factor	62.32	39.13
Macromolecules (Å^2)	62.40	38.64
Ligands (Å^2)	63.94	46.25
Solvent (Å^2)	54.92	44.22

Table S3-4. Data collection and refinement statistics for the crystal structure in Figure 3-11.

	R735Q:2-oxosuberate:Mn:hFX-Asp$_{103}$
Wavelength (Å)	0.9763
Resolution range (Å)	61.9 - 1.699 (1.76 - 1.699)
Space group	P 2_1 2_1 2_1
Unit cell	
a, b, c (Å)	50.36 86.34 123.80
α, β, γ (°)	90 90 90
Total reflections	800477 (65791)
Unique reflections	60223 (5861)
Multiplicity	13.3 (11.2)
Completeness (%)	99.85 (98.92)
Mean I/sigma(I)	14.78 (1.10)
Wilson B-factor (Å^2)	24.91
R-merge	0.08308 (0.7272)
CC1/2	0.999 (0.886)
CC*	1 (0.969)
Reflections used in refinement	60197 (5861)
Reflections used for R-free	1999 (195)
R-work	0.1624 (0.2725)
R-free	0.1837 (0.2989)
CC(work)	0.967 (0.914)
CC(free)	0.969 (0.885)
Number of non-hydrogen atoms	4121
macromolecules	3634
ligands	25
solvent	472
Protein residues	447
RMS(bonds)	0.020
RMS(angles)	1.79
Ramachandran favored (%)	99.55
Ramachandran allowed (%)	0.45
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.26
Clashscore	5.54
Average B-factor	31.10
Macromolecules (Å^2)	30.15
Ligands (Å^2)	24.60
Solvent (Å^2)	38.59

Chapter 4 | Towards time-resolved X-ray crystallography studies on AspH

4.1 Introduction

4.1.1 Methods in structural biology

In 1937 Max Perutz set out on an ambitious project to determine the structure and function of haemoglobin using X-ray diffraction.[261] Twenty years later, combined with work from John Kendrew, a 5.5 Å structure of horse haemoglobin and a 2 Å resolution structure of myoglobin was solved.[31,262] The pair were subsequently awarded the Nobel Prize for Chemistry in 1962, the same year Watson and Crick received the prize for their work on elucidating the structure of DNA.[263] Over the following six decades the field of structural biology has rapidly evolved, underpinned by advances in instrumentation, data collection and computing power. Obtaining a three-dimensional structure from a single protein crystal can now often be achieved in minutes, a massive advance compared to Perutz's two-decade long venture.

X-ray diffraction is the most widely-applied method in structural biology. As of May 2023, nearly 86% of structures in the PDB were determined by X-ray crystallography, compared to 7% from cryo-electron microscopy (cryo-EM) and 7% from NMR. The effectiveness and popularity of X-ray diffraction as a structural biology method has been supported by advances in technology and computing power.[264,265] Liquid handling robots that require very small sample volumes can set up trays to screen thousands of crystallisation conditions in a short amount of time.[266,267] Data collection at synchrotrons has become routine, with very few institutes using copper anode X-ray generators to collect data from crystals, increasing the speed and quality of data collected.[268] The development of improved detectors has increased the resolution achievable. The importance of increased computing power cannot be understated; it has enabled more data to be collected, at higher rates, and allowed for processing of large numbers of data sets increasing throughput and quality of models built.[269,270] Automated processing pipelines can provide reduced data processing within two minutes of data-collection, essentially allowing users to receive real-time feedback on their data quality.[271]

The most commonly used X-ray sources for protein crystallography are synchrotron beamlines. However, recent advances in the development of X-ray free electron lasers (XFELs) have shown significant promise as a new technique to further the field. XFELs deliver femtosecond X-ray pulses generated by a linear accelerator-based X-ray source, in comparison to synchrotron radiation which originates from a circular accelerator system.[272] The X-rays generated by XFELs are nine orders of magnitude brighter than synchrotron sources.[273] Data on protein crystals are collected before the crystals are destroyed by radiation damage, in a concept termed "diffraction before destruction". The beam can be focused to deliver submicron beam sizes allowing for data collection on nanometre-sized crystals.[273] XFEL beamlines have therefore been particularly useful in solving structures of difficult to crystallise samples, e.g. membrane proteins which only crystallise in the form of nanocrystals that are unsuitable for traditional X-ray beams.[274,275]

Other frequently used methods in structural biology include circular dichroism, small angle X-ray scattering, cryo-EM, NMR and more recently computational biology, e.g. AlphaFold and molecular dynamics techniques.[276,277,278] Each of these techniques has drawbacks in terms of sample requirements or data resolution limitations and, most importantly, the experimental conditions are typically incomparable to physiological environments.[279–281] Most structural data ultimately provide an averaged single static snapshot of a macromolecule and may not accurately represent the dynamic system that occurs in living organisms. Conformational changes in proteins can be highly temperature dependent and cryocooling during the X-ray diffraction process can distort these results.[282] Indeed, in this thesis I encountered limitations with cryogenic macromolecular crystallographic structural data. Chapter 3 presented two structures of the AspH variants (R735Q and R735W) which both revealed identical unconventional conformations of 2OG at the active site. However, one variant is catalytically active (R735Q) while the other (R735W) is completely inactive, and therefore it would be incorrect to presume that the ability to bind cosubstrates in a single static structure is a guarantee of successful catalysis.

A study comparing the difference between room-temperature and cryogenic structures of thirty proteins concluded that crystal cryocooling can change the conformation of ~35% of side chains, and therefore cryocooling can conceal movements that are essential for protein function.[283] Issues with over-reliance on cryogenic structures have also been previously reported. Crystalline ribonuclease A undergoes temperature dependent structural transitions that render it unable to bind inhibitors at low temperature, suggesting that a degree of flexibility is required for enzyme function.[284,285] Initial studies of human proline isomerase, cyclophilin A (CYPA), could not rationalise the discrepancies between the structures of conformations obtained by cryogenic X-ray crystallography and transitions between states observed by room-temperature NMR.[286,287,288]

Around 94% of X-ray diffraction structures in the PDB are from samples held at cryogenic temperatures, and these form the basis of datasets used to train artificial intelligence for structure prediction and molecular dynamics modelling.[289,218,290,291] Accurate information on movement within protein structure is therefore not only needed to resolve fundamental mechanistic questions, e.g. the nature of motions during enzyme catalysis, but also to inform more accurate computational studies with applied results, for example, improving inhibitor design in drug development processes. Integrating multiple methods into one experiment has the potential to be more successful in generating a more complete picture of molecular structures in biological systems.[292] The search for comprehensive methods to link structure to biological function continues.[293,292]

4.1.2 Serial X-ray crystallography

During a traditional X-ray crystallography experiment a single protein crystal is subjected to an X-ray beam. The crystal is rotated around one axis in fixed intervals (e.g., 0.1°) until a diffraction pattern for each position has been collected. Crystals are cryo-cooled throughout the process to minimise radiation damage from the X-ray beam; however, this means that structures may not reflect the dynamic processes that occur at room temperature. Serial crystallography methods exploit slurries of microcrystals, in which a single diffraction pattern is obtained from a single

crystal, and each image is merged to generate a dataset from which the full protein structure can be determined. Serial crystallography experiments can therefore be performed at room temperature, as each crystal is only subject to the X-ray beam once before the next crystal is exposed to the beam, so radiation damage to a single crystal should not affect the overall structure. In contrast to conventional X-ray crystallography methods, the emergence of serial crystallography has enabled structural biologists to partly overcome the issue of radiation damage at ambient temperatures.

The development of serial crystallography methods has been underpinned by advancements in XFEL technology. The first structure obtained by XFEL serial methods was of photosystem I (PSI), one of the largest membrane protein complexes, that is involved in the transfer of electrons in oxygenic photosynthesis.[294] Chapman et. al. exposed a stream of PSI nanocrystals to an XFEL beam at room temperature. Over three million diffraction patterns were collected and a structure of PSI could be solved from merging parts of the dataset.[294] Radiation damage could be mitigated as the pulses of radiation used were faster than the timescale of most damaging processes. The end result was a room temperature structure of PSI at a resolution of 2.5 Å and proof of concept that serial room temperature experiments were feasible for structural biologists.[294]

4.1.3 Time-resolved serial femtosecond protein crystallography

The principle of time-resolved crystallography is to trap intermediates of catalysis or other dynamic biological processes along time points on the reaction coordinate, with the reaction occurring *in crystallo*. Whilst time resolved crystallography is possible in conventional crystallography by freeze quenching of crystals (i.e., cryocooling crystals at time points after introduction of a substrate), it is not possible to trap short time frames of fast reactions due to limited diffusion in larger crystals and/or the time taken to remove the crystal from the drop and rapidly cool it. Therefore, methods have been developed to perform time-resolved crystallography without the need to manually quench the reaction. These combine serial femtosecond X-ray crystallography (SFX) with methods of reaction initiation to perform time resolved SFX (tr-SFX).

In an ideal experiment, a reaction would be initiated in protein microcrystals, e.g. by chemical mixing of substrate or an essential cofactor into the crystal slurry, or triggering a reaction by light for photosensitive systems.[295] The microcrystals would then be exposed to the XFEL beam; by controlling the time between the reaction trigger event and exposure to the beam a number of structures at different time points can be obtained throughout the catalytic cycle of an enzyme. Ideally the microcrystal slurry would consist of homogenously sized microcrystals that undergo catalysis at exactly the same rate.[295] The end result of an experiment is a series of snapshots of reaction intermediates that provide a greater understand of mechanistic processes in proteins, therefore opening up the field of 'dynamic structural biology'.

There are many challenges to be faced when designing and preparing for tr-SFX experiments. Firstly, the target protein needs to be a suitable system to study, yielding high quality crystals that diffract to a sufficient resolution to observe atomic changes. Large quantities of protein are needed to optimise crystallisation conditions and microcrystal samples, and therefore it is convenient if it is straightforward to produce and purify the protein. The protein must be amenable to preparation of microcrystalline slurries and must retain catalytic activity *in crystallo* in the optimised precipitant conditions. The target enzyme must catalyse a reaction which proceeds at an appropriate rate and can be triggered by a single event. If chemical mixing is required to initiate the reaction, the rate of diffusion of any small molecules into the protein must be significantly faster than the reaction timescale to allow for homogenous reaction pathways throughout the sample.

Secondly, sample delivery systems must be optimised for time-resolved experiments. Initially, protein structures at XFELs were most commonly obtained by using a gas dynamic virtual nozzle (GDVN) to generate a jet of small crystals in their mother liquor that were projected into the path of the XFEL beam.[294,296–301] However, crystals could be damaged during the spray process and this method typically requires around 10 mL of crystal slurry per structure, and therefore consumes a large amount of sample.[302] Additionally, if the flow rate of the GDVN was not synchronised to the

X-ray pulse repetition rate then a large volume of sample was not subject to the XFEL beam and wasted.[303]

An alternative to continuous jets of sample delivery is the use of small volume droplets. Crystal slurry droplets can be generated by piezo drivers or acoustic ejection systems. Droplet ejection can be synchronised with XFEL pulses and shot in free fall or on tape drive systems, in experimental set ups termed drop-on-demand (DoD) and drop-on-tape (DoT) respectively.[304–306] A comparison of crystal efficiency at Linac Coherent Light Source (LCLS), i.e. crystals used per diffraction pattern, showed that acoustic droplet delivery methods could be three orders of magnitude more efficient than comparable GVDN systems.[306] DoT sample delivery systems can also be coupled with reaction trigger methods, for example by moving the tape through an oxygen chamber prior to beam exposure.

Despite the immense challenges in designing and performing tr-SFX experiments there have been an impressive number of successful tr-SFX trials that have provided detailed mechanistic insights into enzyme catalysis (summarised in Table 4-1). Reactions *in crystallo* have been successfully initiated by both light-triggering and chemical mixing systems in tr-SFX experiments. Oxygen triggering in the case of the *Bos taurus* cytochrome c oxidase was achieved by mixing reduced crystal solutions with oxygen-saturated buffer.[307] This enabled the determination of a key oxygen intermediate and provided information about the proton translocation pathways.[307] More sophisticated methods for timed oxygen exposure have been developed using an acoustic drop delivery system onto a tape drive.[308] The sample can be dispensed onto a tape drive which is then moved through an oxygen gas chamber and subsequently into the path of the XFEL beam.[308] Adjusting the tape speed through the chamber controls the time for which the sample is exposed to oxygen, allowing for the measurement of a number of time points ranging from 0.2 to 8 s.[308] This system has been used to probe the mechanism of O_2 activation in *Chlamydia trachomatis* ribonucleotide reductase R2 (*Ct* RNR) and *E. coli* isopenicillin N synthase (IPNS).[86,308] Microcrystals

of IPNS were exposed to O_2 and crystal structures were collected after 400, 500, 800, 1600 and 3000 ms and allowed for detailed insight into the importance of O_2 induced movements in IPNS.[86] Simultaneous collection of tr-X-ray absorption edge spectroscopy (XES) data also enabled investigation into the Fe oxidation state at each time point.[86]

Table 4-1. Summary of a selection of results in the tr-SFX field. LCLS: Linac Coherent Light Source. SACLA: Spring-8 Angstrom Compact free electron LAser.

Enzyme	XFEL	Sample delivery	Trigger	Timescale
Krokinobacter eikastus Rhodopsin 2[309]	SwissFEL	High viscosity media extruder	Optical pump-X-ray probe	800 fs – 20 ms
Thermosynechococcus elongatus BP-1 Photosystem II[310]	LCLS and SACLA	Acoustic droplet ejector onto tape	Optical pump-X-ray probe	50 µs
Equus caballus Haemoglobin[311]	LCLS	Liquid jet by flow-focusing nozzle	No yet performed but potential for chemical mixing	n/a
Pseudomonas protegens Pf-5 isocyanide hydratase[312]	LCLS	Concentric-flow microfluidic electrokinetic sample holder	Chemical mixing - substrate	15 s – 300 s
Bos taurus Cytochrome c oxidase[307]	LCLS	GDVN	Chemical mixing – O_2 exposure	8s post O_2 exposure
Isopenicillin N synthase (IPNS)[86]	LCLS	Acoustic droplet ejector onto tape drive	Chemical mixing – O_2 exposure	400, 500, 800, 1600 and 3000 ms

The ability to trigger a reaction by the addition of oxygen can be applied to the study of 2OG-dependent oxygenases. Previous tr-SFX studies have attempted to characterise 2OG-dependent oxygenases including FIH and AlkB (unpublished results, Schofield Group).[313] AspH is a human 2OG-dependent hydroxylase which differs from most members of the 2OG-hydroxylase enzyme family due to its unusual Fe(II) binding motif which utilises two histidine ligands instead of the more commonly observed HXD/E…H binding triad. The ability of XFEL sample delivery systems

to trigger an enzyme reaction with O_2 makes AspH a promising candidate for tr-SFX experiments. Obtaining snapshots of intermediates may provide a useful insight into the AspH mechanism, and the role of its unusual active site structure.

4.2 Aims of the work described in this chapter

The work described in this chapter was designed with the long-term goal of performing time-resolved serial crystallography experiments on wild-type $AspH_{315-758}$ (henceforth referred to as AspH) at an XFEL beamline to inform on the mechanism of AspH catalysis. Multiple optimisation steps were required before tr-SFX studies on AspH could be initiated; therefore, the aims of the work described in this chapter were to: (i) optimise the purification of AspH from *E. coli* to generate a scalable process that gives appropriate protein amounts suitable for tr-SFX structural studies, (ii) assess the activity of AspH *in crystallo* using AspH macrocrystals by triggering the hydroxylation reaction with O_2, (iii) optimise AspH crystallisation sample conditions for serial crystallography experiments, (iv) perform synchrotron room-temperature fixed-target experiments on AspH microcrystals, and (v) perform XFEL time-resolved spectroscopic experiments on AspH microcrystals

4.3 Results and discussion

4.3.1 Optimisation of the purification of recombinant AspH

Crystallographic studies, and, in particular time-resolved studies, require large quantities of high purity protein. The original method for purifying His_6-AspH from *E. coli* was developed over a decade ago and required modernization. The protocol initially involved a two-step process with purification achieved by a means of a HisTrap column and then size exclusion chromatography. Typical yields of around 5 mg of pure His_6-AspH per litre of media were attained, which is insufficient for time-resolved crystallographic studies.

A number of modifications were therefore made to the AspH purification. Firstly, protein yields were improved by changing the method of *E. coli* cells lysis from sonication to the use of a high pressure cell disrupter. Secondly, it was necessary to add in an extra step to remove excess metal ions that are co-purified with AspH. XFEL time-resolved experiments can measure both X-ray diffraction data and X-ray absorption data from metal in the sample to establish details of the metal oxidation state and therefore excess metal in solution may interfere with this process. Finally, initial tests revealed that removal of the His_6 affinity tag resulted in faster growing, higher quality AspH crystals.

A new four-step process incorporating these steps was developed and is outlined in this section. Protein quality can vary from batch to batch which can subtly affect crystallisation conditions or crystal quality. To ensure crystallisation conditions for time-resolved studies were reliable, a single large-scale AspH purification was performed and the same AspH batch was used for optimisation of crystallisation conditions and all experiments.

A plasmid encoding the human gene of His_6-AspH was transformed into *E. coli* BL21 DE3 competent cells. A single colony was used to generate an overnight starter culture in 2xYT media supplemented with 30 μg mL^{-1} kanamycin as a selection marker. The overnight starter cultures were diluted 1:100 into baffled flasks that contained fresh 600 mL 2xYT, also supplemented with 30 μg mL^{-1} kanamycin. Cultures were incubated at 37°C with shaking (180 rpm) until an optical density (OD) of 1.0 was reached. The incubator and the cultures were cooled to 18°C and then production of His_6-$AspH_{315-758}$ was induced by the addition of IPTG to a final concentration of 0.1 mM. Cultures were incubated at 18°C for 18 h, then pelleted by centrifugation (8 minutes, 11867 x g, 4°C). Approximately 14 g wet cell mass was obtained per litre of media.

The buffers used in the purification of AspH are outlined in Table 4-2. For the large-scale purification of AspH, 500 g of cell pellet was resuspended in Buffer A (100 mL per 20 g cells), with DNAse I and EDTA-free protease inhibitor tablets (Sigma). The homogeneous suspension was

passed through a high-pressure cell disruptor at 25 kPsi at 4°C. The lysate was centrifuged (30 min, 58,387 x g, 4°C), the supernatant was filtered through 0.45 µm sterile filters and then loaded onto a pre-washed 55 mL HisTrap column at a flow rate of 5 mL min^{-1}.

The column was washed with 10 column volumes (CV) of HisTrap buffer A followed by 10 CV of a mixture of 95%$_{v/v}$ buffer A and 5%$_{v/v}$ buffer B. The His_6-tagged protein was eluted using linear gradient from 20 mM to 500 mM imidazole. Fractions were analyzed on SDS Page 4-12% gels and fractions containing His_6-AspH were pooled. The protein was dialysed for 18 h at 4°C against buffer A, with the addition of 400 units of commercial restriction grade thrombin protease to cleave the His_6-tag. Tag cleavage was monitored using solid-phase extraction mass-spectrometry (SPE-MS). SPE-MS analysis of the reaction mixture was used to check the extent of His_6-tag cleavage. The reaction mixture was loaded onto a HisTrap pre-equilibrated in buffer A after 18 hours of incubation at 4°C. Untagged AspH was collected in the flow through of the loaded fraction and eluted with a buffer A wash.

Table 4-2. The composition of buffers used in the purification of AspH.

Buffer	Composition	pH
HisTrap Buffer A	50 mM HEPES 500 mM NaCl 5 mM Imidazole	7.5
HisTrap Buffer B	50 mM HEPES 500 mM NaCl 500 mM Imidazole	7.5
Size exclusion buffer	50 mM HEPES 150 mM NaCl	7.5

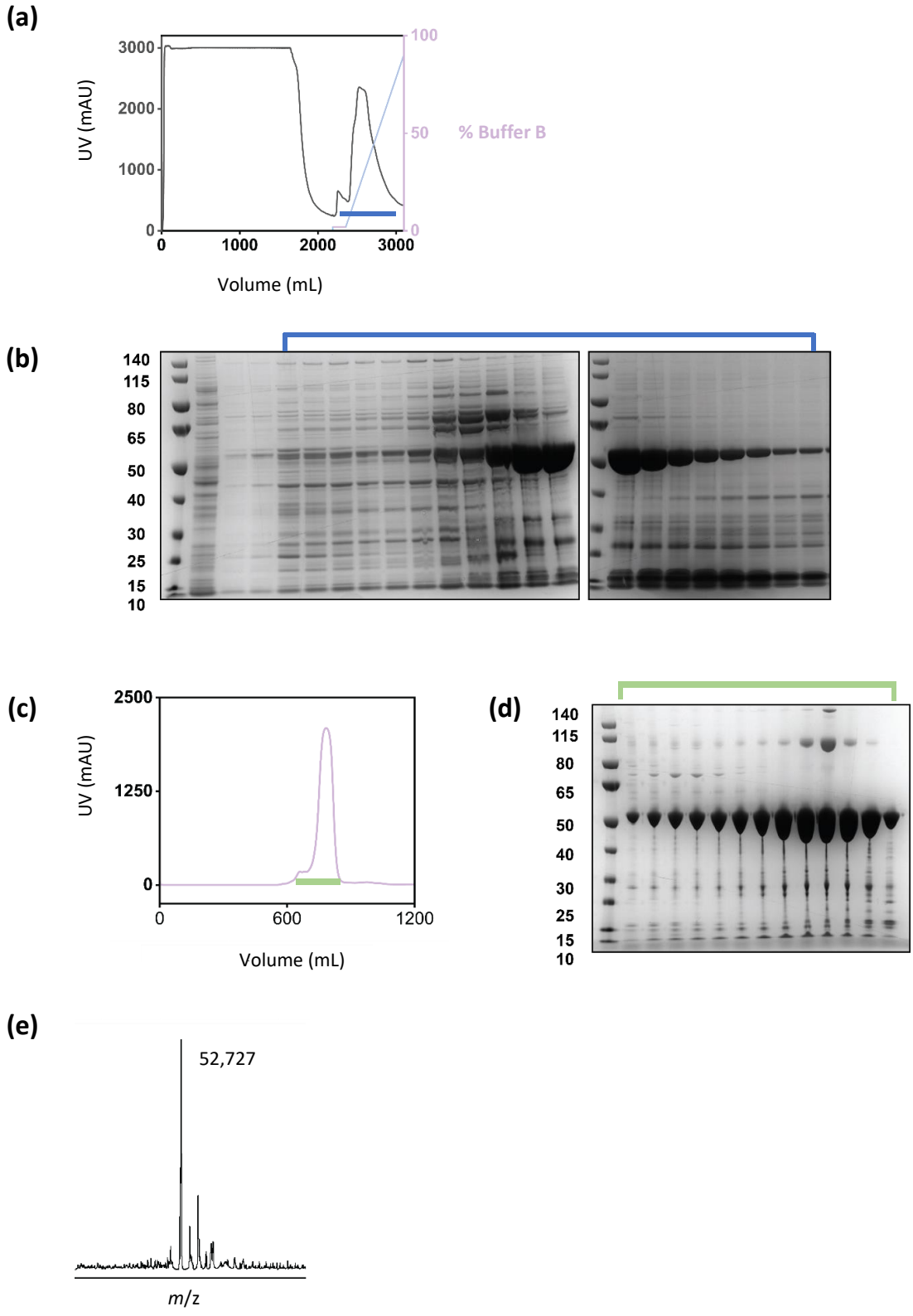


Figure 4-1. Large scale purification of recombinant AspH from *E. coli.* (a) UV trace of HisTrap loading, wash and elution with linear gradient of HisTrap Buffer B. (b) SDS-Page gel of His trap fractions taken from blue line region of UV trace in (a). (c) UV trace from isocratic elution of AspH from 2L SEC and (d) SDS-gel of fractions from green line region of UV trace in (c). (e) MS spectrum of cleaved AspH; calculated mass: 52,727 Da, observed mass: 52,727 Da.

The cleaved protein was then dialysed against 30 mM EDTA and 5 mM 1,10 - phenanthroline in SEC buffer for 2 h at 4°C to remove metal ions bound to AspH. AspH was concentrated and loaded onto a Superdex 75 2 L SEC using a 50 mL SuperLoop. The eluted protein fractions were concentrated to 20.4 mg/mL and aliquoted. Aliquots were flash frozen in liquid N_2 and stored at -80°C.

Using this procedure, a yield of 500 mg purified AspH was obtained, at approximately 8 mg of AspH per litre of 2xYT media. Thus, the protein yield was were improved by about 60 % compared to the previously employed protocol which only yielded His_6-tagged AspH. In future, the purification could potentially be improved by optimisation of the thrombin cleavage step. Approximately 30% of protein remained attached to the His_6 tag after overnight incubation with thrombin protease, suggesting that the cleavage site is not freely accessible to thrombin protease. Subsequently, the sequence of pET28a(+) His_6-AspH was modified to substitute the thrombin cleavage site for a HRV-3C protease recognition site. Complete cleavage of the His_6 tag from AspH was found to occur when using HRV-3C protease. In future it would be advisable to use this construct for large-scale purifications to further improve yields.

4.3.2 Investigations into the structure of AspH using macrocrystals

4.3.2.1 Aerobic AspH crystallography

The first crystal structure of the AspH construct solved in 2014 confirmed that, as predicted by sequence alignments, only two protein histidine ligands were involved in coordinating a metal at the active site.[124] AspH was initially crystallised under aerobic conditions with Mn(II) substituting for Fe(II) to avoid turnover occurring in solution, the 2OG-analogue and AspH inhibitor *N*-oxalylglycine (NOG), and a range of synthetic aspartate-based substrates.[124] NOG was used to obtain a catalytically-inactive stable active site complex. Needle-like AspH crystals formed in the $P2_12_12_1$ space group. A crystal structure of AspH complexed with Mn, 2OG and hFX-Asp_{103} was later also obtained; the structure was refined to final resolution of 2.27 Å (PDB 6YYW).[258] To further

investigate the optimal crystallisation conditions for AspH and obtain a higher-quality 2OG-bound structure, the aerobic crystallography studies were repeated.

AspH was added to solutions of Mn, 2OG and either the synthetic peptide mimics hFX-Asp_{103} or hFX-Asn_{103}, and crystallisation conditions were screened for using the Pact Premier, BCS and LMB broad screens (Molecular Dimensions). Crystals grew after trays were incubated for 24 h at 4°C in a variety of conditions. Crystals grew with a needle morphology, consistent with previously reported results.[124] Needle crystals are advantageous for XFEL studies as their morphology is proposed to facilitate efficient diffusion of O_2 through the crystal lattice and therefore allows a uniform turnover of substrate. Crystals were cryocooled in liquid N_2 with the addition of 25-30 % v/v PEG 400 as a cryoprotectant. Data were collected using MX beamlines i04 or i03 at Diamond Light Source. Crystals from multiple conditions were assessed for diffraction quality, which was determined to be fairly consistent across a large section of the PEG/Anion section of the Pact Premier screen (Figure 4-2). Most combinations of 20% PEG 3350, 0.1 M bis-tris propane pH 6.5, 7.5 or 8.5, and 0.2 M of any of NaBr, NaI, KSCN, $NaNO_2$ resulted in good quality crystals when using either the hFX-Asp_{103} or hFX-Asn_{103} substrate.

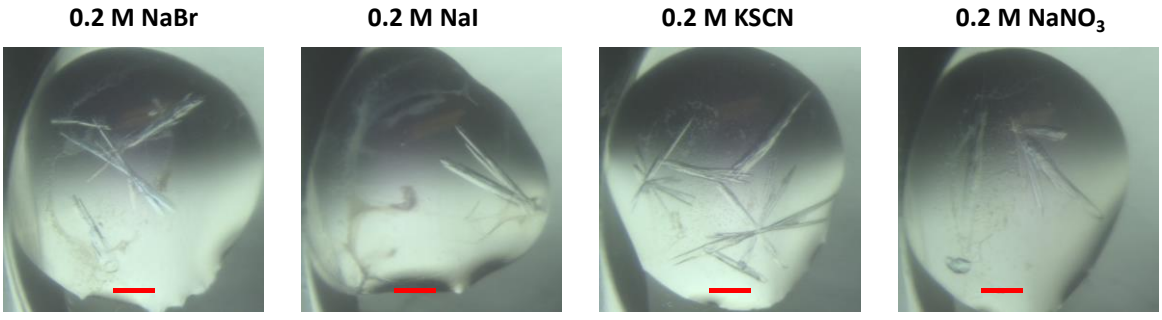


Figure 4-2. Representative images of AspH crystals grown in the Pact Premier broad screen. AspH crystallised in the $P2_12_12_1$ space group. The co-crystallisation solution contained 374 µM AspH, 2 mM 2OG, 1 mM Mn and 800 µM substrate. The precipitant contained 0.1 M bis-tris propane pH 8.5, 20% PEG 3350 and 0.2 M of the labelled anion. Crystal wells containing 0.1 M bis-tris propane at pH 6.5 and 7.5 yielded similar results. The ratio of protein:precipitant was 2:1 (v/v) in 300 nL drops. The red scale bar on photos represents 200 µm. From left to right the wells shown correspond to Pact Premier H2, H3, H4 and H5.

The structure of AspH in complex with hFX-Asp_{103} was refined to a resolution of 1.8 Å, a marked improvement on the 2.27 Å structure previously obtained. The structure was solved by molecular

replacement with a modified PDB 6YYW structure as a search model, in which the cofactors 3-methyl 2OG and Mn were removed from the structure to avoid bias during the molecular replacement process (Figure 4-3).

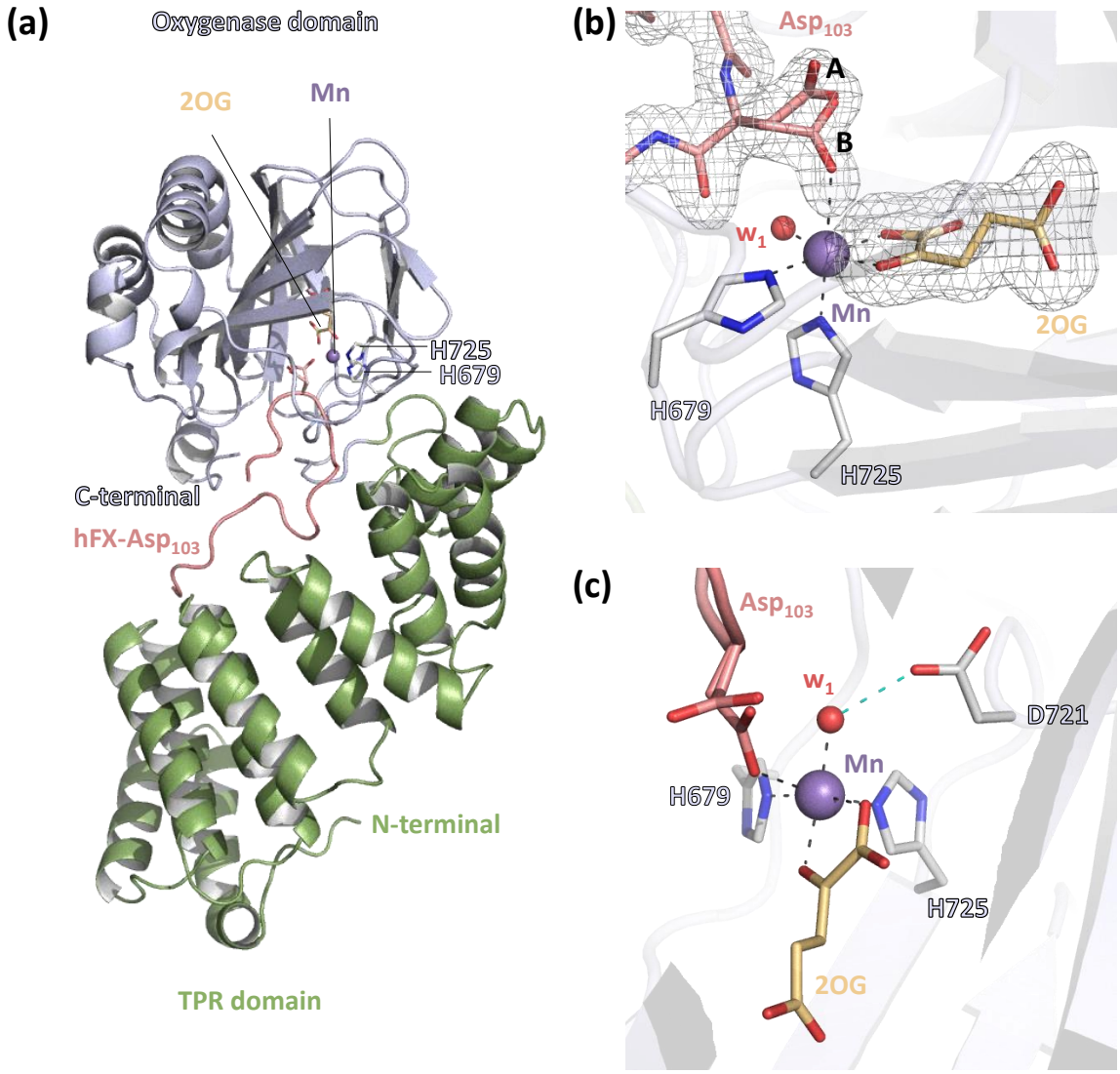


Figure 4-3. View from the structure of an aerobically prepared crystal of AspH complexed with 2OG, Mn and hFX-Asp$_{103}$ substrate. (a) The overall fold of AspH with the TPR domain (green) and oxygenase domain (blue-white). (b) View from the active site with polder omit electron density contoured to 3σ around 2OG and hFX-Asp$_{103}$. Electron density around the substrate Asp$_{103}$ carboxylate shows the two conformations (labelled A and B) that the Asp$_{103}$ can adopt. (c) Alternative view of the active site including D721 which forms a hydrogen bond to a water molecule at the active site (cyan dashes).

The structures of AspH crystallised with Mn, 2OG and the hFX-Asp$_{103}$ substrate were consistent was previous reported results (main chain RMSD = 0.316 Å) .[124] The overall fold of the enzyme construct comprises an oxygenase domain with a double-stranded beta-helix fold (DSBH) and a N-

terminal tetratricopeptide repeat (TPR) domain (Figure 4-3a). Examination of the active site reveals a Mn ion ligated by the side chains of H679 and H725, and bidentate coordination from 2OG (Figure 4-3b). A single water molecule, held in place by hydrogen bonding from D721, is located in the position where a third protein metal binding ligand would typically be found in other 2OG-dependent hydroxylases (Figure 4-3c).

Asp_{103} of hFX-Asp_{103} is observed in two conformations; one with the side chain carboxylate directed away from Mn (conformation A) and one with the Asp_{103} side chain carboxylate positioned to interact with Mn (conformation B) (Figure 4-3b). Substrate conformations A and B have also been detected in the active site structures of AspH variants (Chapters 2 and 3), although at present there is no apparent factor responsible for controlling the substrate conformation observed in a given crystal. In conformation A the substrate Asp_{103} carboxylate is positioned to interact with the side chain of Q627, thus leaving the *pro*-3R-hydrogen on C-3 directed towards Mn, primed for hydroxylation. In conformation B, one oxygen of the Asp_{103} carboxylate acts as a metal-binding ligand. It is likely that this is an inactive conformation as the C-3 methylene of Asp_{103} is positioned away from the metal centre and rearrangement of the substrate to facilitate hydroxylation may be energetically challenging. However, conformation B may have a role in stabilizing the Mn:2OG active site complex, with the substrate Asp_{103} effectively completing the catalytic triad of binding residues that is more commonly observed in the active site of 2OG-dependent hydroxylases.

Investigations into the hFX-Asn_{103} substrate were not performed previously and were therefore carried out for the first time in this work. AspH was co-crystallised with Mn, 2OG and hFX-Asn_{103} using the Pact Premier broad screen. Crystals again crystallised in a needle morphology in the $P2_12_12_1$ space group. Two structures from crystals obtained from two different precipitant conditions were solved using a modified AspH structure (PDB 6YYW), and the structures presented herein were refined to resolutions of 1.8 Å and 1.63 Å (Figure 4-4b and Figure 4-4c respectively).

The overall fold of AspH was not significantly changed by the substitution of hFX-Asp_{103} with hFX-Asn_{103} (Figure 4-3 and Figure 4-4b RMSD = 0.190 Å, Figure 4-3 and Figure 4-4c RMSD = 0.175 Å). However, when AspH was crystallised with hFX-Asn_{103}, the buffer conditions were found to affect the complex formed at the AspH active site. Crystals grown in solutions containing 0.2 M $NaNO_3$, 0.1 M bis-tris propane pH 8.5 and 20% PEG 3350 (Pact Premier well H5) revealed a similar metal complex to that observed with hFX-Asp_{103} (Figure 4-3). The coordination environment around Mn is composed of H679, H725, bidentate coordination from 2OG and two water molecules (Figure 4-4b). However, substitution of $NaNO_3$ for KSCN (Pact Premier wells H4) showed electron density at the active site which could be modelled as a thiocyanate ion coordinating to Mn (Figure 4-4c). The co-ordination of thiocyanate can be confidently modelled with the nitrogen coordinating the Mn due to the significantly larger electron density associated with the sulphur. There was no pH dependence of SCN^- binding as crystals grown in pH 6.5, 7.5 and 8.5 0.1 M bis-tris propane (Pact Premier wells F4, G4 and H4) all exhibited same active site complex. No such binding of SCN^- at the active site was observed under identical precipitant conditions when using hFX-Asp_{103}.

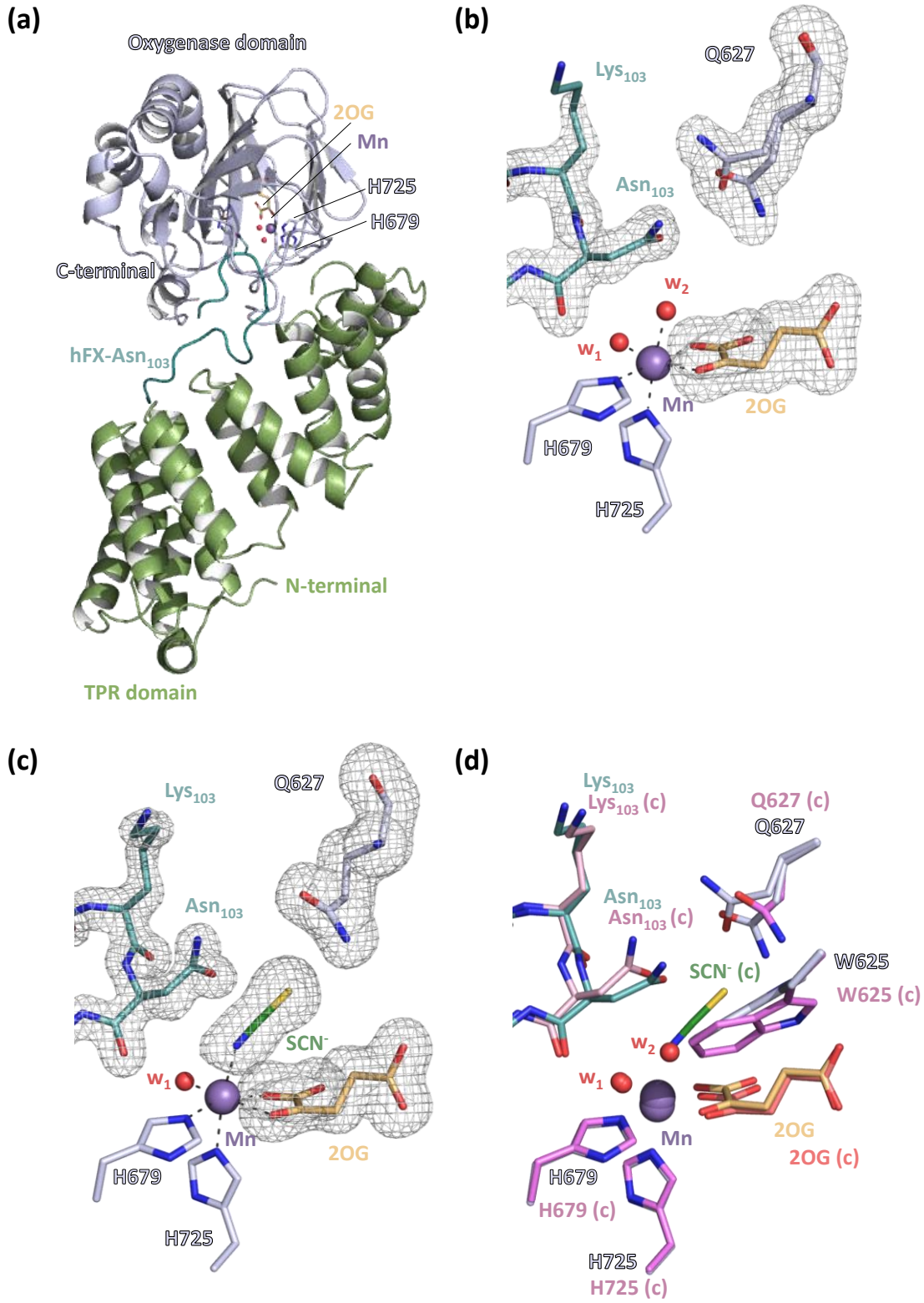


Figure 4-4. Views from the structures of aerobically prepared crystals of AspH complexed with Mn, 2OG and hFX-Asn$_{103}$. (a) The overall fold of AspH in complex with hFX-Asn$_{103}$. (b) View from the active site of AspH:Mn:2OG:hFX-Asn$_{103}$ when the precipitant does not contain KSCN (0.1M bis-tris propane pH 8.5, 0.2 M $NaNO_3$, 20% PEG 3350). (c) View from the active site of AspH:Mn:2OG:hFX-Asn$_{103}$ when precipitant contains KSCN. (d) Superimposition of the active site (b) (in original colouring) and active site (c) (AspH protein residues: pink, 2OG: orange, and substrate: pale pink). Polder omit electron density is carved to 3σ.

Examination of the active site of AspH:2OG:Mn:hFX-Asn_{103} crystals grown in the absence of KSCN (Figure 4-4b) reveals subtle differences of the hFX-Asn_{103} active site compared to the equivalent hFX-Asp_{103} structure. Firstly, substrate conformation B (Figure 4-3b), in which the substrate side chain has a role in ligating Mn, was not observed in the hFX-Asn_{103} structures, possibly reflecting the difference in pKa between the hFX-Asp_{103} and hFX-Asn_{103} substrates. Instead, a second water molecule completes the six-coordinate environment around Mn (w_2). Secondly, electron density around Q627 suggests that the amide side chain is moving between two conformations (Figure 4-4b). Repeated rounds of refinement indicated that this shift is not simply a lateral movement, but a flip in the side chain, effectively keeping the carboxylate oxygen in a similar position with the amine group changing position. The Q627 carboxylate oxygen remains within hydrogen bonding distance (2.81 or 3.01 Å) of the Asn_{103} amine side chain in either conformation. Note that the side chain of Q627 has also been observed to be of importance in other AspH structures, and appears to have a dynamic role in substrate binding.

The inclusion of an SCN^- ion at the active site altered the conformation of a number of secondary sphere residues at the AspH active site. Superimposing the structures of AspH:2OG:Mn:hFX-Asn_{103} with and without SCN^- (Figure 4-4d, superimposition of Figure 4-4b and Figure 4-4c) shows that in the presence of SCN^- the side chain of Q627 has moved away from Mn and is no longer within hydrogen bonding distance to the substrate Asn_{103} side chain (>5 Å from Asn_{103} amine to Q627 amine). Additionally, W625 has shifted away from Q627 and the substrate Asn_{103} side chain has also moved position. The presence of SCN^- increases the distance from Mn to Cβ of Asn_{103} from 4.47 Å to 4.64 Å and the angle between Cα-Cβ-Cγ (∠αβγ) of Asn_{103} decreases from 114.10° to 110.47°. SCN^- likely occupies the oxygen binding site and therefore movement in nearby residues may be reflecting movement at the active site after O_2 coordination. It is possible that, when O_2 coordinates to Fe(II) at the active site during the catalytic cycle, O_2 disrupts a stabilising interaction between Q627 and Asn_{103}/Asp_{103} of the substrate. This allows the substrate Asn_{103}/Asp_{103} to adopt a modified conformation, with a more acute ∠αβγ and Cβ positioned further away from the metal.

This positioning may be more suitable for subsequent hydrogen radical abstraction from Cβ during catalysis. The movement of the Q627 side chain also suggests a possible oxygen-sensing role for this residue.

It is interesting to note that SCN^- coordination at the active site was observed in AspH structures with hFX-Asn_{103}, but not hFX-Asp_{103}. SCN^- from the crystallisation buffer likely accesses the active site via solvent channels in AspH, and it is probable that this is occurs in the AspH:Mn:2OG:hFX-Asp_{103} crystals under the same conditions. It is likely that the interaction between Q627 and the Asp_{103} carboxylate is stronger than the interaction between Q627 and the Asn_{103} amide, and therefore represents a more stable arrangement of side chains *in crystallo.*

4.3.2.2 Assessment of AspH activity *in crystallo*

Anaerobic crystallisation of AspH

To investigate whether AspH retains catalytic activity *in crystallo,* crystallisation of an anaerobic AspH:2OG:Fe:substrate complex was first pursued. Small scale optimisation wells based around the previously successful Pact Premier conditions were set up under a N_2 atmosphere within an anaerobic chamber. AspH was co-crystallised with 2OG, ferrous ammonium iron sulphate (FAS) and hFX-Asp_{103} or hFX-Asn_{103} at 4°C. The crystals were harvested anaerobically into liquid N_2 from wells containing a precipitant mixture of either 0.1 M bis-tris propane pH 7.5, 0.2 M KSCN and 18-22% $_{v/v}$ PEG 3350, or 0.1 M bis-tris propane pH 7.5, 0.2 M NaBr and 18-22%$_{v/v}$ PEG 3350, supplemented with 25% $_{v/v}$ PEG 400 immediately before cryocooling. Crystals were grown in the presence and absence of KSCN in order to investigate the effect of SCN^- ion binding at the active site on hFX-Asn_{103} hydroxylation. Diffraction data were collected using MX beamline i03 at Diamond Light Source and structures were solved by molecular replacement using a modified PDB 6YYW as a search model, with cofactors at the active site removed.

The overall fold of anaerobically prepared AspH is virtually unchanged from that of the aerobically crystalized Mn equivalent structures (RMSD = 0.144 Å for alignment of AspH:Mn/Fe:2OG:hFX-

Asp_{103}, RMSD = 0.147 Å for alignment of AspH:Mn/Fe:2OG:hFX-Asn_{103}, RMSD = 0.156 Å for alignment of AspH:Mn/Fe:2OG:hFX-Asp_{103}:SCN^-). Examination of the AspH active site reveals similar structures to the Mn equivalent presented in Figure 4-3 and Figure 4-4. With either hFX-Asp_{103} or hFX-Asn_{103} substrates, Fe is coordinated by H679 and H725 AspH residues, bidentate coordination from 2OG and either one or two water molecules (Figure 4-5).

Table 4-3. Precipitant conditions used to prepare anaerobic crystals of AspH in complex with 2OG, Fe and hFX-Asp_{103} or hFX-Asn_{103}.

Structure	Precipitant conditions
Figure 4-5a	0.1 M bis-tris propane pH 7.5, 0.2 M KSCN, 18%$_{v/v}$ PEG 3350
Figure 4-5b	0.1 M bis-tris propane pH 7.5, 0.2 M NaBr, 18%$_{v/v}$ PEG 3350
Figure 4-5c	0.1 M bis-tris propane pH 7.5, 0.2 M KSCN, 18%$_{v/v}$ PEG 3350

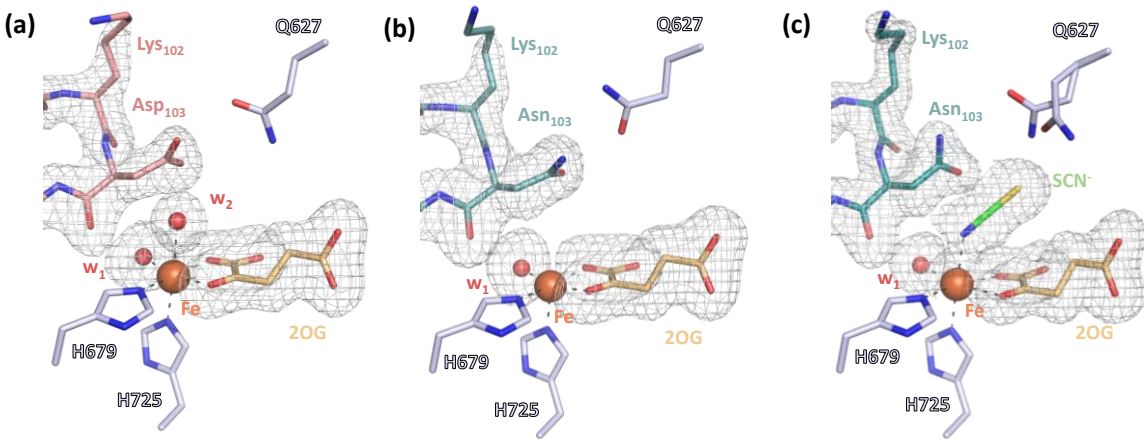


Figure 4-5. Views from the structures of anaerobically prepared AspH crystals complexed with 2OG, Fe and hFX-Asp_{103} (a) or hFX-Asn_{103} (b) and (c). (a) View of the AspH active site in complex with 2OG, Fe and hFX-Asp_{103}. (b) View of the AspH active site in complex with 2OG, Fe and hFX-Asn_{103}. (c) The AspH active site in complex with 2OG, SCN^-, Fe and hFX-Asn_{103}. Polder omit electron density is shown carved to 3σ. The precipitant conditions are outlined in Table 4-3.

It is interesting to note that only a single conformation of the substrate Asp_{103}, conformation A, is observed in the anaerobic AspH:Fe structures (Figure 4-5a). Conformation A is proposed to be the catalytically productive conformation of the substrate Asp. A second water molecule completes the six-coordinate environment around the Fe centre. It is possible that inactive conformation B of

the substrate may be present in AspH crystals as a result of using Mn or related to the oxygen exposure of the enzyme. Consistent with aerobic crystallography, the presence of SCN^- in the buffer did not alter the complex formed at the AspH active site when crystals contained the Asp_{103} substrate.

Anaerobic AspH:2OG:Fe:hFX-Asn_{103} crystals displayed the same properties as the aerobic crystals with respect to the buffer conditions. In the presence of SCN^- ions in the precipitant, an SCN^- ion can be observed at the active site (Figure 4-5c). The secondary sphere residues Q627 and W625 have been displaced by the presence of the SCN^- ion in the same manner as seen in the aerobic structure with Mn (Figure 4-4d). With Br^- ions in the buffer in place of SCN^- no additional ions were observed to coordinate to Fe (Figure 4-5b). Therefore, the coordination of SCN^- to the active site is not an artifact from using Mn as a substitute for Fe or due to oxygen exposure of the aerobically prepared crystals.

Exposure of anaerobic AspH crystals to O_2

Crystals of AspH in complex with 2OG, Fe and either hFX-Asp_{103} or hFX-Asn_{103} substrate were then removed from the glovebox and thus exposed to atmospheric levels of O_2. Crystals were harvested 15 min and 24 h post oxygen exposure. Electron density suggests that there is complete conversion of both substrates to hydroxylated product within 15 min of exposure to atmospheric O_2 levels, showing that the activity of AspH is conserved *in crystallo* (Figure 4-6).

The overall fold of the enzyme in the O_2-exposed crystals is not significantly changed from that of the anaerobic substrate-bound structures (hFX-Asp_{103} structures RMSD = 0.216 Å; hFX-Asn_{103} structures RMSD = 0.264 Å). Interestingly, for both substrates, the hydroxylated product remains bound to the protein and there is no movement in the protein backbone (Figure 4-6). A significant conformational shift occurs in AspH upon substrate binding and reversal of this process to release the product is likely unfavourable *in crystallo.*

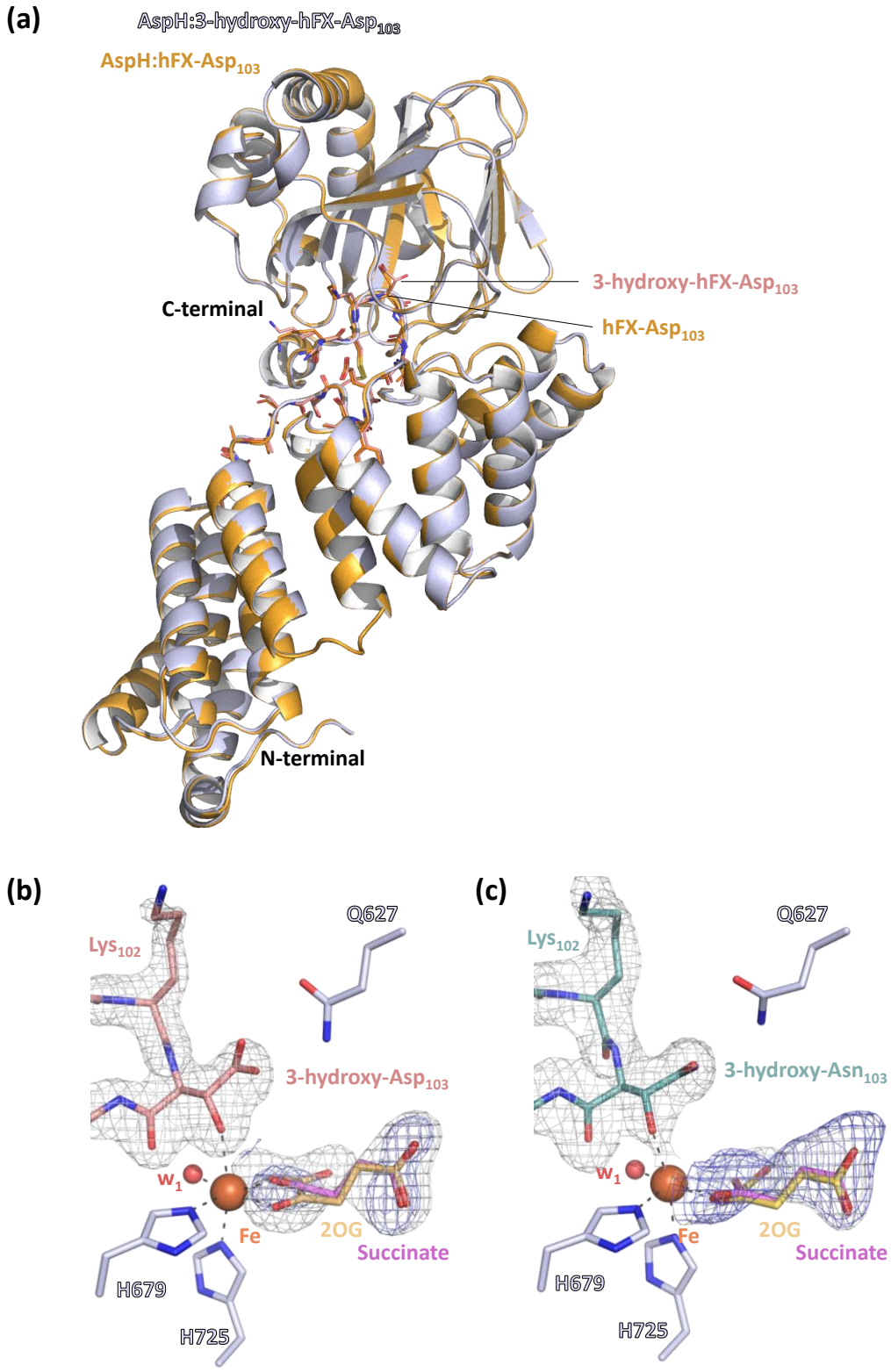


Figure 4-6. Views from the structure of O_2-exposed AspH crystals. (a) Superimposition of the overall structure of AspH:Fe:2OG:hFX-Asp$_{103}$ with AspH:Fe:2OG/Succinate:3-hydroxy-hFX-Asp$_{103}$. (b) A close up view of the active site of AspH:Fe:2OG/succinate:3-hydroxy-hFX-Asp$_{103}$ (c) A close up view of the active site of AspH:Fe:2OG/succinate:3-hydroxy-hFX-Asn$_{103}$. The grey polder omit electron density is carved at 3σ around hydroxyl group of the product and around 2OG. The density carved around succinate is shown in blue.

In both product structures, electron density at the active site clearly shows the presence of 3(*R*)-hydroxylation on Asp_{103} or Asn_{103} (Figure 4-6b and c), as predicted from reported biochemical studies.[123] Modelling conformation B of the Asp_{103} substrate or another water molecule did not fit the electron density around Cβ. Close examination of the stereochemistry of the product Cβ unambiguously confirms stereochemistry of hydroxylation (Figure 4-7). Attempts to model in the 3(*S*)-hydroxy-Asp residue did not convincingly fit electron density.

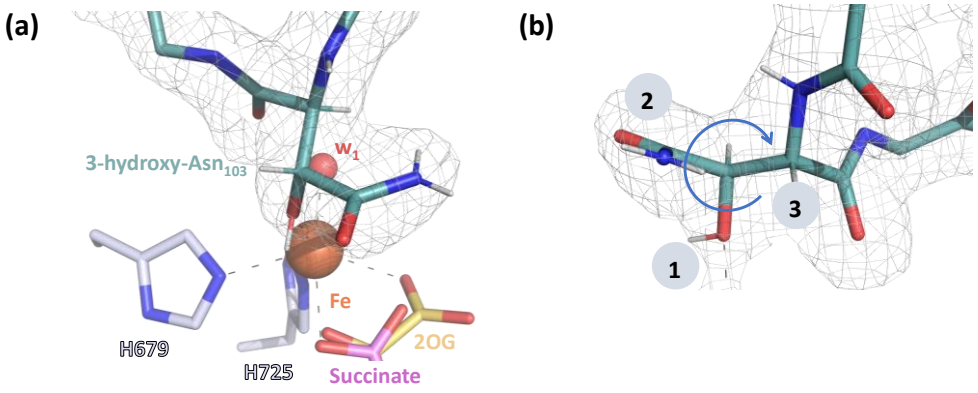


Figure 4-7. Close up views of the stereochemistry of the hydroxylated Asn_{103} residue. (a) Alternative view of the product in Figure 4-6c. (b) Confirmation of (*R*)-stereochemistry of hydroxylation. Numbering of the groups around the Cβ atom in order of priority shows that groups can be ordered in a clockwise direction.

The Cβ hydroxy oxygen is positioned to interact with the metal complex at the active site, occupying the space where conformation B of the Asp_{103} substrate carboxylate can be observed in the hFX-Asp_{103} structures (Figure 4-5) or water 2 in hFX-Asn_{103} structures (Figure 4-4). This product-metal interaction is possibly, at least in part, an artefact of crystallisation as formation of a stable AspH:2OG:Fe:Product complex would hinder turnover in solution. However, given that the biological role of AspH is unclear, it may also reflect a regulatory mechanism.

Although the substrate has been hydroxylated to (3*R*)-hydroxy aspartate, a mixture of 2OG and succinate was modelled into electron density at the AspH active site in both structures (Figure 4-6b and c). Refined occupancies indicate this predominantly consists of 2OG. The shorter length of succinate means it is likely to be less tightly held by coordinating secondary sphere residues, and easily dissociates from the active site after product formation. A large excess of 2OG in the protein

co-crystallisation solution would then enable 2OG to re-coordinate to the metal ion at the active site and form a more stable complex than the succinate bound complex first produced. Additionally, it would not be favourable for catalysis if succinate were to be tightly held at the active site and therefore it is unsurprising that 2OG can displace succinate in the product structures. This proposal is in accord with the inability of succinate to inhibit AspH catalysis (unpublished work from the group).

The nature of succinate binding in the AspH active site

To investigate the nature of succinate binding at the AspH active site, AspH was co-crystallised aerobically with Mn, succinate and hFX-Asp_{103}. The resulting crystals were also soaked in 10 mM succinate solution prior to harvesting (made up of 70%$_{v/v}$ precipitant solution and 30%$_{v/v}$ PEG 400 as cryoprotectant), to ensure detection of succinate at the active site.

The active site of AspH:Mn:succinate:hFX-Asp_{103} revealed a five coordinate Mn complex at the active site, comprising H679, H725, two water molecules and coordination from one of the succinate C1 oxygen atoms (Figure 4-8a). Succinate is positioned to interact with the same secondary residues as 2OG; R688, H690, S668 and R735. The non-metal coordinating carboxylate (C4) of succinate is positioned to interact with R735 (distances 2.7 Å and 2.7 Å), within sufficient distance to form a stable salt bridge, analogous to binding of 2OG (Figure 4-8b). The distance to the H690 imidazole from one of the C1 carboxylate oxygen atoms is 3.3 Å. The other C1 oxygen atom of succinate is positioned to interact with R688 at a distance of 3.3 Å from the R688 carboxylate. The distance from the C1 oxygen of 2OG to R688 is much shorter at 2.7 Å, and therefore succinate is likely to be less strongly bound by this interaction than 2OG. When coupled with the loss of a second coordinating oxygen to the metal center, it is unsurprising that succinate has been partially displaced by excess 2OG in product structures.

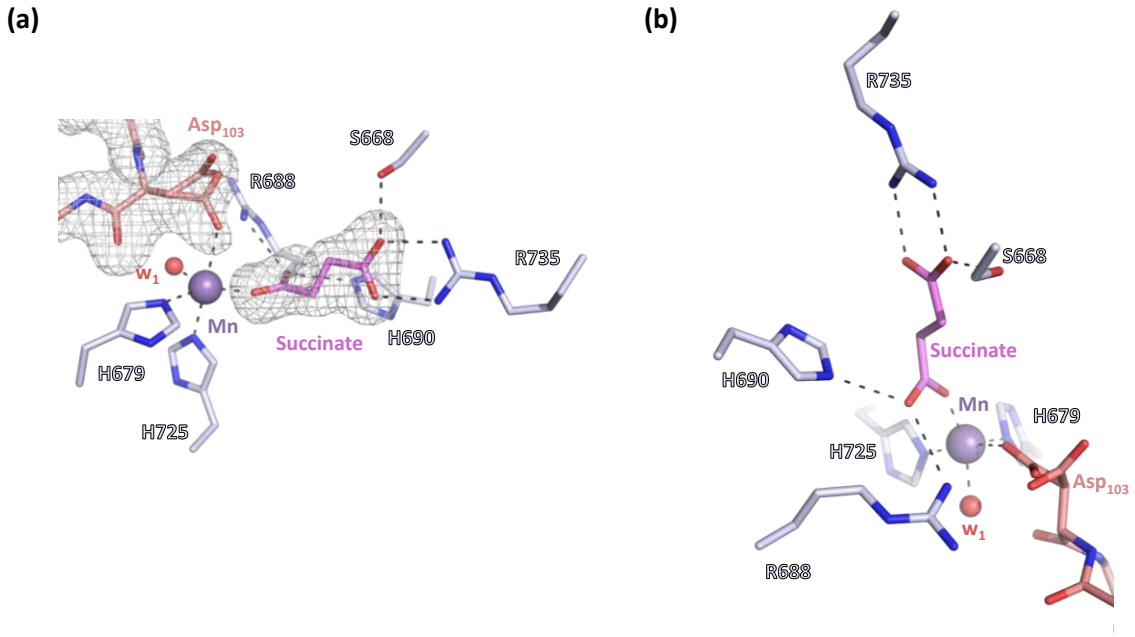


Figure 4-8. View from structure of an AspH crystal in complex with Mn, succinate and hFX-Asp$_{103}$. (a) Polder omit electron density is carved to 3σ around succinate and hFX-Asp$_{103}$. (b) Alternative view to show position of succinate relative to key 2OG binding residues.

In vivo, succinate would only be present at the AspH active site following turnover of substrate to hydroxylated product. As observed in the AspH product bound structures (Figure 4-6), it is possible for the Cβ hydroxyl group to coordinate to the metal centre. It is possible that this complex may affect succinate binding at the active site, perhaps disfavoring formation of a succinate complex, thus acting as a regulatory mechanism to stimulate continuation of the catalytic cycle.

AspH retains catalytic activity *in crystallo*

Confirmation that AspH retains catalytic activity *in crystallo* and that the reaction can be triggered by exposing anaerobically grown crystals to oxygen shows that AspH is a suitable target for tr-SFX experiments. Therefore, optimisation of sample preparation for these methods was next pursued. Limitations in obtaining the synthetic hFX-Asp$_{103}$ and hFX-Asn$_{103}$ substrates during the COVID-19 pandemic lockdown periods meant that the time-resolved studies were only performed using the hFX-Asp$_{103}$ substrate.

4.3.3 Towards time-resolved serial X-ray crystallography on AspH

4.3.3.1 Production of AspH microcrystals

Modern crystallisation techniques have been developed to obtain a small number of large crystals for diffraction, often using small volume sitting or hanging drop vapour diffusions methods. In contrast, serial crystallography experiments require slurries of microcrystals to ensure single diffraction images from single crystals, often in the milliliter scale to obtain enough images for a complete dataset. Optimisation of AspH crystallisation to obtain large quantities of microcrystal production was carried out to reduce crystal size, and to increase both the number of crystals and the volume of crystal slurry. The crystallisation method was consequently optimised from a 300 nL sitting drop crystals to 30-60 µL batch methods to increase volumes of AspH crystals obtained. Microcrystal size and density can be optimised by the addition of seed stocks to the batch crystal wells.

To investigate AspH batch crystal growth, small AspH crystals were first obtained from a 96-well screening tray (Pact Premier) as described previously in 4.3.2. A single well was harvested and the crystals were crushed to form a seed stock (Figure 4-9). Optimisation of seed-stock production determined that use of a single Teflon bead and repeated cycles of vortexing was sufficient to crush samples containing only a few single crystals. However, iterative rounds of seeding using dense samples of batch crystals were not adequately crushed by one Teflon bead. Therefore, batch crystals were crushed by means of a BeadBug microtube high speed homogenizer and 3 mm zirconium beads. The quality of seed stock proved to be fundamental to the quality of microcrystals consequently grown. The AspH co-crystallisation solution (containing 374 µM AspH, 2 mM 2OG, 2 mM Mn and 800 µM hFX-Asp_{103}) was added to precipitant mixtures to a final volume of 30 µL in batch tray wells. Optimisation trials involved adjusting the concentration of PEG 3350, the ratio of protein to precipitant in the wells and the seed stock concentration (Table 4-4).

Table 4-4. Variables altered during batch crystallisation optimisation trials.

Crystallisation variable	Ranges tested
PEG % in precipitant	14%, 16%, 18%, 20%
Ratio of protein:precipitant in well	1:9 1:5 1:3
Seed concentration	Undiluted 1:10 dilution 1:100 dilution

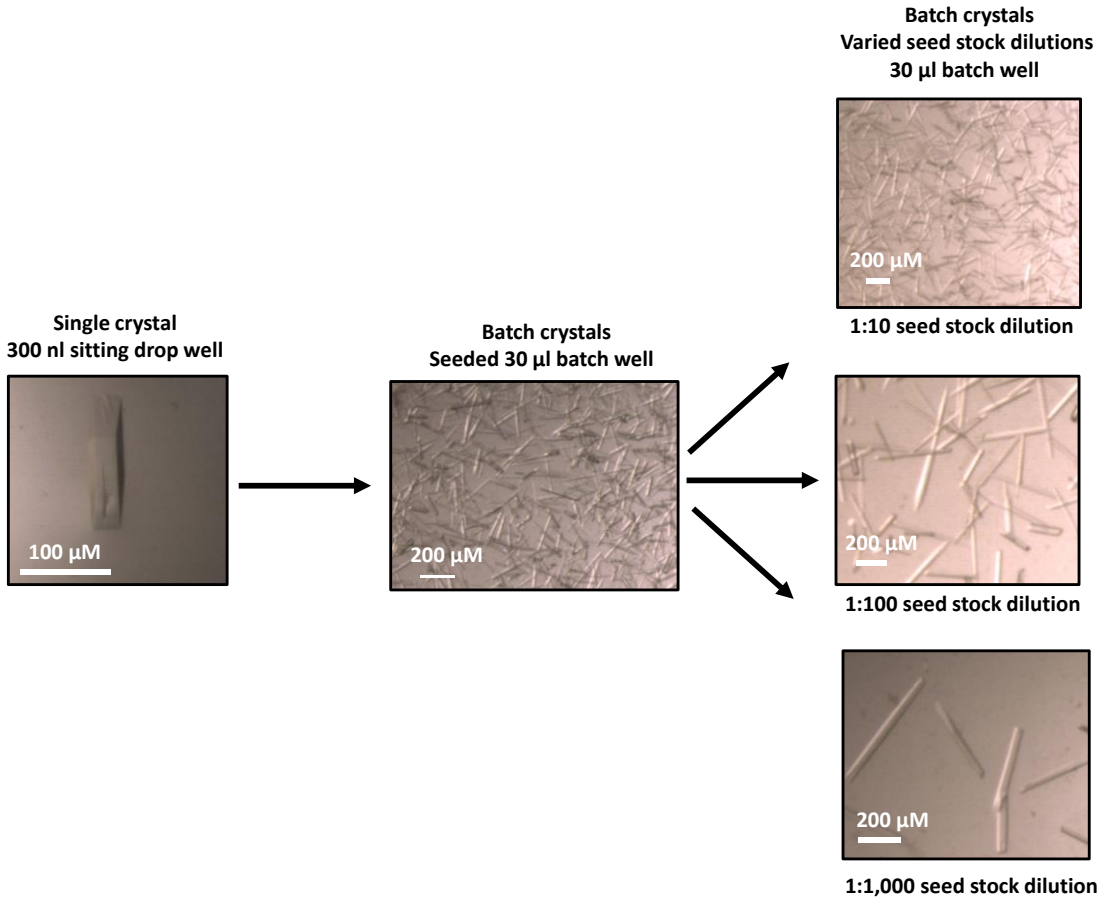


Figure 4-9. The development of AspH seed stocks for the production of microcrystals in batch wells. AspH crystals were first grown in 300 nL wells from 96-well broad screening trays and a single well was crushed to produce seeds (by Teflon seed bead). Batch crystals were then grown by seeding into 30 µL wells containing AspH co-crystallisation solution and precipitant solution. These batch wells were then used to generate seeds (by BeadBug homogenizer) which were applied in another round of batch crystal growth. Higher concentrations of seeds resulted in smaller, denser crystal slurries.

It was found that the optimal protein:precipitant:seed ratio (v/v/v) was 2:1:9, with an AspH concentration of 374 µM, and the precipitant composition was 0.1 M bis-tris propane pH 7.5, 0.2 M of KSCN (or NaBr), 16% PEG 3350 (v/v). Microcrystal production was then also scaled up to 60 µL wells in a 96 well batch plate.

4.3.3.2 Fixed target serial synchrotron crystallography

Although the use of XFEL radiation in room temperature serial crystallography experiments is preferable, XFEL beam time is scarce and requires international shipping of precious samples. There are currently only five working XFELs in the world which are equipped for varying options for sample delivery systems. As such, it is advisable to first test microcrystal diffraction at synchrotrons, which are more widely accessible. Serial crystallographic experiments can be performed at synchrotrons (serial synchrotron crystallography, or SSX), however, crystals at synchrotrons are exposed to a longer duration of X-rays in comparison to rapid XFEL pulses. Sample delivery systems for SSX experiments have evolved from initial experiments that used a simple mesh-screening approach, to more sophisticated hybrid methods involving chips. [314,315] Other sample methods include flowing crystals in a capillary, holding crystals in random positions on a thin silicon wafer, or mounting crystals manually on a support.[316,317]

Methods using chips are increasing in popularity and enable for good quality data collection.[318,319] Crystals are held in apertures in a grid formation on a silicon chip; the beam position can be calibrated to a known point on the chip prior to data collection, allowing each aperture to be located in known X/Y movements.[320] Data collection is rapid and sample handling when setting up the chips is simple and quick. Chips can be prepared under moisture-controlled atmospheres to minimise dehydration damage to the crystals. Crystalline suspension is spread over the chip prior to mounting between mylar film (Figure 4-10).[321,322] It is also possible to use the chip holders without silicon chips, i.e. 'chipless chips', where the sample is directly placed onto the mylar surface. This method has been successfully used in conjunction with anaerobic sample handling set up to investigate oxygen-dependent reactions in three oxygenases: VioC, AlkB and IPNS.[323,324]

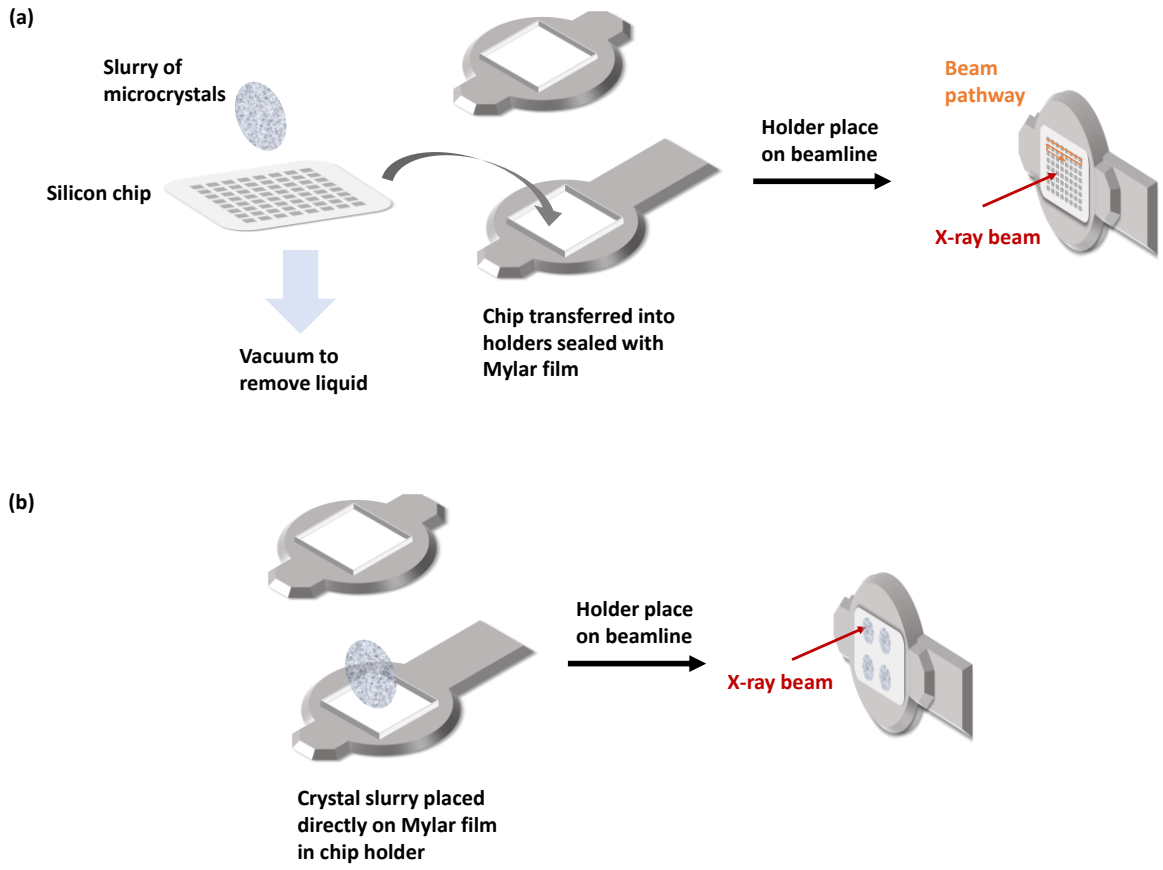


Figure 4-10. Illustration of methods of sample delivery for fixed target SSX at Diamond Light Source. (a) Fixed target sample delivery system using silicon chips. Approximately 150 µL of crystal slurry is spread over the chip and a vacuum applied to draw crystals into the apertures whilst removing excess solution. (b) The fixed target system using "chipless" chips. 1-2 µL drops of crystal slurry are placed directly onto mylar film in holes created by a spacer.

Initial fixed-target SSX experiments on AspH were designed at Diamond Light Source to test the quality of the AspH microcrystal system and to obtain structures of AspH in complex with Mn, 2OG and hFX-Asp_{103} and AspH in complex with Fe and 3-hydroxy-hFX-Asp_{103}. Microcrystals of AspH:2OG:Mn:hFX-Asp_{103} and AspH:2OG:Fe:hFX-Asp_{103} were prepared aerobically and anaerobically respectively. Slurries from batch wells were pooled and the crystals were allowed to settle (Figure 4-11a). Excess precipitant was removed from the suspensions to leave a dense pellet of microcrystals in the tubes (Figure 4-11Figure 4-12a). The AspH:2OG:Fe:hFX-Asp_{103} crystals were then removed from the from the anaerobic chamber and stored at 4°C to allow for oxygen diffusion into the crystals trigger catalysis.

AspH slurries were loaded onto a silicon 'Oxford' chip in a humidity-controlled atmosphere.[325] In accordance with standard procedure, a slight vacuum was then applied to the chip to remove excess precipitant solution and pull the microcrystals into place in the apertures.[325] Unfortunately, no diffraction was observed once the chip was mounted and subjected to the X-ray beam. AspH crystallises in a needle morphology and it is likely that most crystals were pulled through the apertures once a vacuum was applied during the sample handling process. No chips with apertures small enough to catch the needle crystals were available for use and so the alternative method of chipless chips was pursued.

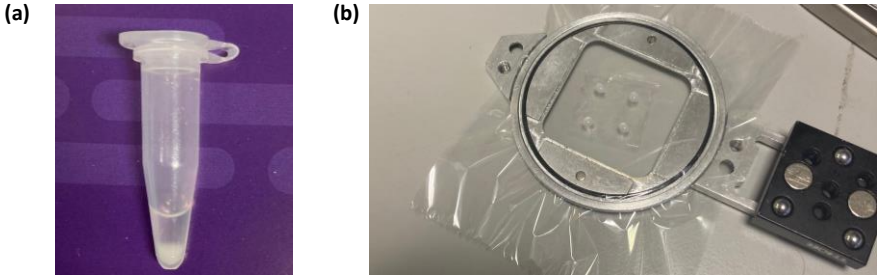


Figure 4-11. Crystal slurries of AspH:Mn:2OG:hFX-Asp$_{103}$ and AspH:Fe:2OG:hFX-Asp$_{103}$ were set up in 60 µL batch plates. (a) The batch crystal wells were pooled in an Eppendorf tube and the microcrystals were allowed to settle to form a dense white pellet with excess precipitant solution on top. (b) 1 µL of the slurry was spread over each hole in a spacer on a mylar film in a "chipless chip" holder. The chip was sealed with a second piece of mylar film in a complementary chip holder.

Mylar film was secured in a chip holder and a spacer containing four positions for sample application was attached to the film (Figure 4-11b). 1 µL of a dense AspH:Mn:2OG:hFX-Asp$_{103}$ crystal slurry was spread onto the mylar film inside the spacer holes and briefly left unsealed to allow for evaporation of excess precipitant solution. The chip was then sealed with another piece of mylar in a chip holder and the sample was placed on the beamline. The beam was calibrated to the top left of each drop and the chip was moved through the beam in raster scan, similar to that of chip data collection methods. The beam was calibrated to the top left of each droplet and set up to collect an array of 100x100 images across each drop. Merging images obtained from two chips enabled the generation of a room temperature structure of AspH in complex with hFX-Asp$_{103}$ substrate, 2OG and Mn (Figure 4-12a).

The highest resolution of wt AspH crystal structures obtained from cryocooled single crystal diffraction was 1.7 Å, although typically the structures were refined to 1.9-2.0 Å. At this resolution, the conformations of side chains and water molecules can be determined. In comparison to the cryogenic data, the resolution of room temperature structures was significantly lower, and the final serial room temperature AspH:2OG:Mn:hFX-Asp_{103} structure was refined to 2.75 Å. At this resolution, it is difficult to resolve details in side chains, water molecules and cofactors. The decrease in resolution between cryogenic and room-temperature data collection is a known phenomenon.[326]

A view from the serial room temperature structure of AspH is shown in Figure 4-12a. The active site of the structure clearly shows Mn coordinated by H679 and H725; however, the rest of the coordination sphere is difficult to determine precisely at this resolution. The conformation of the substrate Asp_{103} carboxylate was modelled as a mixture of both conformations A and B, as electron density between Mn and Cβ of Asp_{103} is ambiguous. There is electron density in the space normally occupied by a 2OG molecule that could indicate the presence of a low occupancy 2OG molecule at the active site. However, this was left unmodelled as 2OG did not fit convincingly in the density.

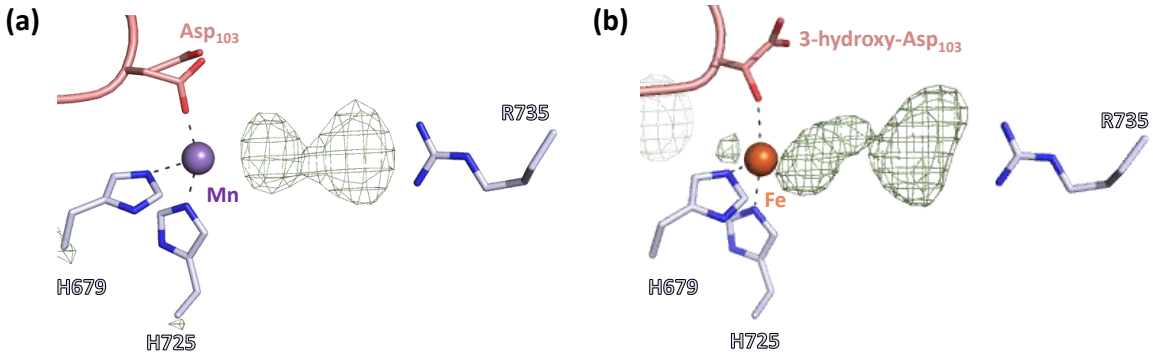


Figure 4-12. Views from the structures of AspH crystals obtained from SSX experiments. (a) The active site of AspH:Mn:2OG:hFX-Asp_{103}. Structure refined to 2.75 Å (b) The active site of O_2 exposed AspH:2OG:Fe:hFX-Asp_{103} crystals. Structure refined to 2.6 Å. mFo-DFc density (green mesh) around Mn or Fe has been left unmodelled in both structures.

Next, the O_2 exposed AspH:Fe:2OG:hFX-Asp_{103} crystals were subjected to the same chipless chip experiments. A 2.6 Å structure generated by merging 18,000 images collected from three chipless

chips was obtained (Figure 4-12b). The AspH crystal slurry was originally prepared anaerobically and then exposed to oxygen. As this method was used to obtain a structure of hydroxylated peptide in single crystals, the structure obtained from SSX was modelled as containing a fully turned-over product complex. However, it is difficult to definitively confirm this observation as a result of the low resolution of the structure. There is still clear evidence for metal bound at the active site but, at this resolution, it is challenging to model 2OG or succinate into the density observed around Fe. It is possible that no succinate or 2OG remains at the active site and water molecules, which are not detected at this resolution, complete the coordination sphere around the Fe centre.

Although the room temperature structures of AspH exhibited a significant decrease in resolution compared to the cryogenically obtained structures, the system still showed promise for time-resolved serial femtosecond crystallography experiments (tr-SFX). Use of an XFEL beam in place of a synchrotron source is known to improve diffraction resolution for room temperature experiments.[326] Structures of adenosine G protein-coupled receptor protein obtained from cryogenic macrocystals, SSX methods and SFX experiments were refined to 1.95 Å, 2.1 Å and 1.7 Å respectively, showing an improvement of 0.4 Å between SSX and SFX data collection on the same sample.[327] Therefore, experimental work to test AspH microcrystal diffraction at an XFEL was pursued.

4.3.3.3 XFEL experiments at Linac Coherent Light Source

XFEL experiments on AspH were planned to be carried out in collaboration with Dr Allen Orville (XFEL hub, DLS) and the groups of Dr Jan Kern and Dr Junko Yano (University of California, Berkley) at Linac Coherent Light Source (LCLS), Stanford. Technical difficulties at the beamline meant that the experiment was postponed from the original date of March 2023, to late September 2023, and I was unable to attend in person. Therefore, crystallisation of AspH and data collection were carried out by Dr Patrick Rabe, Dr Stephen Marshall and Ms Mariska De Munnik.

Initial experiments aimed to obtain room-temperature substrate and product structures of AspH. Using a tape drive sample delivery system, crystal slurry droplets are dispensed onto a krypton tape using an acoustic droplet ejector (ADE) and then the droplets are moved into the XFEL beam on the tape (Figure 4-13).

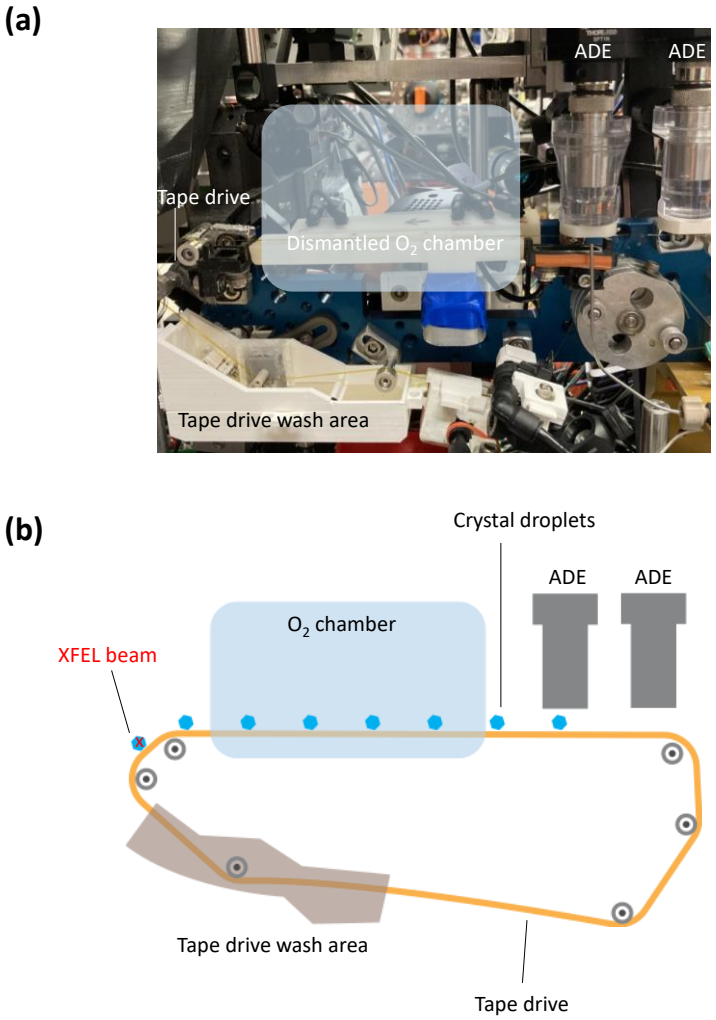


Figure 4-13. The tape drive sample delivery system on the beamline at LCLS. ADE = acoustic droplet ejector. (a) Photograph of the partially-assembled tape drive while not in use. (b) Schematic representation of the tape drive with important features labelled. The tape drive positioned to move anticlockwise (from this perspective). The ADE dispenses drops of crystal slurry (blue hexagons) which are moved through the O_2 chamber. After O_2 exposure, the droplet is then shot by the XFEL beam. The tape drive is then washed, dried and ready for another cycle of data collection.[308]

Crystal slurries of AspH:2OG:Fe:hFX-Asp_{103} were prepared anaerobically and data were collected under an anaerobic atmosphere. Processing of the diffraction data revealed that the sample contained a mixture of crystals with two different lattice parameters (Figure 4-14). These datasets therefore had to be processed separately.

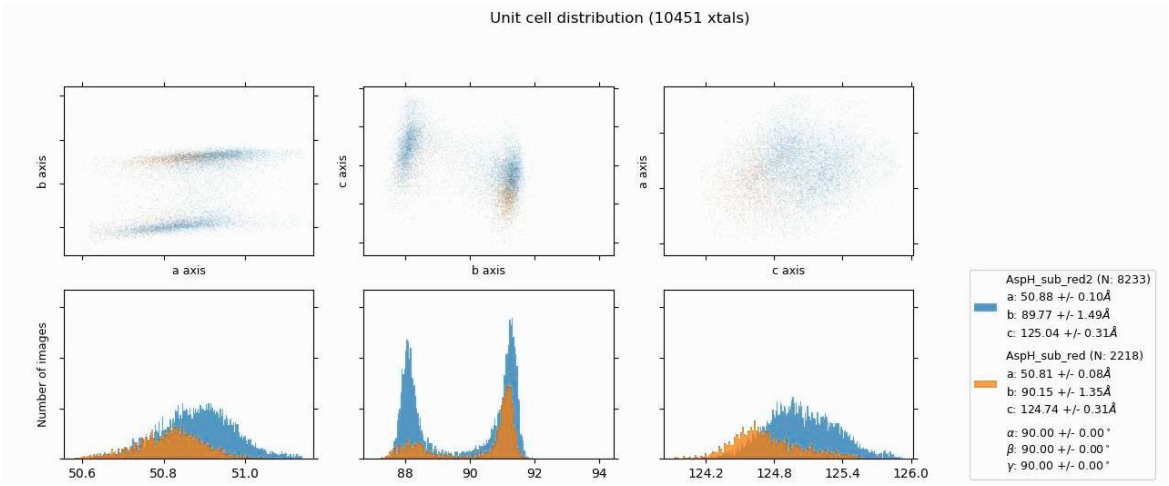


Figure 4-14. The distribution of unit cell parameters of crystals in the AspH microcrystal slurry. The orange and blue parameters represent two different crystal slurries of AspH. The two batches exhibited batch-to-batch variation in the distribution of lattice parameters.

Both of the data sets were refined to a resolution of 2.05 Å. A resolution of 2.05 Å is a significant improvement in resolution from the room-temperature SSX-derived structure (2.75 Å, Figure 4-12a). The overall structure of AspH in complex with 2OG, Fe and hFX-Asp_{103} at RT was not substantially different to the structure determined at cryogenic temperatures (Figure 4-15a, RMSD = 0.262 Å), although the unit cell parameters indicated that there is a contraction of the lattice at cryogenic temperatures (Table 4-5). This was also observed in the SSX-derived dataset.

Table 4-5. Comparison of the unit cell parameters of the AspH crystal unit cell from data collected at cryogenic temperatures and at room temperature (SSX and SFX). Note that there is a contraction of the unit cell at cryogenic temperatures.

Data collection method	Unit cell parameters	
	a, b, c (Å)	α, β, γ (°)
Cryogenic	a = 49.69 b = 86.21 c = 123.38	90, 90, 90
SSX	a = 50.90 b = 91.72 c = 124.72	90, 90, 90
SFX	a_1 = 50.88, a_2 = 50.81 b_1 = 89.77, b_2 = 90.15 c_1 = 125.04, c_2 = 124.74	90, 90, 90

A view of the AspH active site in complex with 2OG, Fe and hFX-Asp_{103} is shown in Figure 4-15b. There are no significant differences in the active site compared to the cryogenically obtained equivalent structure (Figure 4-5). Interestingly, the substrate Asp_{103} side chain carboxylate is only observed in one conformation, and is not directly interacting with the Fe centre. Unfortunately, due to technical difficulties at the beamline, there was no time to collect diffraction data on O_2-exposed AspH crystals in pursuit of a product-bound XFEL-derived structure.

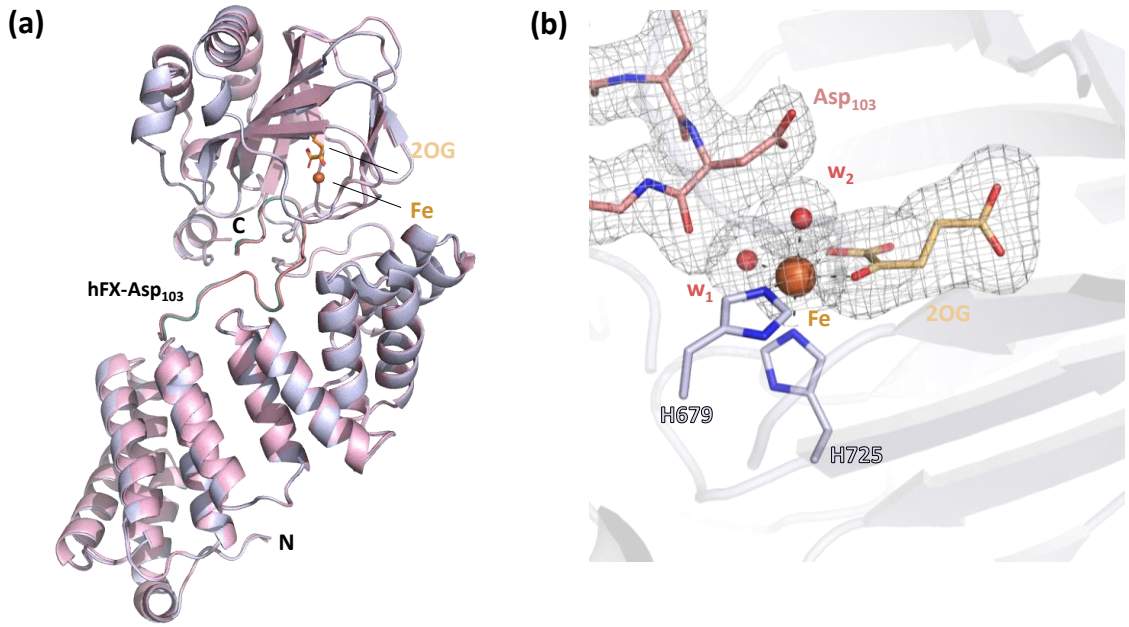


Figure 4-15. View from the crystals structure of AspH in complex with 2OG, Fe and hFX-Asp_{103}. (a) The overall structure of AspH obtained from a single microcrystal at cryogenic temperatures (pink) superimposed on the overall structure of AspH obtained from microcrystals at an XFEL beam line (blue) (RMSD = 0.262 Å). (b) View from the active site of AspH in the SFX-derived room temperature crystal structure. Polder omit electron density is carved to 3σ.

An overlay of the serial XFEL structure of AspH with the cryogenically obtained structure of AspH in complex with product is shown in Figure 4-16. There is a small difference in the relative positioning of Cβ in the product and substrate Asp_{103}, likely due to the observed coordination of the hydroxyl group to the Fe centre. There is no significant difference between the position of the protein residues, H679 and H725, or the position of 2OG at the active site.

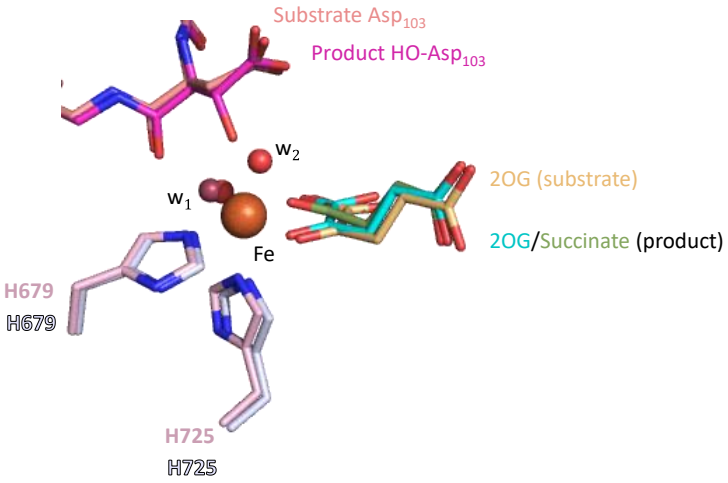


Figure 4-16. An overlay of the XFEL serial dataset substrate structure of wt AspH (Figure 4-15) with the cryogengenically obtained product structure of wt AspH (Figure 4-6b). Figure 4-15 is shown in the original colouring. Figure 4-6b is shown in alternative coluring: protein residues – pale pink, product – bright pink, water – dark red, 2OG – cyan, succinate – green.

4.3.4 Serial crystallography on AspH

A summary of the AspH structures obtained by SSX and the cryogenically obtained macrocrystal equivalent structures is shown in Table 4-6.

Table 4-6. A summary of the structures of AspH obtained by conventional cryogenic crystallography and serial methods. AspH:Substrate refers to AspH:Mn:2OG:hFX-Asp$_{103}$ complexes for cryogen and SSX, and anaerobically prepared AspH:Fe:2OG:hFX-Asp$_{103}$ for SFX. AspH:Product refers to O_2-exposed AspH:Fe:2OG:hFX-Asp$_{103}$ crystals that are modelled as containing AspH:Fe:2OG/succinate:3-hydroxy-hFX-Asp$_{103}$

Complex	Resolution of crystal structures (Å)		
	Cryogenic	**SSX**	**SFX**
AspH:substrate	1.8 (Figure 4-3)	2.75 (Figure 4-12a)	2.05 (Figure 4-15)
AspH:product	1.9 (Figure 4-6)	2.6 (Figure 4-12b)	n/a

Reductions in resolution of 0.95 Å and 0.7 Å between equivalent AspH:Mn:hFX-Asp$_{103}$ and AspH:Fe:3-hydroxy-hFX-Asp$_{103}$ structures respectively were observed. A decrease in structure resolution often occurs when using room-temperature data collection SSX methods compared to

cryogenic data collection as cryogenic structures remove inherent flexibility in the protein to give better diffracting crystals.[326] In similar experiments, structures of *E. coli* oxacillinase-10 (OXA-10) were refined to 1.53 Å using cryogenic macrocrystal data, whereas an equivalent fixed-target SSX-derived structure at Diamond Light Source was refined to 2.3 Å, i.e. a decrease of nearly 0.8 Å was observed when using room-temperature data collection (unpublished results, J Kamps).[328] However, use of an XFEL beam to collect an SFX AspH structure achieved a resolution comparable to that obtained at cryogenic temperatures (2.05 Å). Interestingly, there were no significant differences between the structures of AspH collected at cryogenic temperatures and at room temperature, as evidence by the RMSD values of aligning the carbon backbone (cryogenic and SSX structures RMSD = 0.404 Å, cryogenic and SFX structures RMSD = 0.262 Å).

4.4 Conclusions and future work

This chapter presented work towards establishing an AspH crystal system for use in time-resolved XFEL experiments. Firstly, an improved protein purification method was developed to produce high yields of pure His_6-cleaved AspH. Next, AspH was successfully crystallised under anaerobic conditions, and the structure of the Fe active site complex was revealed to be overall similar to that of the anaerobic Mn equivalent. AspH retains activity in crystallo, as evidenced by the structures of hydroxylated product in the oxygen-exposed macrocrystals**.** Additionally, the crystal structures of product-bound to the AspH active site (Figure 4-6) unambiguously confirmed the reported stereochemistry of hydroxylation catalyzed by AspH.[124] These results are promising and indicate that AspH is a suitable system for further time-resolved femtosecond serial crystallography analysis.

During the crystallisation optimisation process, an SCN^- complex in the AspH active site was observed. The complex was only observed when AspH was crystallised with hFX-Asn_{103} and not hFX-Asp_{103}, indicating a different in the stability of the crystal structures formed with the different substrates. The position of SCN^- binding likely reveals the O_2 binding site at the Fe centre. The

presence of SCN^- did not affect the activity of AspH *in crystallo* and this suggests that the complex was formed reversibly. Future work could be undertaken to crystallise AspH in the presence of other nitrogen-containing anions, e.g. N_3. Additionally, exposure of AspH crystals to gaseous NO may provide further confirmation of the position of the O_2 binding site and investigate whether movement of residues around the active site to accommodate SCN^- binding could reveal O_2 sensing residues.

A system for generating slurries of AspH microcrystals was developed. This proved to be a robust method that yielded scalable volumes of AspH microcrystal slurries. The microcrystals were tested on fixed-target sample delivery systems at Diamond Light Source and structures of substrate and product bound AspH were obtained from serial X-ray crystallography methods. Unfortunately, there was a significant decrease in the resolution when moving from cryogenic macromolecular crystallography to room temperature microcrystal serial synchrotron experiments. It is widely recognised that the inherent flexibility at room temperature is decreased at cryogenic temperatures and gives a more static structure.[326] This is often exemplified by a large increase in the B-factors of crystal structures collected at room temperature compared to 100 K.

Another contribution to lower quality diffraction data possibly arises from the use of chipless chips. The background scattering is increased due to the increased volume of solvent surrounding the crystals, which is negligible in both single crystal data collection and when using silicon chips. It is also possible that the salts and PEG in the crystallisation precipitant solutions started to precipitate upon dehydration of the solution and contribute to noise from powder diffraction or diffuse scattering.

It is likely that the decrease in resolution is due to the room-temperature data collection.[278,326] This is largely in part due to the increased flexibility, as described above, but may also be due to deterioration of the crystal lattice. Screening for initial AspH crystallisation conditions only gave hits when trays were incubated at 4°C. This is likely due to the flexible hinge-loop region between

the oxygenase and TPR domains. AspH crystal stability at room temperature is dependent on precipitant conditions, but AspH crystals will eventually re-dissolve when left at room temperature. Incubating the AspH crystals at room temperature during chip construction and on the X-ray beam may cause the AspH crystals to begin to deteriorate. Commercial crystallisation screens were used to screen for crystallogenesis at room temperature, however these were unsuccessful. Further investigation into this area and establishing room temperature crystallisation conditions that may be more suitable for room temperature data collection.

It has been previously observed that some other protein crystals diffract to a higher resolution using XFEL beamlines, when compared to synchrotron sources.[327–329] An XFEL-derived structure of AspH in complex with 2OG, Fe and hFX-Asp_{103} was obtained and data were refined to 2.05 Å, which is a significant improvement in resolution compared to SSX-derived structures.

These initial results are promising and suggest that there is potential to perform tr-SFX experiments on AspH in future. However, catalysis in a time-resolved manner requires higher resolution to observe the binding of oxygen, the formation of Fe-O intermediates, and the movement of water molecules and side chains. Further work is required to develop a system of homogenous crystals that diffract to a higher resolution; if crystal diffraction is sufficient, experiments could include an oxygen exposed time course to resolve changes at the active site during catalysis.

4.5 Chapter acknowledgements

I am grateful to Dr Allen Orville, Dr Pierre Aller, Dr Anastasya Shilova, Dr Jos Kamps and Dr Tiankun Zhou at the XFEL Hub at Diamond Light Source for assistance with fixed-target serial crystallography experiments and data processing in all serial crystallography experiments. I would like to thank the groups of Dr Jan Kern and Dr Junko Yano for assistance at LCLS. Dr Stephen Marshall, Dr Patrick Rabe and Ms Mariska De Munnik crystallised AspH at LCLS and loaded the samples.

4.6 Supplementary information

Table S4-1. Data collection and refinement statistics for crystal structures of AspH presented in Figure 4-3 and Figure 4-4.

	AspH:2OG:Mn:hFX-Asp$_{103}$	AspH:Mn:2OG:SCN:hFX-Asn$_{103}$
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	43.1 - 1.842 (1.907 - 1.842)	43.18 - 1.6 (1.657 - 1.6)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	49.68 86.19 123.35	49.92 86.08 123.53
α, β, γ (°)	90 90 90	90 90 90
Total reflections	626785 (60807)	939233 (91958)
Unique reflections	46590 (4531)	71024 (7020)
Multiplicity	13.5 (13.4)	13.2 (13.1)
Completeness (%)	99.85 (98.74)	99.95 (99.89)
Mean I/sigma(I)	19.48 (2.11)	20.61 (2.30)
Wilson B-factor (Å^2)	29.04	25.27
R-merge	0.09151 (1.27)	0.06139 (1.179)
CC1/2	1 (0.877)	0.999 (0.815)
CC*	1 (0.967)	1 (0.948)
Reflections used in refinement	46580 (4528)	71019 (7020)
Reflections used for R-free	2190 (223)	3610 (331)
R-work	0.1577 (0.2391)	0.1652 (0.2971)
R-free	0.1944 (0.2584)	0.1818 (0.3050)
CC(work)	0.967 (0.904)	0.965 (0.903)
CC(free)	0.956 (0.906)	0.963 (0.877)
Number of non-hydrogen atoms	3973	4220
macromolecules	3583	3591
ligands	117	32
solvent	337	607
Protein residues	447	447
RMS(bonds)	0.010	0.014
RMS(angles)	1.02	1.26
Ramachandran favored (%)	98.87	99.10
Ramachandran allowed (%)	1.13	0.90
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	1.09	0.55
Clashscore	4.74	2.54
Average B-factor	34.00	28.37
Macromolecules (Å^2)	33.05	26.84
Ligands (Å^2)	54.80	26.99
Solvent (Å^2)	40.84	37.53

Table S4-2. Data collection and refinement statistics for crystal structures of AspH presented in Figure 4-4 and Figure 4-5.

	AspH:2OG:Mn:hFX-Asn$_{103}$	AspH:2OG:Fe:hFX-Asp$_{103}$
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	43.04 - 1.8 (1.864 - 1.8)	50.29 - 1.98 (2.051 - 1.98)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	49.74 85.85 123.64	49.62 86.04 123.94
α, β, γ (°)	90 90 90	90 90 90
Total reflections	671482 (67304)	504607 (50021)
Unique reflections	49877 (4911)	37749 (3698)
Multiplicity	13.5 (13.7)	13.4 (13.5)
Completeness (%)	99.89 (99.86)	99.89 (99.68)
Mean I/sigma(I)	15.60 (1.40)	11.47 (1.01)
Wilson B-factor (Å^2)	31.62	33.36
R-merge	0.09335 (2.083)	0.1262 (1.237)
CC1/2	0.999 (0.695)	0.999 (0.766)
CC*	1 (0.906)	1 (0.931)
Reflections used in refinement	49865 (4905)	37708 (3691)
Reflections used for R-free	2538 (254)	2000 (196)
R-work	0.1866 (0.3360)	0.1971 (0.3113)
R-free	0.2097 (0.3396)	0.2264 (0.3342)
CC(work)	0.956 (0.836)	0.954 (0.846)
CC(free)	0.963 (0.773)	0.947 (0.736)
Number of non-hydrogen atoms	3895	3812
macromolecules	3567	3573
ligands	40	15
solvent	302	228
Protein residues	447	447
RMS(bonds)	0.011	0.006
RMS(angles)	1.11	1.18
Ramachandran favored (%)	98.65	98.65
Ramachandran allowed (%)	1.35	1.35
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.28	0.27
Clashscore	3.82	5.10
Average B-factor	39.55	37.26
Macromolecules (Å^2)	39.02	37.01
Ligands (Å^2)	46.71	26.43
Solvent (Å^2)	45.12	41.64

Table S4-3. Data collection and refinement statistics for crystal structures of AspH presented in Figure 4-5.

	AspH:2OG:Fe:hFX-Asn$_{103}$	AspH:2OG:Fe:SCN$^-$:hFX-Asn$_{103}$
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	42.91 - 1.9 (1.968 - 1.9)	46.08 - 1.7 (1.76 - 1.7)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	49.58 85.62 123.3	49.64 85.89 123.82
α, β, γ (°)	90 90 90	90 90 90
Total reflections	567557 (56280)	784263 (65599)
Unique reflections	42162 (4099)	59116 (5825)
Multiplicity	13.5 (13.7)	13.3 (11.3)
Completeness (%)	99.95 (99.95)	99.94 (99.69)
Mean I/sigma(I)	15.08 (1.40)	17.29 (1.28)
Wilson B-factor (Å^2)	35.02	24.02
R-merge	0.1031 (1.981)	0.07891 (0.8751)
CC1/2	0.999 (0.66)	0.999 (0.864)
CC*	1 (0.892)	1 (0.963)
Reflections used in refinement	42148 (4097)	59096 (5821)
Reflections used for R-free	2000 (195)	2000 (197)
R-work	0.1998 (0.3111)	0.1724 (0.2766)
R-free	0.2212 (0.3290)	0.1968 (0.3251)
CC(work)	0.956 (0.804)	0.965 (0.884)
CC(free)	0.941 (0.767)	0.930 (0.854)
Number of non-hydrogen atoms	3785	4030
macromolecules	3573	3582
ligands	15	18
solvent	201	434
Protein residues	447	447
RMS(bonds)	0.014	0.006
RMS(angles)	1.35	1.13
Ramachandran favored (%)	98.42	99.10
Ramachandran allowed (%)	1.58	0.90
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.27	1.10
Clashscore	4.25	2.54
Average B-factor	40.98	30.37
Macromolecules (Å^2)	40.81	29.31
Ligands (Å^2)	31.64	32.46
Solvent (Å^2)	44.49	39.07

Table S4-4. Data collection and refinement statistics for crystal structures of AspH presented in Figure 4-6 and Figure 4-7.

	AspH:2OG/SIN:Fe:3-hydrox-Asp$_{103}$	AspH:2OG/SIN:Fe:3-hydroxy-Asn$_{103}$
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	43.5 - 1.9 (1.968 - 1.9)	41.52 - 1.9 (1.968 - 1.9)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	50.26 86.83 123.33	50.01 88.21 123.1
α, β, γ (°)	90 90 90	90 90 90
Total reflections	584188 (58280)	588916 (57148)
Unique reflections	43369 (4268)	43745 (4289)
Multiplicity	13.5 (13.7)	13.5 (13.3)
Completeness (%)	99.97 (99.95)	99.87 (99.30)
Mean I/sigma(I)	15.04 (1.61)	12.68 (1.20)
Wilson B-factor (Å^2)	33.67	33.73
R-merge	0.1019 (1.698)	0.1227 (2.047)
CC1/2	0.999 (0.695)	0.999 (0.686)
CC*	1 (0.905)	1 (0.902)
Reflections used in refinement	43363 (4266)	43707 (4271)
Reflections used for R-free	2000 (197)	2231 (243)
R-work	0.1887 (0.2838)	0.2052 (0.3905)
R-free	0.2046 (0.2955)	0.2373 (0.4219)
CC(work)	0.960 (0.844)	0.958 (0.819)
CC(free)	0.960 (0.859)	0.939 (0.717)
Number of non-hydrogen atoms	3951	3841
macromolecules	3623	3565
ligands	85	42
solvent	284	248
Protein residues	447	446
RMS(bonds)	0.003	0.002
RMS(angles)	0.60	0.58
Ramachandran favored (%)	98.64	98.86
Ramachandran allowed (%)	1.36	1.14
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.80	1.10
Clashscore	3.16	3.54
Average B-factor	40.60	43.41
Macromolecules (Å^2)	40.27	43.13
Ligands (Å^2)	40.00	45.27
Solvent (Å^2)	44.88	47.15

Table S4-5. Data collection and refinement statistics for crystal structures of AspH presented in Figure 4-8.

	AspH:Succinate:Mn:hFX-Asp$_{103}$
Wavelength (Å)	0.9763
Resolution range (Å)	50.41 - 1.8 (1.865 - 1.8)
Space group	P 2_1 2_1 2_1
Unit cell	
a, b, c (Å)	50.09 86.43 124.12
α, β, γ (°)	90 90 90
Total reflections	679552 (69348)
Unique reflections	50757 (4980)
Multiplicity	13.4 (13.9)
Completeness (%)	99.87 (99.26)
Mean I/sigma(I)	19.30 (2.02)
Wilson B-factor (Å^2)	31.60
R-merge	0.07521 (1.454)
CC1/2	1 (0.807)
CC*	1 (0.945)
Reflections used in refinement	50693 (4967)
Reflections used for R-free	1999 (196)
R-work	0.1914 (0.3467)
R-free	0.2254 (0.3785)
CC(work)	0.911 (0.518)
CC(free)	0.861 (0.560)
Number of non-hydrogen atoms	3927
macromolecules	3600
ligands	84
solvent	287
Protein residues	446
RMS(bonds)	0.004
RMS(angles)	1.00
Ramachandran favored (%)	98.87
Ramachandran allowed (%)	1.13
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.54
Clashscore	2.79
Average B-factor	35.37
Macromolecules (Å^2)	34.91
Ligands (Å^2)	44.05
Solvent (Å^2)	39.95

Table S4-6. Data collection and refinement statistics for crystal structures of AspH obtained by fixed target serial synchrotron crystallography (Figure 4-12). DLS = Diamond Light Source.

	AspH:Mn:hFX-Asp$_{103}$	**AspH:Fe:3-hydroxy-hFX-Asp$_{103}$**
Data collection method	Fixed target chipless chip at DLS	Fixed target chipless chip at DLS
Resolution range (Å)	62.43 - 3.2 (3.315 – 2.7)	73.33 - 2.6 (2.693 - 2.6)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	50.90 91.28 124.87	50.83 90.83 124.25
α, β, γ (°)	90 90 90	90 90 90
Reflections	10050 (961)	18059 (1661)
Multiplicity	6.6 (2.5)	64.03 (11.55)
Completeness (%)	97.97 (91.56)	98.07 (90.54)
Mean I/sigma(I)	1.135 (0.353)	3.5 (1.05)
Wilson B-factor (Å^2)	64.96	54.66
R_{split}	42.8 (107)	17.9 (58)
CC1/2	0.83 (0.05)	0.96 (0.47)
Reflections used in refinement	9895 (889)	17995 (1627)
Reflections used for R-free	991 (91)	1800 (162)
R-work	0.2402 (0.3239)	0.1770 (0.2851)
R-free	0.2965 (0.3785)	0.2482 (0.3771)
Number of non-hydrogen atoms	3555	3586
macromolecules	3536	3560
ligands	1	1
Protein residues	447	446
RMS(bonds)	0.003	0.013
RMS(angles)	0.49	1.52
Ramachandran favored (%)	95.49	96.58
Ramachandran allowed (%)	4.29	3.19
Ramachandran outliers (%)	0.23	0.23
Rotamer outliers (%)	0.28	2.47
Clashscore	4.48	5.84
Average B-factor	55.01	60.35
macromolecules	55.04	60.44
ligands	42.04	77.14
solvent	50.55	46.77

Table S 4-7. Data collection and refinement statistics for the AspH crystal structure obtained at Linac Coherent Light Source (LCLS) at room temperature (Figure 4-15).

	AspH:2OG:Fe:hFX-Asp$_{103}$
Data collection method	XFEL beamline at LCLS
Resolution range (Å)	26.6 - 2.05 (2.123 - 2.05)
Space group	P 2_1 2_1 2_1
Unit cell	
a, b, c (Å)	50.79 88.16 125.16
α, β, γ (°)	90 90 90
Reflections	36053 (3567)
Multiplicity	44.4 (7.9)
Completeness (%)	99.89 (100.00)
Mean I/sigma(I)	4.2 (0.5)
Wilson B-factor (Å^2)	35.61
R_{int}	14.6 (114)
CC1/2	0.99 (0.13)
Reflections used in refinement	36037 (3567)
Reflections used for R-free	1387 (137)
R-work	0.1779 (0.3089)
R-free	0.2130 (0.3255)
Number of non-hydrogen atoms	3721
macromolecules	3575
ligands	15
solvent	135
Protein residues	447
RMS(bonds)	0.006
RMS(angles)	0.76
Ramachandran favored (%)	98.42
Ramachandran allowed (%)	1.58
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.36
Clashscore	2.97
Average B-factor	47.41
macromolecules	47.44
ligands	30.49
solvent	48.02

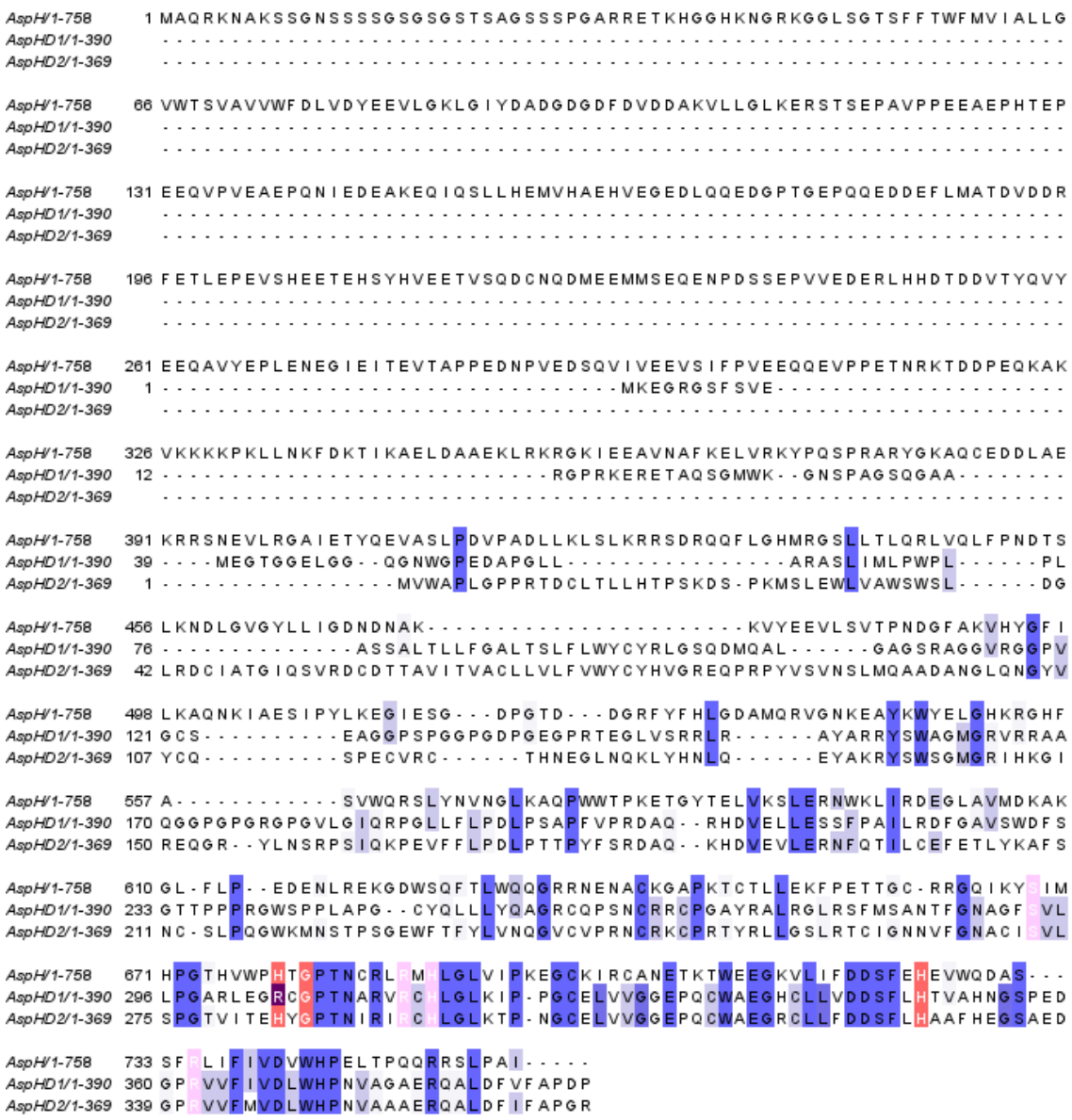
Chapter 5 | AspH homologue: Aspartate-β-hydroxylase domain-containing protein 1

5.1 Introduction

2OG-dependent oxygenases typically have a conserved Fe(II) binding motif at the active site, comprising a HXD/E…H triad.[330] Interestingly, however, this triad is not present in all 2OG-dependent oxygenases with the most common divergence from the triad being observed in the 2OG-dependent halogenases.[331] In the halogenases, the acidic residue of the triad is substituted for a glycine residue, thus creating a vacant coordination site at the Fe(II) centre allowing for halogen ion binding during catalysis;[330] note that 2OG-dependent halogenases have not yet been identified in humans. Deviations from the triad are reported to be less common in 2OG-dependent enzymes that catalyse reactions other than halogenations, with AspH being a prominent example of a human hydroxylase, while Plant Homeodomain (PHD) Finger Protein 2 (PHF2) and Hairless are proposed to be human pseudoenzymes that do not possess the HXD/E…H triad.[122,123,332]

AspH is a 2OG-dependent hydroxylase which employs the same metal binding pattern as the 2OG-dependent halogenases, i.e. an HXG…H dyad, but which does not catalyse protein halogenation in vitro. As shown in Chapter 2, the metal binding site of AspH is remarkably flexible and its catalysis can proceed with only a single metal binding residue originating from the protein. These observations raise questions as to whether other members of the 2OG-dependent superfamily that bear an HXG…H motif or an XXG…H/HXG…X motif exist, but have previously been missed from searches that are only based on the metal-binding triad alignments.

Aspartate-β-hydroxylase domain containing proteins 1 and 2 (AspHD1 and AspHD2) are predicted potential human 2OG-oxygenases. Sequence alignments suggest that both AspHD1 and AspHD1 possess Fe(II)-binding motifs that are unusual for 2OG-dependent oxygenases (Figure 5-1). AspHD1 is predicted to have a dyad of RXG…H residues at the active site, while AspHD2 may exhibit an HXG…H dyad that is also observed in AspH (Figure 5-2a).


AspH/1-758 1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMVIALLG
AspHD1/1-390 - - -

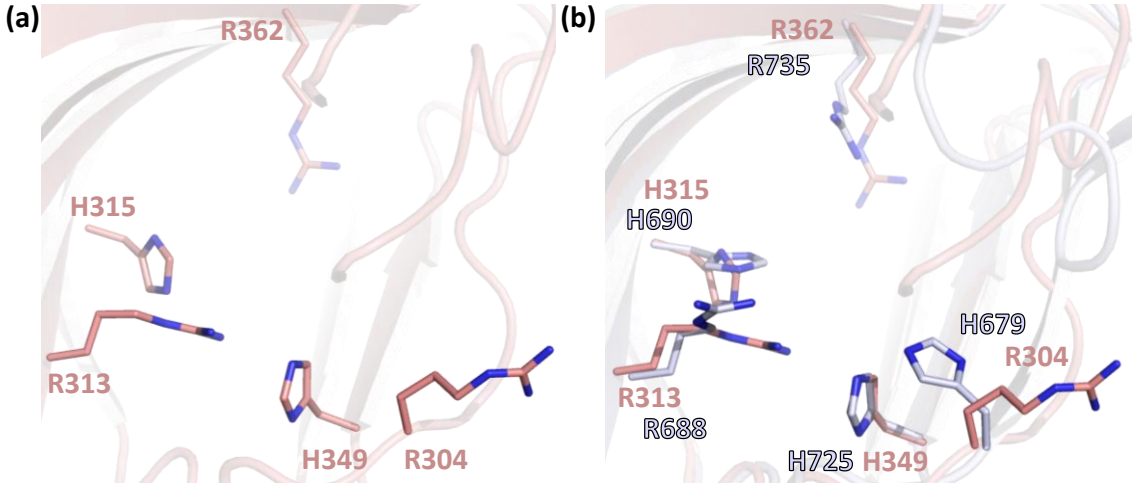


Figure 5-2. The AlphaFold predicted active site structure of AspHD1 with an unusual RXG…H Fe(II)-binding motif. (a) The predicted active site structure of AspHD1. (b) An overlay of the predicted AspHD1 active site structure with the active site of AspH.

The relatively high pKa of the positively-charged guanidinium side chain of L-arginine results in coordination properties which, in general, are expected to disfavour effective metal ligation. Analysis of metal binding residues in the PDB revealed that only three arginine-metal interactions have been observed in reported protein structures, that is in *E. coli* biotin synthase (BioB), a human carbonic anhydrase I variant and *Bacillus caldovelox* L-arginase (Figure 5-3).[333] In the structures for BioB and *B. caldovelox* L-arginase there are deviations from ideal geometries which suggest the metal-side chain interactions are relatively weak.[333]

E. coli BioB is an S-adenosylmethionine (SAM)-dependent enzyme that catalyses the final radical mediated reaction in the biotin biosynthetic pathway.[334] BioB contains an Fe_4S_4 cluster that is characteristic of the SAM superfamily, and additionally an Fe_2S_2 cluster which is postulated to act as a sulfur source.[335] A BioB crystal structure has revealed an unprecedented coordination environment around the Fe_2S_2 cluster, comprising three cysteine residues and an arginyl guanidinium group (R260) (Figure 5-3a).[335] In other Fe_2S_2 cluster-coordinating proteins, a cysteine or histidine residue is involved in cluster ligation rather than an arginyl residue.[335] Berkovitch et al. noted that there is evidence that guanidines can bind some metal ions in solution at neutral pH, and proposed that the unusual environment of the BioB active site would alter the pKa of R260,

resulting in an uncharged guanidino group which can act as a metal binding ligand.[335] It is also possible that a protonated arginine residue can coordinate a metal ion, but via a guanidinium nitrogen that is not protonated.

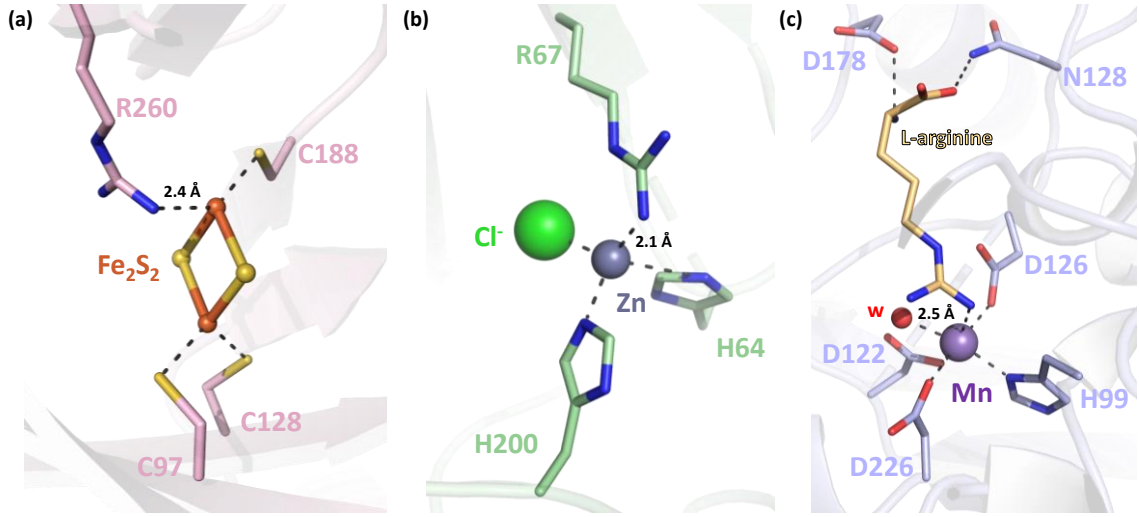


Figure 5-3. Evidence for arginyl guanidinium as a metal-coordinating group in proteins. (a) The coordination of the Fe_2S_2 cluster in BioB (PDB 1R30)[335] (b) The active site of carbonic anhydrase I variant 1 in the presence of an excess of Zn(II) (PDB 1JV0)[336] (c) Substrate binding in *B. caldovelox* L-arginase (PDB 3CEV).[337]

Carbonic anhydrase (CA) catalyses the interconversion of carbon dioxide and bicarbonate.[11] A naturally occurring human variant of CA I, contains a single point mutation that results in the substitution of H67 for R.[336] Crystallisation studies have revealed that in the absence of stochiometric amounts of Zn(II), R67 is not involved in metal-binding, and the H67R variation does not alter the overall structure of Ca I. By contrast, in the presence of an excess of Zn(II), the corresponding $Zn(II)_2$-Ca I complex was obtained and in which the side chain of R67 adopts a conformation enabling the coordination to a Zn(II) ion (Figure 5-3b).[336] Ferraroni et al. have proposed that H64 and H200, which are adjacent to R67, may enhance the nucleophilicity of the R67 guanidinium group so enabling it to act as a metal coordinating residue.[336]

Mn(II)-dependent *B. caldovelox* L-arginase catalyses the hydrolysis of L-arginine to L-ornithine and urea.[337] Crystallographic analysis of L-arginase in complex with substrate L-arginine and Mn(II) indicated that the substrate may be coordinated to the Mn(II) ion through a guanidinium

nitrogen.[337] Subsequent reanalysis of this structure revealed that the Mn(II)-arginine interaction is longer and distorted compared to inorganic model complexes deposited in the Cambridge Structural Database and concluded that it is unlikely that the arginine has a significant contribution in binding the Mn(II) ion.[333]

5.2 Aims of the research described in this chapter

Biomedical research is likely to continue to inform of the importance of AspHD1 and AspHD2. Additionally, characterisation of the predicted unusual Fe(II)-binding motif in AspHD1 may have implications for the assignment of the 2OG-oxygenase superfamily. However, there are no reports on the structural or biochemical characterization of AspHD1 or AspHD2. Thus, the work presented in this chapter aimed to (i) define a system for the efficient recombinant expression of human AspHD1 to enable its biochemical characterisation and (ii) to assess the potential of arginine guanidinium groups as a metal-coordinating moiety using the AspH H679R variant as a model system.

5.3 Results and discussion

5.3.1 AspHD1 construct design

The structure and biochemical functions of AspHD1 are unknown, and its relevance in human disease is incompletely understood. Thus, to investigate the Fe(II) binding mode in AspHD1 and to inform on whether AspHD1 is catalytically active, studies were initiated to obtain a soluble and crystallisable AspHD1 construct. AspHD1 expression trials were carried out using *E. coli* as an expression system, because recombinant protein production in *E. coli* is typically fast, inexpensive and it is easy to screen multiple constructs and conditions. Additionally, methods for the synthesis of *E. coli* plasmids are efficient, with many online tools and laboratory kits available to manipulate fragments of DNA and create desired plasmid. The wild-type $AspH_{315-758}$ (wt AspH) construct used in this thesis was also recombinantly produced in *E. coli*. Considering the relatively high sequence similarity of AspH and AspHD1, *E. coli* appeared to be a suitable system for trial expression.

The gene encoding for human AspHD1$_{1\text{-}390}$ was synthesized by GeneArt (Life Technologies) with a codon-optimised sequence for *E. coli* codon usage. Translation of rare codons in *E. coli* can impair ribosomal activity and consequently lead to misfolded protein.[338] Three different AspHD1 gene constructs of different length (AspHD1$_{1\text{-}390}$, AspHD1$_{96\text{-}390}$, AspHD1$_{206\text{-}390}$) were cloned into a modified pet28a(+) vector by means of Gibson Assembly (AspHD1$_{1\text{-}390}$ and AspHD1$_{96\text{-}390}$), or deletion of bases from the AspHD1$_{1\text{-}390}$ construct (Table 5-1).[339] The pet28a(+) vector was previously modified to substitute the thrombin protease recognition sequence for a 3C-protease recognition sequence to improve the efficiency of His$_6$-tag cleavage during purification, as described in Chapter 4.

The three constructs made were: (i) the full-length gene encoding the AspHD1$_{1\text{-}390}$ construct to assess expression levels of the whole protein, (ii) a construct in which the predicted N-terminal transmembrane domain was removed (AspHD1$_{96\text{-}390}$) to avoid potential issues with expression of transmembrane proteins in *E. coli*, and (iii) a shorter construct (AspHD1$_{206\text{-}390}$) containing only the predicted oxygenase domain. Analysis of the AspHD1 sequencing using the tool Predictor of Natural Disordered Regions (PONDR) predicted a highly disordered region in the mid-section of the protein and this may impair the synthesis of correctly folded protein in *E. coli* (Figure 5-4).[340]

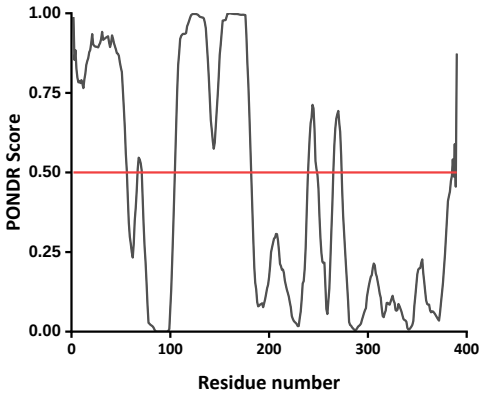


Figure 5-4. Predicted disorder within the AspHD1 sequence, calculated using the online tool Predictor of Natural Disordered Regions (PONDR).[340] A score above 0.5 is indicative of a highly disordered region within the protein. Predicted transmembrane region residues 73-95, predicted cytoplasmic region residues 96-390. The shortest AspHD1 construct was designed to eliminate the predicted highly disordered region from 100 to 200 while preserving the predicted oxygenase domain.

Table 5-1. AspHD1 constructs and cloning methods used. The modified 3C-pET28a(+) vector refers to a standard pET28a(+) vector (Addgene 69864-3) in which the thrombin recognition site and subsequent multiple cloning site were substituted for a 3C-protease recognition site. Full sequence available in the appendix.

Name	Cloning method	Vector	Tag	MW (kDa)
$AspHD1_{1-390}$	Gibson assembly	3C- pET28a(+)	His_6	44.7
$AspHD1_{96-390}$	Gibson assembly	3C- pET28a(+)	His_6	35.4
$AspHD1_{206-390}$	PCR deletion of bases from $AspHD1_{96-390}$ vector	3C- pET28a(+)	His_6	24.4

5.3.2 Initial AspHD1 expression trials

The vectors encoding the genes for $AspHD1_{1-390}$, $AspHD1_{96-390}$ and $AspHD1_{206-390}$ were transformed into BL21 DE3 competent cells. 1 mL cultures were grown at 37°C until an OD_{600} of 0.4-0.8 was reached. Recombinant protein production was induced by addition of IPTG to a final concentration of 0.1 mM. Cultures were incubated at 18°C for 16 hours to avoid issues with low protein solubility often observed at higher induction temperatures.

1 mL of each culture was pelleted and cell pellets were lysed using detergent (BugBuster) and lysonase protein reagent (Sigma). Following incubation at room temperature for 20 min, the samples were centrifuged (14,462 x *g*, 15 min, 4°C) and the soluble fraction was removed. The insoluble pellet was resuspended in 8 M aqueous urea. The soluble and insoluble fractions were run on two equivalent SDS-PAGE gels. One gel was stained with QuickBlue and one gel was used for Western Blot analysis (Figure 5-5). QuickBlue staining of the SDS-Page gel allows for rapid visualization of overexpressed bands in the soluble and insoluble fractions. Western Blot is a slower method but allows for detection of specific tags in the sample in case of low levels of expression. Western Blot analysis used a primary antibody of an anti-His_6 Mouse antibody and a secondary antibody of a GFP tag to allow for detection of the His_6 tag included in all three AspHD1 constructs.

The stained SDS-Page gels revealed overexpression of AspHD1 constructs in the insoluble samples (Figure 5-5a). The soluble fractions contained a mix of bands and it was not possible to determine

whether AspHD1 was expressed in soluble form from the gel. Western Blot analysis confirmed that the likely overexpressed bands in the insoluble fractions contained a His_6 tag and were therefore the correct constructs (Figure 5-5b). However, the tested AspHD1 constructs were not observed in the soluble protein fractions.

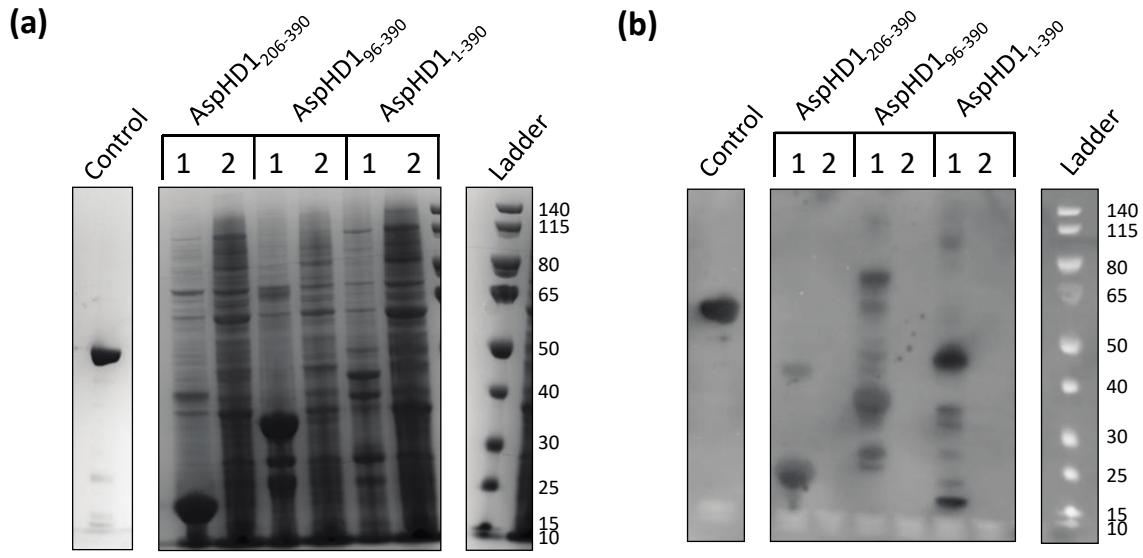


Figure 5-5. SDS-Page (a) and Western blot (b) analysis of soluble and insoluble protein fractions of three AspHD1 constructs expressed in BL21 DE3 (0.1 mM IPTG, 18°C, 16 hours). 1 = insoluble fraction, 2 = soluble fraction. Control = 1 mg/ml purified His_6-$AspH_{315-758}$.

Although *E. coli* can be used as an expression system to produce soluble protein, induction of expression of a target gene for protein synthesis can also lead to the formation of incorrectly folded proteins, which consequently aggregate into insoluble inclusion bodies.[341] The AspHD1 sequence contains eleven cystine residues, some of which are likely to form disulfide bonds. Disulfide bonds can be difficult to synthesize correctly in *E. coli* due to the reducing environment of the cytoplasm where most are proteins synthesized, and therefore proteins containing many disulfide bonds have a propensity to be misfolded.[342] There are possible ways to overcome solubility issues with protein expression in *E. coli* and address the issue of correctly folding disulfide bonds, including *E. coli* strain optimisation and the use of fusion tags.[342] Alternatively, inclusion bodies can be purified and re-folded, however this process needs optimizing for each protein and there are a large number of variables that need to be considered; there is also a history of failure for refolding of 2OG-

oxygenases.[343] Therefore, optimisation of *E. coli* strain and the addition of fusion tags were initially explored to remedy the problem of insoluble recombinant protein production.

5.3.3 AspHD1 expression in *E. coli* optimisation: the competent cell strain

E. coli strains have been developed to overcome common issues associated with expression of proteins in *E. coli*, including translation of rare codons, correct folding of disulfide bonds and expression of membrane proteins. SHuffle® T7 competent *E. coli* cells have been engineered to express the disulfide bond isomerase DsbC to facilitate production of correctly folded disulfide-containing proteins.[344] Lemo21 cells allow for tunable levels of T7 protein expression using L-rhamnose and have been used to produce membrane proteins.[345] Arctic Express cells are engineered to co-express low temperature adapted chaperone proteins, and allow for induction temperatures as low as 4°C which may help increase soluble protein yields.

Thus, as it was hypothesized that the folding of disulfide bonds may the cause of misfolded AspHD1 in *E. coli*, vectors encoding AspHD1$_{1\text{-}390}$, AspHD1$_{96\text{-}390}$ and AspHD1$_{206\text{-}390}$ were transformed into Shuffle T7 competent cells. Small scale expression trials were performed as described previously and protein expression was assessed by His$_6$ tag antibody staining.

Initial results from analysis of cell lysate and pellets on an SDS-Page gel and Western Blot appeared to indicate that some AspHD1$_{1\text{-}390}$ and AspHD1$_{96\text{-}390}$ may be present in the soluble portion of the cell lysate (Figure 5-6). A larger scale expression trial was attempted with both constructs; however, unfortunately the initial results could not be reproduced and it is possible that the initial Western Blot reflected a false positive result, possibly resulting from mis-loading samples into wells.

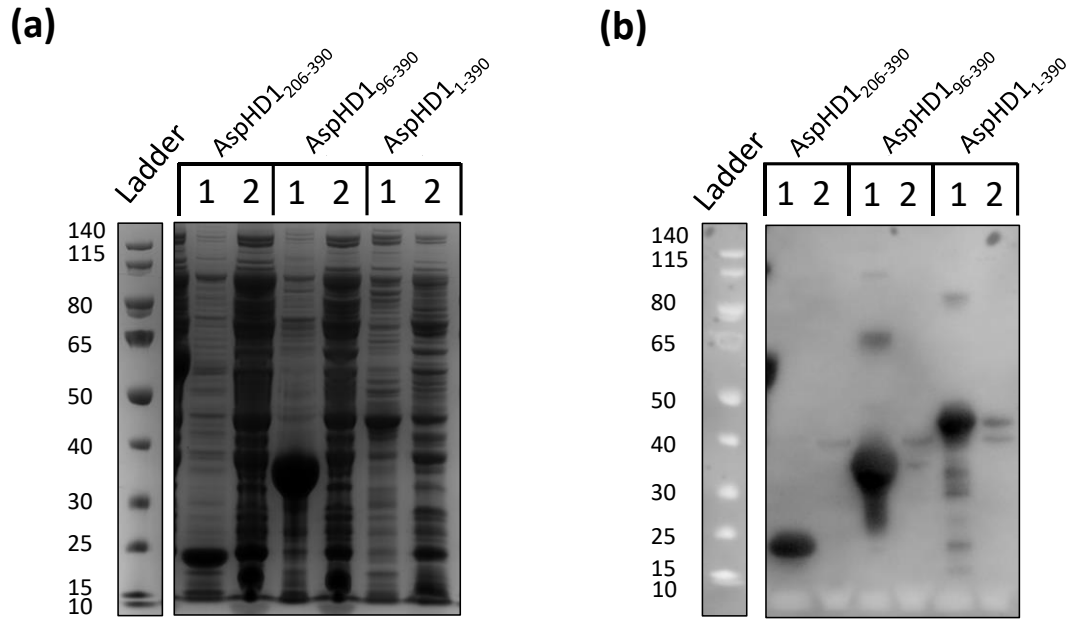


Figure 5-6. SDS-Page (a) and Western Blot (b) analysis of soluble and insoluble cell lysate fractions of three AspHD1 constructs expressed in SHuffle® competent cells. 1 = insoluble fraction, 2 = soluble fraction.

Expression of the AspHD1$_{206-390}$ construct was additionally trialed in Lemo21 and Arctic Express competent cells; however, no protein production in the soluble fraction was observed with any of the competent cell strains used. It was thus concluded that optimizing competent cell strain is not a successful strategy to improve expression of correctly folded AspHD1, regardless of the construct length. Hence, the fusion tag was optimised next.

5.3.4 AspHD1 expression in *E. coli* optimisation: fusion tags

The addition of fusion tags to proteins has been shown to enhance the solubility of difficult to express constructs.[346] Some of the most commonly used tags include maltose binding protein (MBP), small ubiquitin-like modifier (SUMO), glutathione (GST) and thioredoxin.[346] However, the use of fusion tags can also negatively affect protein production in *E. coli* as the activity of the desired protein may be compromised by the presence of a large fusion tag and removal of the tag during the purification process may leave the protein prone to aggregation.

The shortest AspHD1$_{206-390}$ construct was chosen for further trials with fusion tags, as the construct did not include the transmembrane domain or the predicted highly disordered region, both

features that may also have been hindering production of correctly folded soluble AspHD1 in *E. coli*. The sequences for MBP and SUMO fusion tags were cloned into the $AspHD1_{206-390}$ construct using Gibson Assembly to give vectors encoding for the MBP-His_6-$AspHD1_{206-390}$ and SUMO-His_6-$AspHD1_{206-390}$ constructs (Table 5-2).[339] Fusion tags were cloned into the vectors at the N terminal of the protein sequence as the calculated structures predicted the C-terminal to be near the active site. Constructs were transformed into BL21 DE3 competent cells and AspHD1 protein production was checked for as described in 5.3.2 (Figure 5-7).

Table 5-2. Constructs of $AspHD1_{206-390}$ with fusion tags

Construct	Expected MW (kDa)	Resistance Marker	Vector
His_6-MBP-$AspHD1_{206-390}$	62.9	KAN	3C-pET28a(+)
His_6-SUMO-$AspHD1_{206-390}$	33.0	KAN	3C-pET28a(+)

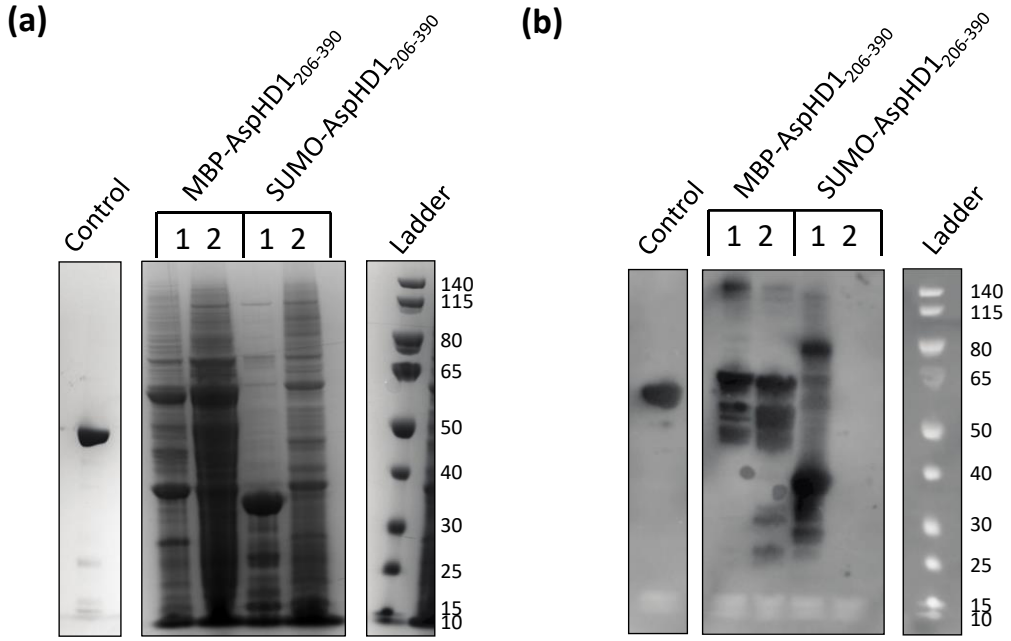


Figure 5-7. Western Blot analysis of $AspHD1_{206-390}$ solubility tagged constructs. 1 = insoluble protein fraction, 2 = soluble fraction, control = His_6-$AspH_{315-758}$. Expression of soluble His_6-MBP-$AspHD1_{206-390}$ can be observed in MBP lane 2.

A relatively high expression level was observed with the N-terminal SUMO-tagged SUMO-His_6-$AspHD1_{206-390}$ construct, which, however, was entirely observed in the insoluble fraction (Figure

5-7). His_6–MBP–$AspHD1_{206-390}$ expressed in both the soluble and insoluble fractions (Figure 5-7). Around 50% total protein was soluble, and this could potentially be optimised to improve yields. The His_6-MBP-$AspHD1_{206-390}$ construct was thus taken forward to probe larger scale production and purification.

5.3.5 Purification of MBP-His_6-$AspHD1_{206-390}$ produced in *E. coli*

The use of detergents, i.e. BugBuster, in expression trials is useful for quick analysis of a large number of samples; however, the results can be misleading as detergents have a tendency to solubilise the majority of cell debris. Therefore, a small-scale purification of MBP-His_6-$AspHD1_{206-390}$ using sonication as a manual cell lysis method was attempted. Following this, affinity purification exploiting the MBP tag was trialled. Buffers used in the purification trial are outlined in Table 5-3.

A plasmid encoding MBP-His_6-$AspHD1_{206-390}$ was transformed into BL21 DE3 competent cells. An overnight starter culture was used to inoculate 600 mL of 2YT media, supplemented with 30 µg/mL kanamycin. The culture was grown at 37°C with shaking (180 rpm) until an O.D. of 0.6 was reached. Expression of the MBP-His_6-$AspHD1_{206-390}$ gene was induced in with 0.1 mM IPTG and the culture was incubated at 18°C overnight. Cells were pelleted by centrifugation (8 minutes, 11867 x *g*, 4°C).

The cell pellet was resuspended in 100 mL $30g^{-1}$ of MBP Buffer A (Table 4), with DNAse I and protease inhibitor tablets. Cell lysis was achieved by high pressure cell disruption (French Press 25 kPSi) and the lysate was clarified by centrifugation (30 min, 58,387 x g, 4°C). The supernatant was filtered (0.45 µm filters) and loaded onto a 5 mL MBPTrap (Cytivia), pre-equilibrated with MBP buffer A. After lysate loading was complete, the column was washed with MBP buffer A until the UV trace dropped below 30 mAU. The MBP-tagged protein was then eluted by washing the column with 100% MBP buffer B (Figure 5-8a). The loaded lysate, the flow through from loading, and the elution fractions were analysed by SDS-PAGE (Figure 5-8b).

Table 5-3. Composition of buffers used in purification of His_6-MBP-$AspHD1_{206-390}$.

Buffer	Composition	pH
MBP Buffer A	20 mM Tris 200 mM NaCl 1 mM EDTA	7.4
MBP Buffer B	20 mM Tris 200 mM NaCl 1 mM EDTA 10 mM Maltose	7.4

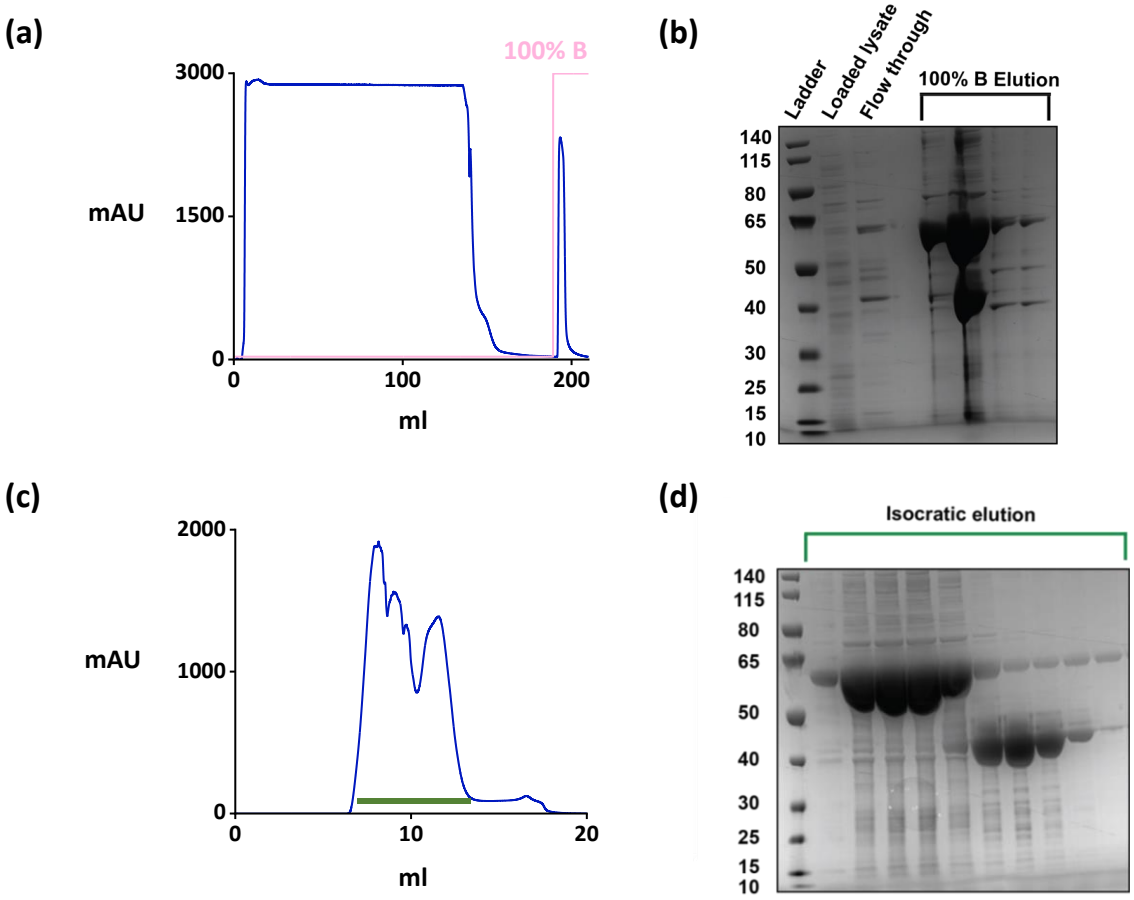


Figure 5-8. Purification of MBP-His_6-$AspHD1_{206-390}$. (a) UV trace of loading of lysate onto MBP-Trap followed by washing and then elution of protein with 100% MBP Buffer B. (b) SDS-PAGE of loaded fractions and elution fractions from MBP-Trap. (c) UV-trace of isocractic elution of purification of protein fractions on a GL10/300 S75 size exclusion column. (d) Corresponding SDS-PAGE gel of fractions eluted.

The SDS-Page analysis indicated an overexpressed mass similar to the expected MBP-His_6-$AspHD_{206-390}$ construct mass (62.9 kDa). One fraction precipitated out of solution after elution

which was then centrifuged and the soluble portion was combined with other fractions. The elution fractions were concentrated to 0.5 mL and loaded onto a GL 10/300 S75 size-exclusion column for further purification. The isocratic elution (Figure 5-8c) revealed that proteins of different masses were present in the mixture. Analysis of the fractions by SDS-Page (Figure 5-8d) revealed two protein species being present in the fractions; one larger of a mass ~60 kDa, and one smaller of a mass ~40 kDa. Unfortunately, following the size exclusion step, protein in all fractions precipitated from solution and soluble protein could not be recovered from any fraction.

The band in the SDS-Page gel located around at 40 kDa could correspond to a truncated His6-MBP construct (expected MW = 43 kDa) as the protein was only eluted from the MBP-Trap after the addition of maltose and therefore may contain the MBP-tag. This observation suggests that translation may be incomplete during the production of the target construct in *E. coli*, and, although translation of the initial part of the gene encoding for the His_6 and MBP tags is possible, synthesising the target protein joined to the tag is unfavourable in *E. coli*. The fractions containing the 63 kDa construct were not subjected to further analysis, e.g. mass spectrometry, due to aforementioned issues with precipitation. However, the positive Western blot result in the expression trials and the fact that this mass was only eluted from the MBP-Trap following a wash with high maltose concentration provides evidence that this was the desired construct. Buffer optimisation, including the use of appropriate additives can improve protein stability and solubility during the purification process. Osmolytes such as glycerol can prevent aggregative, arginine can bind directly to residues in the target protein and assist in stabilising the target, and reducing agents can prevent oxidation and consequent aggregation of proteins containing multiple cysteine residues.[347–349] Unfortunately, due to time constraints on the project these optimisation processes were not attempted, however this avenue can be advisable to pursued in future.

Evidence from the expression trials described in this chapter indicate that *E. coli* is likely an unsuitable expression system for the recombinant production of AspHD1. It is advisable to attempt

expression trials in another system, e.g. baculovirus or mammalian cells. Both systems are likely more suited to correctly synthesising disulphide bonds which is often an issue when attempting to produce correctly folded soluble proteins in in *E. coli*.

5.3.6 A model system to assess arginine as a metal binding residue

In order to investigate the non-canonical Fe(II)-binding motif predicted to be present in AspHD1, the analogous H679R $AspH_{315-758}$ variant (henceforth referred to as H679R AspH) was produced. The H679R mutation was introduced into the gene encoding for wt His_6-$AspH_{315-758}$ by site directed mutagenesis (the mutagenesis was performed by Dr Inga Pfeffer). The H679R AspH variant has the same predicted binding dyad as AspHD1 and could therefore be a suitable model system to investigate whether an arginine can ligate a metal ion at the active site.

The H679R variant was produced recombinantly in *E. coli* and purified as described for other AspH variants in Chapter 2 (Figure 5-9).[236] wt AspH can be overexpressed in *E. coli* and point-mutations to the enzyme do not significantly affect the expression levels. Therefore, it was possible to successfully purify the H679R variant and use it as a model system for coordination in AspHD1.

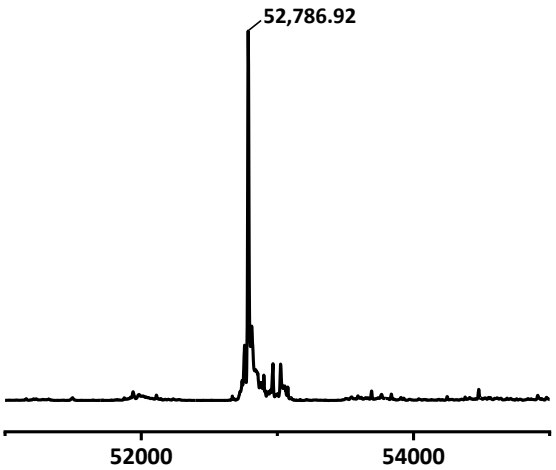


Figure 5-9. The deconvoluted MS spectrum of His_6- tagged cleaved H679R AspH variant following purification. Calculated mass: 52,767.34 Da, observed mass: 52,786.92 Da.

Firstly, the activity of the H679R variant was assessed by means of real-time SPE-MS monitoring of substrate and product formation as previously described in Chapters 2 and 3. The activity of the H679R AspH variant in the absence and presence of NaCl was measured in buffer (25 mM HEPES,

pH 7.5; Figure 5-10). The activity of the H679R variant was measured concomitantly with the activity of the H679A variant (Figure 5-10a) and wt AspH (Figure 5-10b).

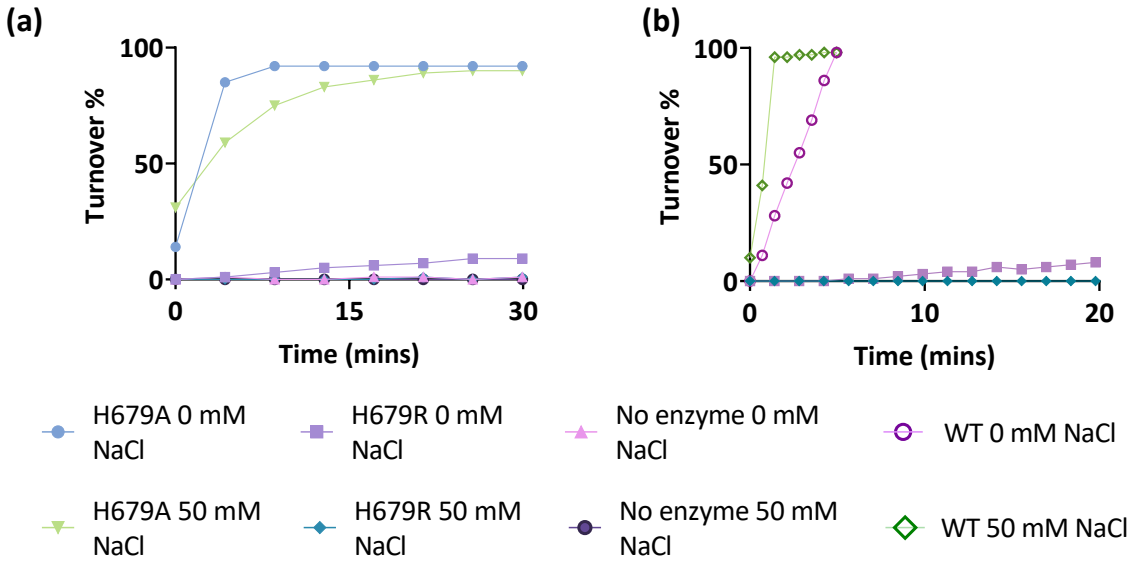


Figure 5-10. SPE-MS monitoring H679R AspH catalysis. (a) Activities of the H679R and H679A variants. Low levels of turnover by the H679R variant could be detected in 25 mM HEPES pH 7.5 buffer. After 20 mins, no detectable product was formed in 25 mM HEPES pH 7.5 mM buffer which contained 50 mM NaCl. (b) A comparison of the activity of the H679R variant with wt-AspH in 25 mM HEPES pH 7.5, with and without 50 mM NaCl. Assay conditions: 100 µM L-ascorbic acid, 200 µM 2OG, 4 µM hFX-Asp_{103}, 20 µM ferrous ammonium iron sulphate and 0.2 µM enzyme.

Approximately 20 minutes post addition of the H679R variant to the reaction assay, <10% of substrate had been converted to product in 25 mM HEPES pH 7.5 buffer. No detectable levels of product were formed when 50 mM NaCl was added to the buffer. Thus, the H679R variant displays a significant decrease in activity compared to the wt AspH, which catalysed the full conversion of substrate to product in ~5 min under the same conditions (Figure 5-10b). The level of activity displayed by the H679R variant is also lower than that of both the H679A (Figure 5-10a) and H725A variants (Chapter 2), suggesting that the H679R variation disrupts the ability of wt AspH to perform efficient catalysis.

The H679R AspH variant was co-crystallised with Mn(II), 2OG and the hFX-Asp_{103} substrate. Cross-seeding with wt AspH crystals was used to enhance the rate of crystal formation and improve diffraction data quality. Analysis of the H679R AspH:Mn:hFX-Asp_{103} complex structure confirmed

the presence of an arginine residue at position 679. Despite co-crystallisation with Mn(II) and 2OG, no metal or 2OG was observed to bind to the H679R AspH active site.

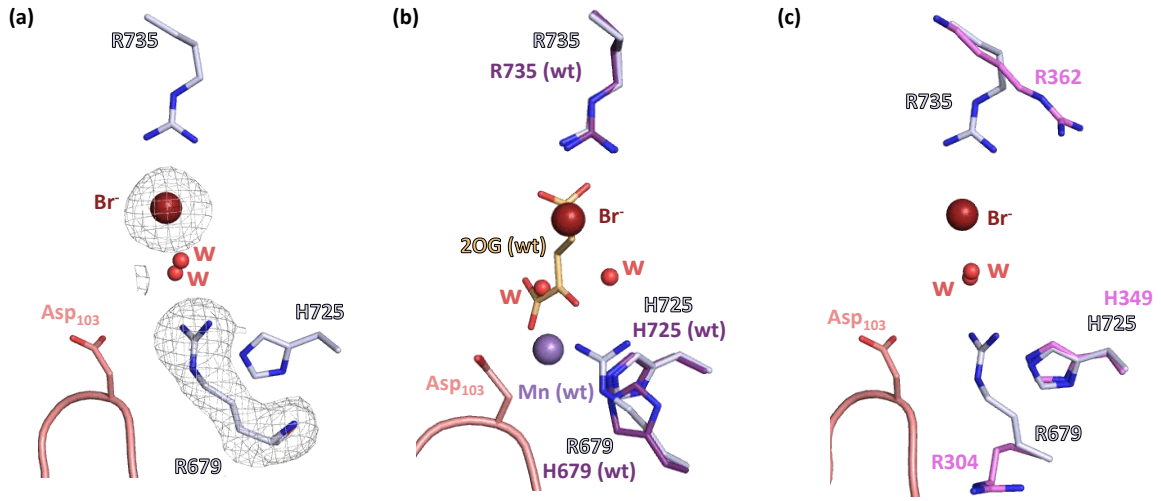


Figure 5-11. View from the active site of the H679R AspH variant crystal structure. (a) Evidence for a bromide ion originating from the precipitant can be observed at the active site. Conditions: 0.1 M bis-tris propane pH 7.5, 0.2 M NaBr, 20 % PEG 3350, wt AspH seeds. Polder omit electron density is carved to 3σ around the R679 and Br^-. (b) an overlay of the H679R active site with the wt AspH active site, co-crystallised with Mn and 2OG. Water molecules omitted for clarity. H679R: blue-white, wt-AspH: deep purple. (c) Overlay of H679R AspH active site (blue-white) with the calculated AspHD1 active site (pink).

Instead, a larger ion, most likely Br^- originating from the buffer (buffer composition: 0.2 M sodium bromide, 0.1 M Bis Tris propane pH 8.5 and 20 $\%_{w/v}$ PEG 3350), is the only unassigned group observed at the active site. By contrast with the calculated AspHD1 structure, the side chain of R679 orients into the active site space which is occupied by a metal ion in wt AspH:metal complex structures (Figure 1-10). Thus, the guanidinium group of R679 likely hinders the metal ion from binding in the AspH active site.

The crystallisation process was repeated with addition of a soaking step immediately prior to harvesting of the crystals. The crystals were soaked in solutions containing 10 mM 2OG and 10 mM Mn(II) for 5 min. In the H679R AspH structures obtained from these crystals the Mn(II) ion and 2OG density can clearly be observed at the active site (Figure 5-12).

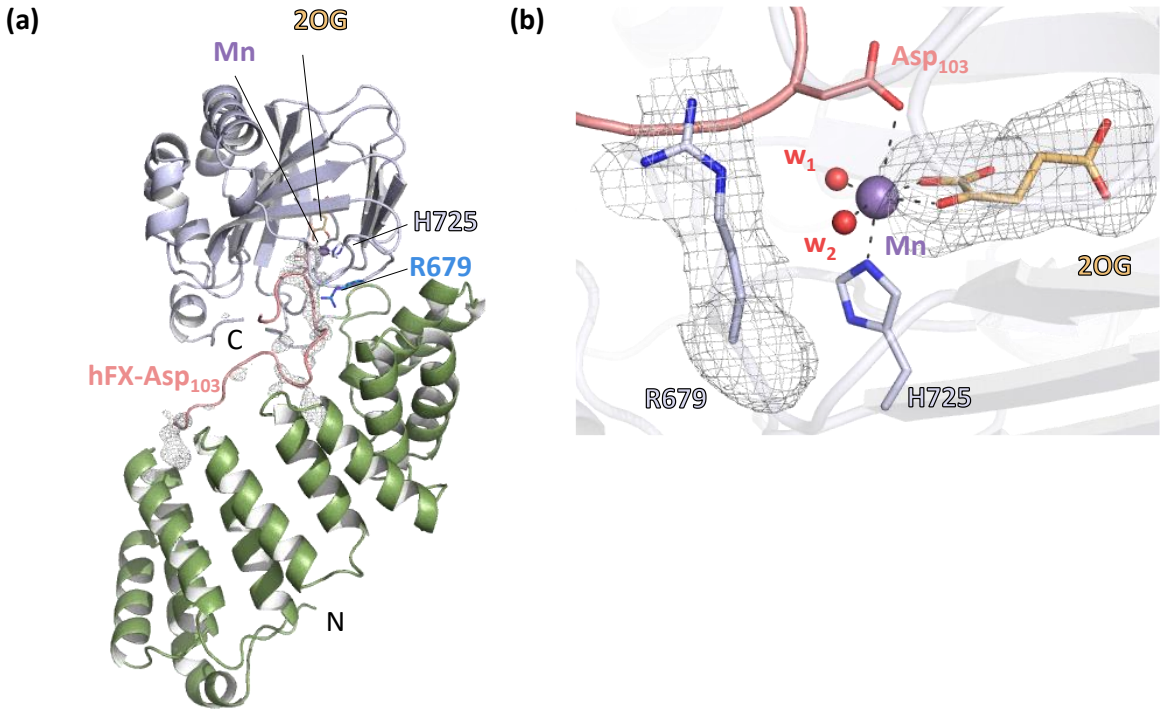


Figure 5-12. View from a crystal structure of the AspH H679R variant in complex with hFX-Asp$_{103}$ substrate, Mn(II) and 2OG. (a) Overview of the H679R variant overall fold with Polder omit electron density carved to 2σ around the substrate. The substrate is present, although there is poor electron density for the substrate outside of the active site. R679 is highlighted in cyan to show position of side chain outside of the active site. (b) Close up of active site with most likely position of R679

Interestingly, the introduction of arginine at the active site appears to have affected the substrate binding mode. There is a significant reduction in electron density around the substrate compared to the wt AspH structures, suggesting it is relatively low in occupancy. Additionally, the electron density surrounding the side chain of R679 is weak and it is impossible to resolve the exact location of the guanidinium group. The conformation modelled of R679 into the active site is based on iterative rounds of refinement that assigned this conformation to be the most likely one. It is possible that there is a clash between the longer guanidinium side chain of R679 and the substrate Asp$_{103}$ carboxylate group for space at the active site.

Additionally, there is a shift in the backbone acidic loop adjacent to the active site, consisting of residues 615-620 (RMSD for whole construct alignment: 0.625 Å; RMSD for 615-620 loop alignment: 3.657 Å) (Figure 5-13).

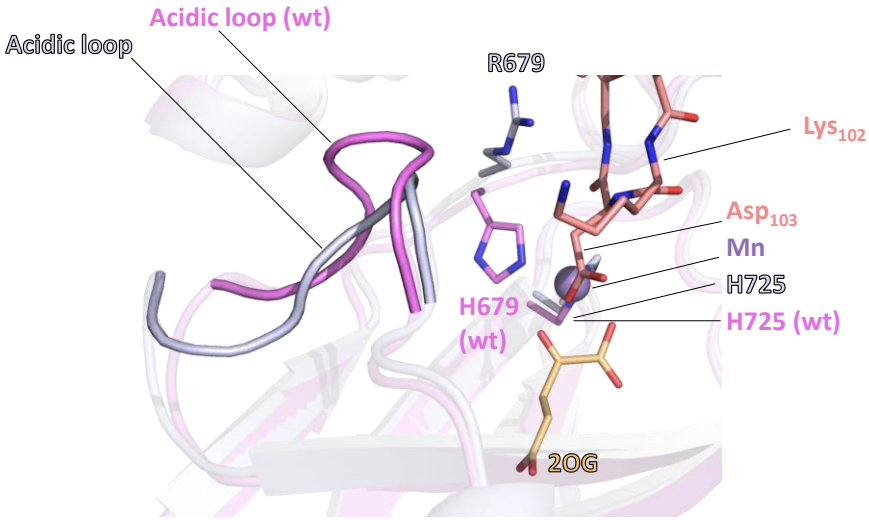


Figure 5-13. Overlay of the H679R active site with wt AspH active site. Pink = wt AspH, blue-white = H679R AspH. Cartoon representations of the acidic loop consisting of residues 615-620 located adjacent to the active site are highlighted to show the movement in the H679R structure. Wt AspH cofactors (Mn, 2OG), wt substrate and active site water molecules omitted for clarity.

Substrate binding in AspH likely follows an induced fit mechanism and the side chains of acidic residues within this loop undergo substantial conformation shifts that are likely to be essential for secure substrate coordination. Upon substrate binding in wt AspH, the side chains of E613, D616 and E617 shift to form interactions with the main chain of the substrate Asp_{103} and the side chain of the substrate Lys_{102}.[124] If the side chain of R679 is indeed projecting out of the active site as the electron density indicates, it would disrupt the conformation of the protein backbone residues in this loop and affect the ability of the H679R variant to bind substrate. A less tightly held substrate would thus reduce the clarity of the electron density around the substrate, as observed, and may also be responsible for the reduction in the catalytic efficiency of the H679R variant.

A comparison of the H679R AspH:Mn:2OG:substrate complex structure with the H679R AspH apo structure suggests that it is unfavourable for AspH to employ an arginine at residue 679 as a metal binding ligand; in the H679R structure the R679 guanidium group does not chelate the metal and the H679R variant manifests a reduction in catalytic efficiency compared to wild-type AspH. This could be due to unfavourable steric factors or destabilising electrostatic interactions. As shown by the overlay of the wt AspH active site with the H679R active site (Figure 5-13), the R679 side chain is longer than the H679 imidazole group, and without Mn present the R679 side chain extends into

the space normally occupied by the metal ion. Therefore, it may be sterically unfavourable for R679 to adopt a conformation that positions the guanidinium group at an appropriate angle to coordinate to Mn, as supported by the reduction in substrate occupancy (Figure 5-12).

Secondly, the active site of AspH is likely to be unsuitable environment for generating appropriate electrostatic interactions to allow arginine to act as a metal binding ligand. The guanidinium side chains of arginine residues are usually protonated at physiologically relevant pH values. The movement of the R679 side chain out of the active site when Mn(II) is present (Figure 5-12) may be a result of electrostatic repulsion between the charged guanidinium group and the Mn(II) ion. The capacity of other proteins to employ arginine as a metal binding ligand, e.g. BioB or the CA I variant (Section 5.1, Figure 5-3), is postulated to be due to extenuating circumstances in these systems. In the case of BioB, it was suggested that the high number of potential hydrogen-bonding partners in close proximity to a metal coordinating arginine could alter its pKa and allow it to serve as a metal ligand.[335] Similarly, the ability of the CA I variant to use arginine to coordinate Zn(II) was suggested to be due to nearby residues enhancing the nucleophilicity of the arginine ligand.[336] The active site of AspH is likely to have no such extenuating circumstances, and indeed, other nearby arginine residues, R735 and R688, are assumed to be protonated based on the interactions the side chains form to 2OG. Future investigations could research the effect of pH on the conformation of R679, however, the pH at which R679 is likely to be unprotonated (>12) would be above the pH at which the enzyme has been denatured so this may be challenging to pursue.

Residues at the AspHD1 active site are computationally predicted to adopt comparable conformations to the AspH active site residues and it is possible that AspHD1 would encounter similar challenges when forming an Fe:2OG:AspHD1:Substrate complex. Difficulty in forming an active site complex may hinder catalysis and AspHD1 may not be an active oxygenase enzyme. However, it is impossible to predict based on the currently available structural data whether the overall fold and active site of AspHD1 is more accommodating to an arginine metal ligand than

AspH. As described in other oxygenases, it is possible that secondary sphere interactions at the AspHD1 active site substantially alter the local environment around R304, thus altering the pKa and/or the orientation of the side chain and making it amenable to metal ion coordination.[334,336] Examination of the predicted active site of AspHD1 does not reveal any obvious nearby candidates that could deprotonate R304; however. there are regions of the sequence in which the structure has not been predicted and these could change the overall fold of AspHD1 to alter the environment around R304.

Secondly, even if R304 is not involved in metal coordination at the AspHD1 active site, AspHD1 may still be able to function as an effective catalyst. As discussed in Chapter 2, the AspH active site is remarkably flexible and mutations to key Fe(II)-ligating residues H679 and H725 do not render the enzyme inactive. AspH can catalyse the hydroxylation of substrate using only a single His residue as a metal binding ligand. It is possible that AspHD1 can function similarly, perhaps employing only H349 to ligate Fe(II) at the active site. R304 may have an alternative role at the active site; the conformational flexibility of the side chain may assist in substrate recognition. Binding of any AspHD1 substrates, if they do exist, could cause conformational changes in the overall fold of the enzyme that allow R304 to be positioned in a productive conformation. With the correct substrates, R304 could be involved in substrate stabilisation outside of the active site with H349 providing the sole enzyme based Fe(II) ligand during catalysis. Binding of incorrect substrates to AspHD1 may lead to steric clashes between R304 and the substrate, as observed in the AspH H679R variant structure, and therefore prevent turnover occurring. Understanding the pathways that AspHD1 is involved with, and therefore identifying potential substrates will be key to further assessing the function of the AspHD1 active site and the role, if any, of R304 as a metal binding ligand.

5.4 Conclusions and future work

Biomedical research has recently discovered the presence of *ASPHD1* and *ASPHD2* in clinically relevant systems, for example the overexpression of *ASPHD2* tumour stem cell microenvironments.[221] Characterisation of the proteins encoded by *ASPHD1* and *ASPHD2* is therefore of interest, however, little is currently known about their structure and function.

AspHD1 is predicted to be a human 2OG-dependent hydroxylase possessing an unusual RXG...H active site metal binding motif (Figure 5-2a). This work attempted to express the human protein AspHD1 in *E. coli* for use in biochemical characterisation studies. Despite varying construct length, the competent cell expression system and the fusion tag of the constructs, correctly-folded soluble protein could not be recovered from *E. coli*. Due to time constraints, re-folding of protein formed in inclusion bodies was not attempted and expression trials in other organisms were not performed.

Previous studies have reported that up to 75% of human proteins can be successfully expressed in *E. coli*, however only a quarter of these are found in an active soluble form.[350] The results from this work are further evidence for the challenges in using *E. coli* as an expression system for human proteins. *E. coli* is likely an unsuitable expression system for AspHD1, possibly due to the number of cysteine residues in the sequence which are difficult to fold correctly in the *E. coli* cytoplasm. Future attempts to produce recombinant AspHD1 should use alternative expression systems, such as baculovirus or mammalian cell lines. Although these are more expensive and it can take longer to produce the target protein, there is a greater chance of obtaining correctly folded, soluble human protein from these systems. AspHD1 and AspHD2 are likely to be of increasing interest in biomedical research and therefore it will be important to be able to characterise these proteins *in vitro*.

Computationally generated structures of AspHD1 predict the active site to look remarkably similar to that of AspH. As soluble, correctly folded AspHD1 could not be recovered from expression trials

in this work, a variant of $AspH_{315-758}$ was produced to assess the capacity of arginine as a metal binding residue in a related system. The H679R AspH variant was purified, expressed and crystallised.

A crystal structure of the apo-protein confirmed the presence of the H679R variation at the active site. Crystallisation of the H679R variant with cofactors Mn and 2OG was challenging. A relatively low occupancy of substrate was observed in the H679R AspH:Mn:2OG:hFX-Asp_{103} complex as revealed by analysis of the electron density. The lack of density of around R679 side chain makes it difficult to assess whether the residue is involved in metal binding, but suggests that it is conformationally flexible. Movement in the protein backbone adjacent to the active site suggests that the arginine side chain may be disrupting the normal conformation of this region, which, in turn hinders the ability of H679R to effectively bind substrate. These data indicate that R679 is unlikely to be productive in metal binding at the active site of AspH.

Due to time limitations on this project no further kinetic or thermodynamic characterisation of the AspH H679R variant was performed. It is possible to conclude that in AspH, R679 is unable to coordinate Fe(II) at the active site. However, the AspH H679R variant still retains low levels of catalytic activity, likely due to the fact that AspH can catalyse hydroxylation with a single metal binding ligand at the active site, as evidenced by the H679A and H725A variants (Chapter 2). There is a similar predicted geometry of residues around AspHD1 active site as observed in AspH, but the overall fold of AspHD1 could be more accommodating to an Arg residue than in AspH. Alternatively, AspHD1 may be an effective catalyst without employing R304 as an Fe(II) binding ligand, as determined to be possible in the AspH H679R model system.

In conclusion, although further work is required, the results described here further support the capacity of 2OG oxygenases, and in particular, AspH, to accommodate unusual Fe(II) binding coordination modes whilst maintaining the potential for catalysis. Thus it could be of further

investigation on potential 2OG oxygenases with atypical active sites, e.g. PHF2, JARID and Hairless, especially since some of these have been shown to be of biological importance.

5.5 Chapter acknowledgements

The plasmid encoding the H679R AspH variant was produced by Dr Inga Pfeffer.

5.6 Supplementary information

Table S5-1. Data collection and refinement statistics for the crystal structures of the H679A variant (Figure 5-11 and Figure 5-12).

	H679R:hFX-Asp$_{103}$	**H679R:2OG:Mn:hFX-Asp$_{103}$**
Wavelength (Å)	0.9763	0.9763
Resolution range (Å)	50.17 - 2.45 (2.538 - 2.45)	61.63 - 2.301 (2.383 - 2.301)
Space group	P 2_1 2_1 2_1	P 2_1 2_1 2_1
Unit cell		
a, b, c (Å)	49.6188 86.0423 123.504 90	50.499 85.9894 123.255
α, β, γ (°)	90 90	90 90 90
Total reflections	269850 (27859)	330636 (33619)
Unique reflections	20137 (1974)	24536 (2394)
Multiplicity	13.4 (14.1)	13.5 (14.0)
Completeness (%)	99.85 (99.75)	99.89 (99.58)
Mean I/sigma(I)	5.74 (0.85)	11.25 (1.53)
Wilson B-factor (Å^2)	43.17	40.03
R-merge	0.3039 (2.141)	0.1672 (1.162)
CC1/2	0.993 (0.757)	0.998 (0.847)
CC*	0.998 (0.928)	0.999 (0.958)
Reflections used in refinement	20112 (1970)	24517 (2391)
Reflections used for R-free	1998 (195)	1203 (122)
R-work	0.1920 (0.2614)	0.1911 (0.2434)
R-free	0.2519 (0.3138)	0.2493 (0.2984)
CC(work)	0.958 (0.852)	0.961 (0.885)
CC(free)	0.911 (0.784)	0.942 (0.779)
Number of non-hydrogen atoms	3647	3761
macromolecules	3516	3627
ligands	66	11
solvent	91	123
Protein residues	446	447
RMS(bonds)	0.004	0.008
RMS(angles)	0.95	0.88
Ramachandran favored (%)	97.29	97.07
Ramachandran allowed (%)	2.71	2.71
Ramachandran outliers (%)	0.00	0.23
Rotamer outliers (%)	1.13	1.04
Clashscore	6.14	5.11
Average B-factor	47.52	46.48
Macromolecules (Å^2)	47.47	46.55
Ligands (Å^2)	59.59	35.01
Solvent (Å^2)	44.03	45.40

Chapter 6 | Conclusions and future work

Conclusions and future work

The AspH Fe(II) binding motif and mechanism of hydroxylation

The work described in this thesis aimed to investigate the mechanism of hydroxylation catalysed by AspH, and determine the significance of the unusual active site Fe(II)-ligating dyad. The work described in **Chapter 2** predominantly explored Fe(II)-binding in AspH by use of mutagenesis studies that substituted the active site Fe(II)-ligating histidine residues for alanine residues, affording the H679A and H725A AspH variants. Activity assays using the H679A and H725A variants established that AspH retains substantial levels of catalytic activity while employing only a single Fe(II)-ligating residue at the active site (Figure 2-4). This unexpected finding may have implications for the functional assignment of 2OG-dependent oxygenases.

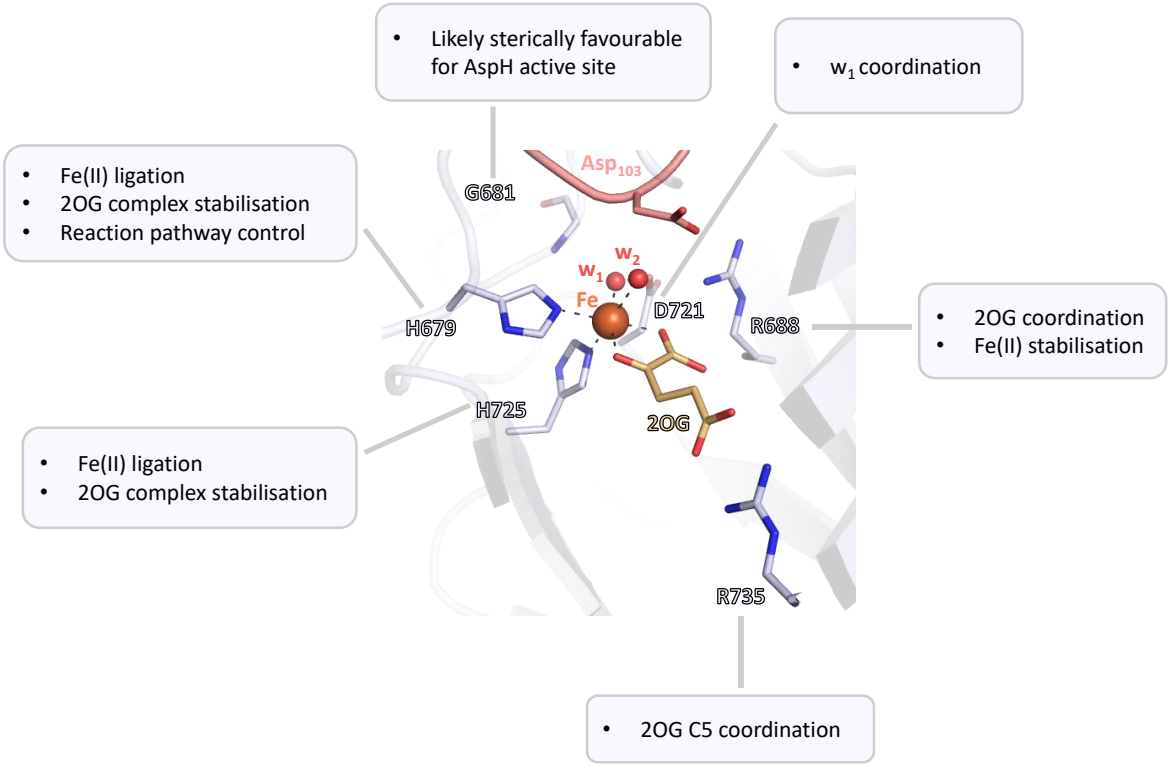


Figure 6-1. A summary of the roles of AspH active site residues based on the work described in this thesis. Chapter 2 provided evidence for the roles of H679, H725 and G681. The roles of R735 and R688 were examined in Chapter 3.

Surprisingly, evidence from kinetic studies indicated that substitution of the either H679 or H725 to A had a greater effect on the 2OG-binding capacity of AspH than on Fe(II)-binding (Table 2-4). In

particular, the H725A variant exhibited significantly impaired 2OG-binding. These data suggest a role for the H679 and H725 residues in stabilising the formation of an AspH:2OG:Fe(II) complex at the active site.

Interestingly, the addition of third active site ligand to restore the typical hydroxylase Fe(II) triad in the G681D and G681E variants appeared to impair AspH catalysed hydroxylation activity more than removal of an Fe(II)-ligating histidine (Figure 2-17). No detailed characterisation of the G681D and G681E variants was performed; however, this reduction in activity could possibly be attributed to the introduction of unfavourable steric clashes around the active site. A summary of the activity of different AspH Fe(II) motifs is shown in Table 6-1.

Table 6-1. A summary of the Fe(II)-binding motifs in AspH variants explored in this thesis and the effect on AspH activity. AspH activity variant stated as a percentage of the initial rate of hydroxylation of wt AspH under optimised assay conditions (See Chapter 2, Figure 18 for complete timecourse).

Number of Fe(II)-binding residues	Motif	AspH activity
0	$A_{679}XG_{681}...A_{725}$	Inactive
1	$H_{679}XG_{681}...A_{725}$	~10% of wt activity
	$A_{679}XG_{681}...H_{725}$	~30% of wt activity
2 (wt)	$H_{679}XG_{681}...H_{725}$	Optimal hydroxylation activity
3	$H_{679}XD_{681}...H_{725}$	~10% wt activity
	$H_{679}XE_{681}...H_{725}$	~3% wt activity

Use of oxygen isotope labelling revealed the significance of the AspH histidine dyad in directing the reaction pathway (Figure 2-15). wt AspH likely follows the consensus mechanism for hydroxylation, with incorporation of oxygen into the product originating from O_2 only (Figure 6-2, grey pathway). By contrast, the substitution of H679 to A enabled the reaction to proceed via an alternative reaction pathway which resulted in approximately 50% of oxygen in the product derived from O_2 and 50% from solvent (Figure 6-2, blue alternative pathway). The exact nature of the final radical

recombination step could not be determined from current data; however, future work could be performed to characterise intermediates in these pathways.

The discovery that the H679A variant can catalyse hydroxylation of a hydroxy group originating from the solvent raises the possibility of engineering an AspH variant that can catalyse the substitution of other functional groups onto substrates. The functionalisation of inactivated carbon sp^3 centres is of interest from a biocatalytic perspective. Future work, based on prior investigations that successfully engineered a hydroxylase to perform halogenation, could involve designing an AspH variant that can catalyse the substitution of functional groups other than OH onto the Asp and Asn substrates.

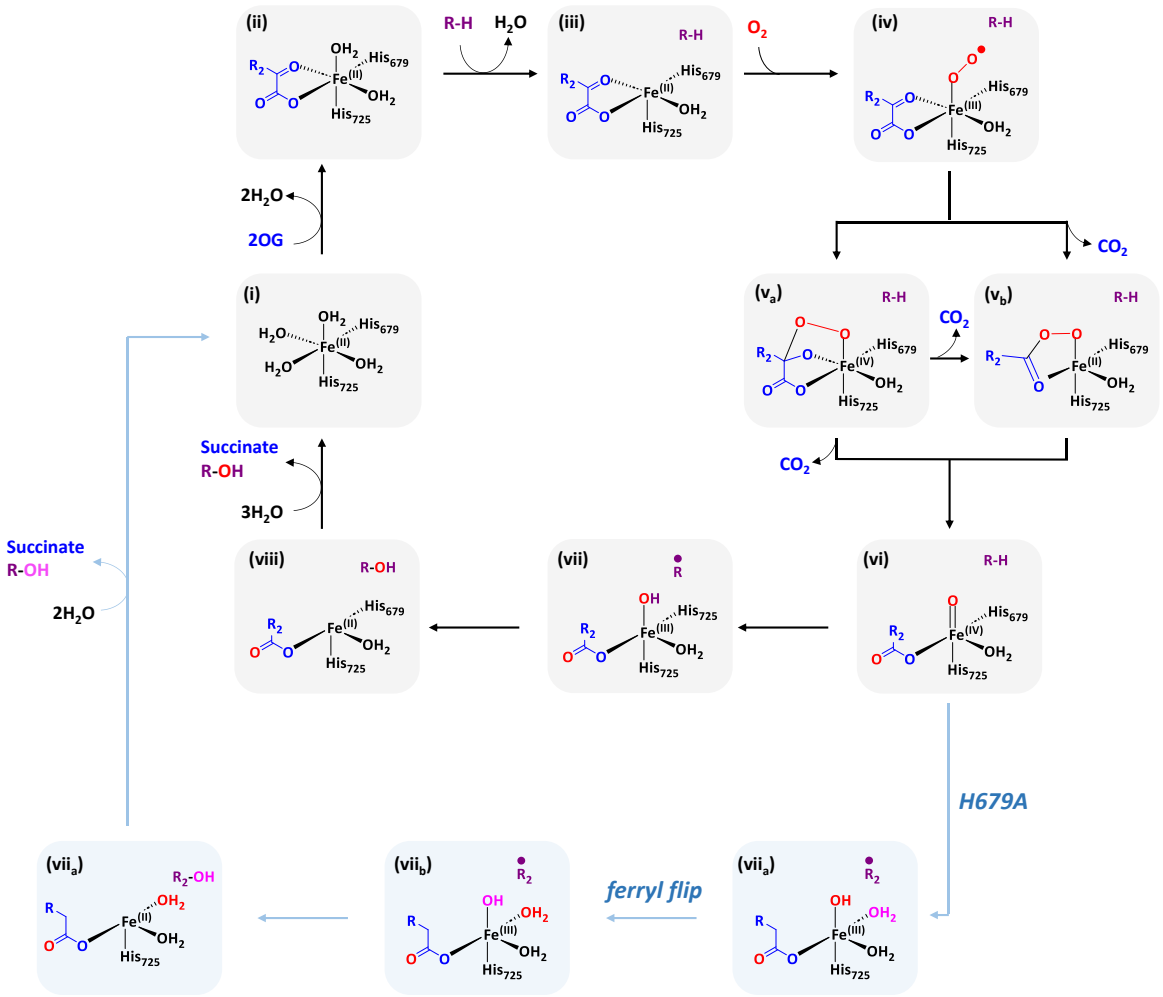


Figure 6-2. The catalytic mechanism for AspH catalysed hydroxylation. The final steps of the H679A variant mechanism is simplified as the exact nature of the proposed recombination step is unknown, but likely to involve rearrangement of intermediates (vi) or (vii$_a$). See Chapter 2, Section 2.3.3 for more discussion.

The work described in **Chapter 4** established the foundations for future work to investigate the mechanism of AspH catalysis by room temperature time-resolved serial femtosecond crystallography (tr-SFX) at an X-ray free electron laser (XFEL) beamline. AspH was found to retain catalytic activity *in crystallo,* making it a suitable target for the development of tr-SFX experiments. A robust, scalable method for the production of AspH microcrystals for serial crystallography was developed. Subsequently, a structure of AspH in complex with Fe, 2OG and hFX-Asp_{103} was obtained at an XFEL and refined to a resolution of 2.05 Å (Figure 4-15).

Tr-SFX experiments could be particularly useful in probing the order of substrate and cofactor binding at the AspH active site. The carboxylate side chain of the hfX-Asp_{103} Asp_{103} is observed in two conformations in AspH crystal structures. It is hypothesised that one of these, in which the Asp_{103} carboxylate coordinates to the metal centre, is unproductive for catalysis, but may have a role in stabilising metal binding at the active site. The substrate Asp_{103} could effectively complete the catalytic triad of metal ligating residues prior to catalysis occurring.

Future tr-SFX experiments could involve performing drop-on-tape experiments which expose droplets of anaerobically prepared AspH:2OG:Fe(II):substrate microcrystals to oxygen. Depending on the time scale of the binding processes, it maybe be possible to resolve whether a mixture of Asp_{103} conformations are observed during Fe(II) binding, and then whether Asp_{103} subsequently adopts a single, catalytically-productive conformation upon exposure to oxygen.

AspH as a therapeutic target for FDLAB syndrome

FDLAB syndrome is a rare genetic disorder which has been linked to mutations in the ASPH gene.[210] Several reported pathogenic mutations in the ASPH gene, including R688Q, R735Q and R735W, are located in the oxygenase domain and were thus predicted to affect the functioning of AspH.[210] The work described in **Chapter 3** unambiguously confirmed the correlation between the FDLAB-linked pathogenic mutations R688Q, R735Q and R735W and impaired hydroxylation AspH activity. The

R735W AspH variant was found to be catalytically inactive, while the R688Q and R735Q variants retained some catalytic activity but substantially less than that exhibited by wt AspH (Figure 3-3).

Treatments that aim to restore normal levels of AspH function could be therefore be pursued in future. One such treatment may be the use of 2OG-derivatives which exploit alternative interactions at the AspH active site. The use of 2OG-derivatives to restore AspH variant activity is further proof that the concept of chemical co-rescue is a promising avenue for treatment of genetic diseases that result in impaired enzyme activity (Table 3-5).[257] Future investigations could test the efficacy of the 2OG-derivatives *in vivo*. Treatments may also require the optimisation of the small molecule delivery system and would need to ensure the 2OG-derivatives are specific to AspH and do not affect the functioning of other human 2OG-dependent oxygenases.

The 2OG-oxygenase catalytic triad

Evidence that AspH is catalytically active with a single Fe(II)-binding ligand raises questions around the assignment of 2OG-oxygenases based on active site motifs. Other 2OG-hydroxylases with unusual Fe(II) binding motifs have been reported, e.g. PHF2 or Hairless, but their catalytic activity is under debate and further work is needed in this area.

The work described in **Chapter 5** aimed to expand the understanding of the consensus 2OG-oxygenase binding triad and evaluate the potential of arginine as a Fe(II)-binding residue. The worked aimed to characterise AspHD1, a predicted oxygenase of which sequence alignments suggest an RXG…H active site metal-binding dyad. Although AspHD1 was not characterised due to difficulties with recombinant protein expression in *E. coli*, a model system gave some insight into arginine as a metal-binding residue. In the AspH model system, that is the H679R variant, arginine was not a productive Fe(II)-ligating ligand. Previous mutagenesis studies on the 2OG-hydroxylase prolyl-4-hydroxylase also established that substitution of the Fe(II)-binding histidine residues to positively charged residues, e.g. arginine, inactivated the enzyme.[351] However, other arginine metal binding interactions are highly dependent on the surrounding environment and this could

also be the case in AspHD1.[333] Further work on characterising AspHD1 and AspHD2 should be performed in future to fully investigate the unusual predicted oxygenase Fe(II)-binding motifs.

It may be argued that AspH (and some, but not all, halogenases, e.g. SyrB2) are special cases in that their binding of their oligomeric peptide substrates may promote Fe(II) binding. Indeed, this may well be the case for AspH where the substrate Asp/Asn residue can complex the Fe(II), likely in a catalytically unproductive manner. However, some halogenases, e.g. WelO5, have a non-oligomeric small molecule substrate.[352]

Artificial engineering of surface residues on cytochrome cb_{562} (cyt cb_{562}) has provided evidence for self-assembly of proteins driven by metal templating. Histidine residues were substituted onto the surface of cyt cb_{562} and subsequent self-assembly of protein oligomers was driven by coordination to Zn(II) ions by the histidine residues.[353–356] It is possible that coordination to metal ions can could be the driving force for this phenomenon to occur naturally in proteins and their peptide substrates.

The operation of a "substrate template" active site, however, also raises the possibility that metal-chelating active sites might be formed by two proteins. This possibility will not be easily identified by sequence studies on single proteins. If correct, it would mean the number of metal-enzyme active sites could be much larger than previously predicted. This proposal may also help inspire new types of non-enzyme metal-catalysts. Note some such catalysts are biomedically integral, e.g. bleomycin, which is used in cancer treatment.

Chapter 7 | Materials and methods

7.1 General procedures

Frequently used chemicals were purchased from Sigma Aldrich or Thermo Fisher Scientific, unless stated otherwise. Synthetic peptides were purchased from GL Biochem (Shanghai) Ltd. Ultra-pure MilliQ water, purified by a Milli-Q Reference Water Purification System, was used in assays, buffers and media. Autoclaved MilliQ water, referred to as sterile MilliQ, was used in making sterile solutions of antibiotics, IPTG and for all DNA cloning protocols. Antibiotics were purchased from Apollo Scientific.

7.2 Molecular biology methods

7.2.1 General methods

Aseptic procedures were followed when handling *E. coli* cultures. All procedures were carried out in a Laminar Flow Hood using autoclaved pipette tips or sterile filter tips and autoclaved microfuge tubes. Growth media, pipette tips and tubes were autoclaved for 20 minutes at 121°C before use. Components for making 2YT media and agar plate (2xYT, Tryptone, Yeast Extract and Agar) were purchased from Sigma Aldrich. Terrific Broth Autoinduction media with trace elements was purchased from Formedium. Kanamycin stocks were prepared to a final concentration of 30 mg ml^{-1} in sterile water and chloramphenicol stocks were prepared to a final concentration of 40 mg ml^{-1} in 70% ethanol solutions. *E. coli* competent cells New England Biolabs (NEB) 5-alpha (C2987H), BL21 (DE3) (C2527H), LEMO21 (DE3) (C2528J) and SHuffle (C3026J) were purchased from New England Biolabs. XL10 Gold cells (AGLS200314) were purchased from Agilent. BL21 (DE3) pLysS cells (Novagen) were purchased from Merck.

7.2.2 DNA concentrations and sequencing

DNA concentrations were measured using a NanoDrop One Microvolume UV-Vis spectrophotometer. DNA sequencing was performed by Eurofins Genomics. Standard T7 forward and T7 reverse primers were used to sequence inserts in the pET28a(+) vector.

7.2.3 Plasmids and genes

The pET28a(+)-His_6-$AspH_{315-758}$ plasmid containing the human gene encoding $AspH_{315-758}$ was a gift from Dr Inga Pfeffer. Site-directed mutagenesis to produce the AspH variants H679A, H725A, R735W, H679R, G681E and G681D was performed by Dr Inga Pfeffer. Site directed mutagenesis to generate the AspH G434V variant was performed by Ms Cynthia Hou. The full length $AspHD1_{1-390}$ gene was codon-optimised for expression in *E. coli* and synthesised by Life Technologies. Vectors containing genes encoding MBP and SUMO fusion tags were a gift from Dr Klemensas Simenlis. The sequence encoding for a 3C protease recognition site was amplified from a pCold I vector which was a gift from Dr Patrick Rabe.

7.2.4 Primers

Primers for use in Gibson Assembly were designed using the NEBuilder online tool (https://nebuilder.neb.com/#!/).[339] Primers were for use in site directed mutagenesis, insertions, deletions or substitutions of DNA were designed using the NEBase Changer online tool (https://nebasechanger.neb.com/). Custom made oligos were supplied by Life Technologies. Sterile MilliQ water was used to dissolve desalted primers to a stock concentration of 100 mM.

Table 7-1. Primers used for substitutions, deletions or insertions of DNA.

Assembled construct	Amplified DNA	Primers
$AspHD1_{206-390}$	$AspHD1_{96-390}$ plasmid	5'-CGTCATGATGTTGAACTGCTGG-3' 5'-CGGGCCCTGAAACAGCAC-3'
3C-pET28a(+)-His_6-$AspH_{315-758}$	PET28a(+)-His_6-$AspH_{315-758}$	5'-TTTCAGGGCCCGCATATGGCTAGCATGACTGG-3' 5'-CAGCACTTCCAGGCCGCTGCTGTGATGATG-3'

Table 7-2. Primers used for site-directed mutagenesis of AspH. *Designed by and mutagenesis performed by Dr Inga Pfeffer. **Designed by and mutagenesis performed by Ms Cynthia Hou.

Mutation	Primers
H679A*	5’ – CTACGTGTGGCCGGCCACAGGGCCC-3’ 5’ – GGGCCCTGTGGCCGGCCACACGTGAG-3’
H725A*	5’ – GATGACTCCTTTGAGGCCGAGGTATGGCAGG-3’ 5’ – CCTGCCATACCTCGGCCTCAAAGGAGTCATC-3’
H679R*	5'-CTCACGTTGGCCGAGGACAGGGCCC-3' 5'-GGGCCCTGTCCTCGGCCACACGTGAG-3'
R688Q	5’-CTCCAGATGCACCTGGGCTTG-3’ 5’-GTGCATCTGGAGCCTGCAGT-3’
R735Q	5’-CTTTCCAGCTGATATTCATCGTGG-3’ 5’-GAATATCAGCTGGAAAGATGAGG-3’
R735W*	5'-GGATGCCTCATTITTGGCTGATATTCATCGTG-3' 5'-CACGATGAATATAGCCAGAAAGATGAGGCATCC-3'
G434V**	5’-GTTCATATGAGAGGTTCCCTGCTTACCCTG-3’ 5’-CTCATATGAACTAGAAATTGTTGCCTGTCTGAGCGACGCTTC-3’
G681D*	5'-CTGCAGTITGTGGGATCTGTGTGCGGCCAC-3' 5'-GTGGCCGCACACAGAACCCACAAACTCAG-3'
G681E*	5'-CTGCAGTITGTGGGTTCTGTGTGCGGCCAC-3' 5'-CGCAGCCTTCCAGCATGATAGTAAGCGGC-3'

Table 7-3. Primers used for Gibson Assembly of plasmids.

Assembled construct	Amplified DNA	Primers
pET28a(+) AspHD1$_{1-390}$	AspHD1 synthetic gene	5'-AAGTGCTGTTTCAGGGCCCGATGAAAGAAGGTCGCGGTAG-3' 5'-TGCTCGAGTGCGGCCGCTTATTACGGATCCGGTGCAAAAAC-3'
	pET28a(+) plasmid	5'-TAAGCGGCCGCACTCGAG-3' 5'-CGGGCCCTGAAACAGCAC-3'
pET28a(+) AspHD1$_{96-390}$	AspHD1 commercial gene	5'-GCAAATGGGTTATCGTCTGGGTAGTCAGG-3' 5'-CCGCTTACTATTACGGATCCGGTGCAAAAAC-3'
	pET28a(+) plasmid	5'-GGATCCGTAATAGTAAGCGGCCGCACTC-3' 5'-CCAGACGATAACCCATTTGCTGTCCACC-3'
MBP-AspHD1$_{206-390}$	pET28a(+) AspHD1$_{206-390}$	5'-ACCGGATCCGTAACGGATCCGAATTCGAG-3' 5'-CATCATGACGTCCACTTCCAATATTGGATTG-3'
	MBP gene	5'-TGGAAGTGGACGTCATGATGTTGAACTGC-3' 5'-GGATCCGTTACGGATCCGGTGCAAAAAC-3'
SUMO-AspHD1$_{206-390}$	pET28a(+) AspHD1$_{206-390}$	5'-TACATCATCATCATCATCACATGTCGGACTCAGAAGTC-3' 5'-CCCTGAAACAGCACTTCCAGTCCACCAATCTGTTCTCTG-3'
	SUMO gene	5'-CTGGAAGTGCTGTTTCAG-3' 5'-GTGATGATGATGATGATGTAC-3'

7.2.5 PCR

PCR was carried out using the NEB Q5® High-Fidelity DNA Polymerase kit and set up according to manufacturer's recommendations (Table 7-4). Annealing temperatures were calculated based on primer sequences using the NEB T_m calculator. Thermocycling was performed using a BioRad Thermocycler and a routine PCR protocol is outlined in

Table 7-5. The extension time and annealing temperatures were adjusted for each reaction depending on the primers and length of amplicon.

Table 7-4. Composition of PCR reactions.

Component	50 µl reaction	Final concentration
5X Q5 Reaction Buffer	10 µL	1X
10 mM dNTPs	1 µL	200 µM
10 µM Forward Primer	2.5 µL	0.5 µM
10 µM Reverse Primer	2.5 µL	0.5 µM
Template DNA	variable	< 1,000 ng
Q5 High-Fidelity DNA Polymerase	0.5 µL	0.02 U/µl
Sterile Water	to 50 µL	n/a

Table 7-5. Thermocycling conditions used in a typical PCR experiment.

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds
Denaturation	98°C	5–10 seconds
Annealing	*50–72°C	10–30 seconds
Extension	72°C	20–30 seconds/kb
25-35 Cycles		
Final Extension	72°C	2 minutes
Hold	4°C	

DpnI Digestion

Digestion of template DNA in PCR reactions was achieved by the addition of 1 µL of DpnI (NEB) to 25 µL of PCR product and incubation of the mixture at 37°C for 60 minutes.

7.2.6 Agarose gel electrophoresis

A 50x TAE buffer stock was purchased from Life Technologies. 1x stocks were prepared from this stock prior to use (final concentrations 40 mM Tris, 20 mM acetic acid and 1 mM EDTA, pH 8.6).

Agarose gel preparation
Agarose gel electrophoresis was used to analyse PCR reactions and separate PCR products for gel extraction and purification. 1% (w/v) Agarose gels were made by adding agarose powder to a solution of 1x TAE buffer and the suspension was microwaved until all agarose powder had dissolved. SyberSafe DNA dye (10,000 stock used to generate 1x final concentration) was added to the liquid agarose. The agarose was then allowed to set fully in a gel casting tray.

Sample preparation and gel electrophoresis
Samples were mixed with 6x DNA loading dye (Life Technologies) prior to loading on the gel. 5-20 μL of sample was added to each well alongside 5 μL ul of a DNA ladder (Life Technologies). Gels were run at 100 V for 50 minutes in 1x TAE buffer.

Imaging
Agarose gels were imaged under UV light using a Bio-Rad ChemiDoc MP Universal Hood III and images were analysed using ImageLab 5.2.1 software.

DNA gel purification
Bands were visualised using the UV imager, extracted and then purified using GeneJet Gel Extraction Kit (ThermoFisher) according to the manufacturer's protocol. Bands were excised from the agarose gel and dissolved in binding buffer (100 μL of buffer per 100 mg of gel). The solubilized gel solution was added to a spin purification column. The column was washed with 700 μL of wash buffer prior to DNA elution by addition of 50 μL of elution buffer to the column.

7.2.7 Minipreps

Minipreps were carried out according to the manufacturer's recommended protocol using a GeneJet Plasmid Miniprep kit (Thermo Fisher). A single colony from an agar plate was grown in 2xYT media containing the appropriate selection antibiotic at 37°C overnight. Cells from 4.5 mL of culture were pelleted by centrifugation (5 minutes, 14,462 x *g*, 4°C). The pellet was resuspended in 250 μL of resuspension buffer before addition of lysis buffer (250 μL) and neutralisation buffer (350 μL). Samples were centrifuged (5 mins, 14,462 x *g*, and the supernatant was added to spin

purification column. The column was washed with three rounds of 500 µL wash buffer. DNA was eluted from the column with 50 µL of elution buffer. Plasmids were stored at -20°C.

7.2.8 Gibson assembly protocol

DNA cloning was performed using the Gibson Assembly method.[339] Vectors and inserts used to generate fragments are outlined in Table 7-6. Fragments were amplified by PCR to generate 20 base-pair overlaps. The template DNA was digested by Dpn1 and then PCR products were purified by gel extraction. NEB HiFi DNA Assembly kit (E2621S) was then used according to the manufacturer's protocol to ligate DNA fragments. Components were mixed as outlined in Table 7-7 and incubated at 50°C for 30 minutes. Products were then transformed into DH5α competent cells.

Table 7-6. Constructs produced by Gibson Assembly. The origin of the vector backbone and insert for each construct is stated. The modified pET28a(+) vector contains a 3C protease recognition site in place of the thrombin protease recognition site between the His_6-tag and protein sequence.

Construct	Vector backbone	Insert
pET28a(+) $AspHD1_{1-390}$	Modified pET28a(+) from $AspH_{315-758}$ vector	Synthetic *E. coli* codon optimised full length AspHD1 gene
pET28a(+) $AspHD1_{96-390}$	Modified pET28a(+) from $AspH_{315-758}$ vector	Synthetic *E. coli* codon optimised full length AspHD1 gene
MBP-$AspHD1_{206-390}$	pET28a(+) $AspHD1_{206-390}$	MBP gene
SUMO-$AspHD1_{206-390}$	pET28a(+) $AspHD1_{206-390}$	SUMO gene

Table 7-7. Components of the NEBuilder HiFi DNA Assembly kit used in Gibson Assembly of plasmids.

Component	Volume
DNA Molar Ratio	vector: insert = 1:2
Total Amount of Fragments	3 µL
NEBuilder HiFi DNA Assembly Master Mix	10 µL
Deionized H_2O	7 µL

7.2.9 Insertions, deletions and substitutions of DNA sequences

The NEB Kinase, Ligase and DpnI enzyme kit (KLD) (M0554S) was used to insert, delete or substitute sections of DNA. Constructs produced by this method are outlined in Table 7-8.

Table 7-8. Constructs produced by insertion, substitution or deletions.

Construct	Template vector	Modification
$AspHD1_{206-390}$	pET28a(+) $AspHD1_{96-390}$	Deletion of bases encoding $AspHD1_{96-205}$
pET28a(+) His_6-3C-$AspH_{315-758}$	pET28a(+) His_6-thrombin-$AspH_{315-758}$	Substitution of section of plasmid backbone encoding thrombin protease recognition site for a 3C protease recognition site

The target section of DNA was amplified by PCR. Template DNA was then digested using Dpn1 and the PCR products were purified by agarose gel extraction. The manufacturer's protocol for the NEB KLD kit was followed to generate the product and the composition of the reactions is shown in Table 7-9. The reaction mixtures were incubated at 37°C for 30 minutes. DH5 alpha competent cells were transformed by the plasmids.

Table 7-9. Composition of KLD Enzyme Kit reactions for insertions, substitutions or deletions of DNA.

Component	Volume
PCR Product	1 µL
KLD Reaction Buffer (2X)	5 µL
KLD Enzyme Mix (10X)	1 µL
Nuclease-free Water	3 µL
Total Volume	10 µL

7.3 Protein expression and purification

7.3.1 Agar plates

Agar plates were prepared according to Table 7-10. The agar was autoclaved at 121°C for 20 minutes and allowed to cool before addition of the appropriate selection antibiotic. Agar was poured into petri dishes and allowed to set in a sterile laminar flow hood. Plates were stored at 4°C until required.

Table 7-10. Composition of agar used to make agar plates.

Component	Mass in 300 mL of MilliQ (g)
Tryptone	3
Yeast Extract	1.6
NaCl	3
Agar	4.5

7.3.2 Media

The majority of *E. coli* growths were carried out using 2xYT media. 2xYT media was freshly prepared for using 2xYT media pellets (Melford) dissolved in MilliQ water to give final concentrations 16 g L^{-1} tryptone, 10 g L^{-1} yeast extract and 5 g L^{-1} NaCl. Terrific Broth (TB) autoinduction media was used for the growths of cell pellet for AspH G681E and AspH G681D variants due to issues with bacteriophage contamination in the fermentation room. Autoinduction TB media was prepared by dissolving 55g of TB powder per litre of MilliQ, with the addition of 5 mL of glycerol per litre of media. All growth media was sterilised at 121°C for 20 minutes in an autoclave.

7.3.3 Transformations

Transformations were carried out according to manufacturer's protocol for each competent cell strain. 1-3 μL of plasmid was incubated with competent cells on ice for 20-30 minutes. Cells were heat shocked at 42°C for the duration recommended by the manufacturer. The mixture was

returned to ice, SOC outgrowth media (NEB) was added and the cells were incubated at the recommended temperature with shaking (180 rpm) for 1 hour. The sample was spread on agar plates containing the appropriate selection antibiotic and incubated overnight at the recommended temperature.

Table 7-11. Competent cell strains used.

Strain	Manufacturer	Antibiotic	Use	Heat shock (s)	Outgrowth (°C)
DH5a	NEB	n/a	Plasmid isolation	30	37
XL10 Gold	Agilent	n/a	Plasmid isolation	30	37
BL21 (DE3)	NEB	n/a	Protein expression	10	37
LEMO	NEB	Chloramphenicol	Protein expression	30	37
SHuffle	NEB	N/a	Protein expression	30	30
BL21 pLysS	NEB	Chloramphenicol	Protein expression	30	37

7.3.4 Starter cultures

Overnight starter cultures were used to inoculate fresh media for protein expression. 100 mL of 2YT media was supplemented with the appropriate selection antibiotics. A single colony from a transformation was picked using an inoculating loop, and added to the media. The media was incubated at 37°C overnight with shaking (180 rpm).

7.3.5 Expression trials

Expression trials were performed in 96 deep well plates to screen multiple cell strains and IPTG concentrations in small volumes. An overnight starter culture was used to inoculate 100 mL of fresh 2YT media, pretreated with selection antibiotic. When the desired OD_{600} was reached, 1.5 mL of culture was transferred to a 96-deep well plate. Cultures were treated with the desired

concentration of IPTG and incubated at the temperature planned for each expression trial. The plates were sealed with a gas permeable membrane. 1 mL of each culture was pelleted (10 mins, 14,462 x *g*, 4°C). The pellet was resuspended in 50 µL of BugBuster protein extraction reagent (Merck) containing 1 µL per 1 mL BugBuster Lysonase bioprocessing reagent (Merck), and pellets were allowed to solubilize at room temperature for 20 minutes. The samples were then centrifuged (15 mins, 14,462 x *g*, 4°C) and the soluble fraction was removed. The insoluble pellet was resuspended in the same volume of 8 M urea. The soluble and insoluble fractions were analysed by SDS-PAGE gels.

7.3.6 Large scale protein expression

Flasks
Baffled 2 L Erlenmeyer flasks were used for the majority of cell growths as these resulted in higher cell yields than unbaffled flasks.

2xYT media
An overnight starter culture was set up as previously described. 6 mL of starter culture was added to 600 mL of 2xYT media, supplemented with the appropriate selection antibiotics. For growths in 2xYT, cultures were grown at 37°C with shaking (180 rpm) until the $O.D_{600}$ reached 1.0-1.1. The incubator temperature was cooled to 18°C and then expression of the target gene was induced by addition of IPTG (final concentration 0.1 mM for all AspH and AspH variant growths). Cultures were incubated at 18°C with shaking (180 rpm) for 16-18 hours. Cells were pelleted by centrifugation (8 minutes, 11867 x g, 4°C) and cell pellets were stored at -80°C until protein purification.

Autoinduction terrific broth media
An overnight starter culture was prepared in 2xYT media as described previously. 6 mL of starter culture was added to 600 mL of autoinduction terrific broth (AIM TB Media, Formedium) which had been supplemented with the appropriate selection antibiotics. Cultures were grown at 37°C with shaking (180 rpm) for 4 hours, the incubator temperature was then reduced for 25°C for 4

hours, and then reduced again to 18°C for 16-18 hours. Cells were pelleted by centrifugation (8 minutes, 11867 x g, 4°C) and cell pellets were stored at - 80°C until protein purification.

7.3.7 His_6-tagged protein purification

Purifications were carried out on AKTA FPLC systems controlled by Unicorn software. All buffers were filtered through 0.2 µm filters and degassed prior to use.

Purification buffers
Buffers used in His_6-tagged protein purifications are outlined in Table 7-12.

Table 7-12. Buffers used in His_6-tagged protein purification.

Buffer	Composition
Buffer A	50 mM HEPES, 500 mM NaCl, 5 mM imidazole, pH 7.5
Buffer B	50 mM HEPES, 500 mM NaCl, 500 mM imidazole, pH 7.5
Size exclusion buffer	50 mM HEPES, 150 mM NaCl, pH 7.5

Purification of AspH and AspH variants (removal of His_6 tag)
Cell pellets were resuspended in Buffer A (100 mL per 25 g of wet cell pellet). Approximately 1 µg mL^{-1} bovine pancreas DNAse I grade II (Merck) and protease inhibitor tablets (1 tablet per 50 g of cell pellet, Roche) were added to the lysate. The cell pellet was stirred until a homogenous mixture was obtained. The lysate was passed through high pressure cell disrupter (Constant Systems) at 25 kPsi at 4°C. The lysate was centrifuged (30 minutes, 58387 x *g*, 4°C), the supernatant was removed and filtered through 0.45 µm filters (Fisherbrand), then loaded onto a 5 mL HisTrap (Cytiva), which was pre-equilibrated with 5 column volumes (CV) of Buffer A. The column was washed with 10 CV buffer A, 10 CV of a mixture of 5%$_{v/v}$ Buffer B and 95%$_{v/v}$ Buffer A, then the protein was eluted by means of a linear gradient from 5%$_{v/v}$ Buffer B to 100%$_{v/v}$ Buffer B over 20 CVs. Location and purity of the desired protein was checked for using SDS-PAGE analysis. Fractions containing the target protein were pooled and restriction grade thrombin protease (Merck) was added. The mixture was dialysed against Buffer A at 4°C overnight using SnakeSkin Dialysis Tubing (Thermo Fisher) with a

3.5 kDa molecular weight cut off. The protein was then loaded onto a HisTrap pre-equilibrated with buffer A and the cleaved protein was collected in the flow through. The protein was concentrated to a volume of 2 mL using 15 mL 30 kDa Amicon spin concentrator columns, then loaded onto a size exclusion column (Superdex 75 (S75) resin, 300 mL) using a 2 mL loop. The size exclusion column was pre-equilibrated in size exclusion buffer. Size exclusion chromatography produced pure, cleaved protein which was concentrated using spin concentrator columns, aliquoted, and flash frozen in liquid N_2 before storage at -80°C.

His_6-AspH purification (His_6 tag included)

Initial protein purifications did not include removal of the His_6 tag. The initial treatment of cell pellet and elution from HisTrap column was carried out as described before. After the HisTrap elution, fractions containing the target protein were concentrated to 2 mL and loaded onto a size exclusion column (S75, 300 mL). Protein was eluted in size exclusion buffer, aliquoted and then flash frozen in liquid N_2 for storage at -80°C.

$AspH_{315-758}$ purification for time-resolved crystallography

Large scale purification of $AspH_{315-758}$ for the purpose of time-resolved serial crystallography experiments was carried out as outlined in Chapter 4.

7.3.8 MBP-tagged protein purification

Affinity purification of MBP-tagged proteins was carried out as described in Chapter 5. A 5 mL MBP-Trap (Cytivia) was used for affinity purification and the protocol was followed according to the manufacturer's recommendations.

7.3.9 SDS PAGE analysis

NuPAGE 4-12% Bis-Tris pre-cast protein gels (Thermo Fisher) were used for SDS-PAGE analysis of protein samples.

Sample preparation
Loading dye was prepared from 4x NuPAGE LDS Sample Buffer (Merck) and β-mercaptoethanol (2.5%$_{v/v}$ final concentration). 10 µL of protein sample was added to 10 µL of loading dye and samples were then heated to 90°C for 5 minutes.

Sample loading and gel running
6 µL of sample was added per well alongside one well containing PAGERule prestained protein ladder (Thermo Fisher). Gels were run for 45 minutes at 180V in 1x MOPS running buffer (20x stock obtained from Thermo Fisher).

Gel analysis
SDS-Page gels were stained using Instant Blue solution (Abcam) on a rocker at room temperature for 15 minutes. Background staining was removed by substitution of the stain for MilliQ water. Gels were imaged using a BIO-RAD ChemiDoc MP Universal Hood III and analysed in ImageLab 5.2.1.

7.3.10 Western Blots

Materials
The presence of a His_6-tag was assessed by use of a primary anti-His_6 mouse monoclonal antibody (ProteinTech). The secondary antibody, used to detect the anti-His_6 antibody, was anti-mouse IgG, HRP-linked horse antibody (Cell Signaling Technology). Western blot transfer sandwiches consisted of a nitrocellulose membrane (Amersham™ Protrane 0.2 M) and Whatman 3 mm cellulose chromatography papers (Whatman). Amersham enhanced chemiluminescence (ECL) detection reagents were purchased from Cytivia. Skimmed milk powder (Tesco) was dissolved in PBS-T for use as a blocking solution. Buffers were prepared as described in Table 7-13.

Table 7-13. Composition of buffers used in Western Blot analysis

Buffer	Components
Transfer buffer	25 mM Tris 200 mM glycine 20% v/v Methanol
PBS-T	10 mM Na_2HPO_4 1.8 mM KH_2PO_4 137 mM NaCl 2.7 mM KCl 0.1% w/v Tween
Blocking solution	5% skimmed milk dissolved in PBS-T

Membrane transfer

SDS-PAGE gels were prepared as previously described. A transfer sandwich comprising a sponge, 3 sheets of Whatman paper, the membrane, 3 sheets of Whatman paper and a sponge. The transfer was run in ice cold transfer buffer at 60V for 1 hour. Protein transfer was checked by staining of the membrane in Ponceau-S stain for 5 minutes, after which the stain was removed and the membrane was washed with PBS-T.

Membrane treatment

The membrane treatment steps were performed on a shaker unless stated otherwise. The membrane was incubated with blocking solution for 1 hour at room temperature. The blocking solution was then discarded and a new solution containing the anti-His_6 primary antibody in blocking solution (1:1000 (v/v) dilution) was added to the membrane. The membrane was incubated with the primary antibody for 18 hours at 4°C. The primary antibody solution was removed and the membrane was then washed with PBS-T (3x 10minute washes). A solution of secondary antibody (anti-mouse) in blocking solution (1:10,000 (v/v) dilution) was then added to the membrane. The membrane was incubated with the secondary antibody for 1 hour at room temperature. The membrane was washed again (3x 10 minute PBS-T washes). The membrane was developed with ECL solution, mixed 1:1 (v/v) according to the manufacturer's protocol. The

membrane was incubated in the dark for 5 minutes before imaging in a Bio-Rad ChemiDoc MP Universal Hood III.

7.4 Crystallography

7.4.1 Crystallisation buffer preparation

Buffer stocks were prepared from chemical solids dissolved in MilliQ and then filtered through 0.2 µm filters. 50% PEG solutions were purchased pre-made from Hampton Research.

7.4.2 Seed stock preparation

In order to improve crystal quality, grow crystals of some AspH variants, or obtain slurries of microcrystals, seeding methods were used.

Seed Bead method

Seeds from single crystals were prepared by using a PTFE seed bead (Hampton, HR2-320) and vortexing for 30 s, followed by 30 s incubation on ice for a total of 10 cycles.

BeadBug method

Seeds from batch crystal wells were prepared using a BeadBug Microtube homogeniser (Benchmark Scientific). Batch crystal wells were pooled and transferred into 2 mL capped tubes containing 3 mm diameter zirconium beads (Benchmark Scientific). The crystals were then crushed by the BeadBug for 60 s intervals with 60 s rest on ice until the seed stock was judged to be homogeneous by examination under the microscope. Crushing was performed at 4°C.

7.4.3 Crystallisation methods

Enzyme preparation

AspH and all AspH variants were typically crystallised with 2OG, Mn and substrate. The method for preparing the enzyme, cofactor and substrate solution was as follows: A 60 µL enzyme aliquot (at 20.4 mg mL^{-1}) was thawed on ice. 1.5 µL of 100 mM Mn solution and 1.5 µL of 100 mM 2OG solution was added to the enzyme solution. The enzyme and cofactor mixture was added to 0.3 mg of solid substrate. The mixture was incubated on ice for 15 minutes.

Broad screens

Broad crystallisation screens were dispensed (25µL) into 3 lens 96 well low-profile crystallisation plate reservoirs (SwisSci) by an Integra Viaflo 384 96 channel pipette. Protein and precipitant drops were set in wells by Mosquito (SPTLab Tech), in ratios 1:1, 1:2 and 2:1 (v/v) of protein: precipitant.

Table 7-14. Broad crystallisations screens used.

Broad Screen	Manufacturer
Pact Premier	Molecular Dimensions
PEG/ION	Hampton Research
BCS	Molecular Dimensions
LMB	Molecular Dimensions

Sitting drop trays

Conditions which produced crystals in 96 well broad screen trays were used for further crystal growth in 24 well sitting drop trays (Hampton). A screening tray was developed using Hampton "Make a Tray" programme, with optimisation of two parameters along the x and y directions. Well solutions were prepared from stock solutions, as directed by Hampton. 200 µL of solution was added to each reservoir and total volumes of 3 µL per drop of protein and precipitant were used.

Hanging drop trays

Hanging drop crystallisation trays were set up manually to increase the size of crystals grown. 24 well VDX plates with pre-greased wells were used for hanging drop crystallisation. A precipitant screen, based around a successful 96 well broad screen conditions, was developed using the Hampton "Make a Tray" programme, with variation in two parameters. 500 µL of precipitant solution was added to the hanging drop tray reservoirs. 3 µL crystal drops were set up on OptiClear plastic circle cover slides (Hampton).

Batch crystallisation

Crystal slurries for room temperature serial synchrotron X-ray diffraction (SSX) experiments were prepared using seed stocks that were first generated from single crystals using the Seed Bead

method. Seeds were prepared as described previously. A precipitant solution (0.1 M KSCN, 0.1 M bis-tris propane pH 7.5 and 16% PEG 3350) was mixed with the protein and cofactor solution (prepared as above) and seed stock in the ratio 2:1:9 (protein:seed stock:precipitant) (v/v/v). The mixture was aliquoted into 30 µL batches in flat bottom 96 well small volume chimney well plates (Greiner), the plates were sealed and then incubated at 4°C for 18 hours.

To generate larger volumes of seed stock, the 30 µL batch wells were then used as seed stock for further rounds of crystallisation. Seed stock was prepared using a BeadBug as described previously. The protein, precipitant and seed mixture were prepared as described above. Batch wells were scaled up to 60 µL volumes in larger volume chimney well plates. Trays were sealed and incubated at 4°C for 16 hours.

Seeded broad crystallisation screens

A single 300 nL well of wild-type AspH crystals from a broad screen SwisSci plate was harvested in 30 µl of well solution, and crushed using a SeedBead kit as previously described. This stock was then diluted 1:50 (v/v) in the precipitant well solution and used as a seed stock. Seeded 96-well broad screens were set up using the Mosquito liquid handling robot. 30 nL of seed stock was mixed with 3 different ratios of protein and precipitant solution, where seed stock was used in place of 30 nL of precipitant solution.

7.4.4 Crystal harvesting

Crystals were harvested using cryoloops (Hampton) into liquid N_2 and stored at cryogenic temperatures. Crystals were cryoprotected in solutions of 30% PEG 400 and 70% precipitant solution (v/v).

Cofactor soaking

Soaking was used to aid binding of cofactors at the active site in crystals. A mixture of 70% well solution and 30% PEG 400 (v/v) was prepared as cryoprotectant and then a 100 mM stock of cofactor was diluted 1:10 (v/v) in the cryoprotectant solution. Crystals were moved into drops of

the 10 mM cofactor in cryoprotectant solution and allowed to incubate in the solution for 5-10 minutes before freezing in liquid N_2.

7.4.5 Data collection and processing

Data collection
Data were collected at 100K at Diamond Light Source using beamlines I03 or I04 using the unattended data collection mode.

Data processing
Data were indexed, integrated and scaled using either the xia2 dials or xia2 3dii data processing pipelines to give $CC_{1/2}$ values greater than 0.5 and/or I/σ to greater than 1. AspH structures were solved by molecular replacement using the PhaserMR function in PHENIX.[357] The search models for structures were either PDB ID 6YYW, or a modified version of PDB ID 6YYX in which the cofactors Mn and 3-methyl 2OG were removed before use.

Refinements
Crystal structure models were refined by iterative cycles of crystallographic refinement using phenix.refine and manual modification in COOT.[264,358] Refinement details are summarised for each crystal structures in tables in the corresponding chapter supplementary information sections.

7.4.6 Anaerobic crystallography

Anaerobic experiments were carried out in a glovebox (Belle Technologies). All solutions were saturated with gaseous N_2 and left in the anaerobic chamber port overnight to ensure oxygen removal. Protein samples were transferred to the glovebox before use and purged in the transfer port. Solid, pre-weighed cofactor aliquots were transferred to the chamber and solutions were made using deoxygenated MiliQ water. AspH crystallisation trays were incubated at 4°C by means of a fridge inside the glovebox.

7.5 Analytical methods

7.5.1 RapidFire MS

MS analyses of assays were carried out using a RapidFire RF 365 high-throughput sampling robot (Agilent) attached to an iFunnel Agilent 6550 accurate mass quadrupole time-of-flight (Q-TOF) mass spectrometer. The mass spectrometer was operated in the positive ionization mode. Samples were injected onto a C4 solid phase extraction (SPE) cartridge following aspiration under vacuum for 0.6s). The cartridge was washed with 0.1%$_{v/v}$ formic acid (5.52 s, 1.5 mL/min), then elution of the peptide was achieved by washing the cartridge with 0.1%$_{v/v}$ formic acid in 80/20$_{v/v}$ acetonitrile/water (5.5 s, 1.5 mL/min). The parameters for subsequent mass spectrometer analysis are outlined in Table 7-15.

Table 7-15. Mass spectrometer parameters for analyses of AspH variant inhibition assays.

Parameter	Value
Capillary voltage	4000 V
Nozzle voltage	1000 V
Fragmentor voltage	365 V
Gas temperature	280C
Gas flow	13 L/min
Sheath gas temperature	250 C
Sheath gas flow	12 L/min

The *m*/z +4 charge states of the substrate and product were used to extract ion chromatogram data and peak areas were integrated using RapidFire Integrator software (Agilent). The data were exported as a csv file and analysed in Microsoft Excel. The turnover % was calculated by using the equation: $\%\ Turnover = 100\ X\ \frac{Product}{Product+Substrate}$

7.5.2 Xevo MS

High-performance liquid chromatography (HPLC) grade solvents were obtained from Sigma-Aldrich. LC-MS analyses on AspH ^{18}O labelled assays were performed using an Acquity UPLC system coupled to a Xevo G2-S Q-TOF mass spectrometer. 5 µL of sample was injected onto an ACE Equivalence 3 C18 column (50 x 2.1 mm) using an Agilent 1290 series autosampler. Samples were eluted by washing with water and MeCN, supplemented with 0.1%$_{v/v}$ formic acid, at a flow rate of 0.3 mL/min. The elution gradient was 95 % water (0 – 1 min), 5-95 % (1 - 8 min), 95 % water (8 - 10 min). Data were processed using MassLynx MS software (Waters).

7.5.3 NMR

Oxidative decarboxylation of 2OG uncoupled from substrate oxidation was measured using a Bruker AVIII 700 NMR machine combined with a 5mm ^{1}H ($^{13}C/^{15}N$) inverse cryoprobe. Assays contained 200 µM 2OG, 20 µM FAS and 1 µM enzyme in 50 mM aqueous Tris-d_{11} containing 10%$_{v/v}$ D2O pH 7.5.

7.6 Assays

7.6.1 General procedures

Substrate

Synthetic substrates based on the recognition sequence of human Factor X (hFX) were designed as substrates for AspH assays. Peptides were synthesised by GL Biochem Shanghai.

General assay conditions

Assay conditions were optimised for each AspH variant. A selection of buffer pH and concentrations were screened to determine the conditions under which each variant was most active. Unless otherwise stated, the assay conditions are outlined in Table 7-16.

Table 7-16. Assay conditions used for AspH and variant assays. Variants marked with * were not fully optimised but conditions shown were used for initial tests of activities. 2OG – 2-oxoglutarate; LAA – L-ascorbic acid; FAS – ferrous ammonium iron sulphate; hFX-Asp_{103} – synthetic AspH substrate based on the sequence of human coagulation factor X.

Enzyme	Buffer composition	Final concentration of reactant (*μM*)				
		LAA	2OG	hFX-Asp_{103}	FAS	Enzyme
WT AspH	25 mM HEPES pH 7.5, 50 mM NaCl	100	20	4	20	0.1
H679A	25 mM HEPES pH 7.5	100	2000	4	20	0.2
H725A	25 mM HEPES pH 7.5	100	2000	4	20	0.2
G681D*	25 mM HEPES pH 7.5	100	2000	4	20	0.2
G681E*	25 mM HEPES pH 7.5	100	2000	4	20	0.2
G434V	25 mM HEPES pH 7.5, 50 mM NaCl	100	20	4	20	0.1
R688Q	25 mM MES pH 6.0	400	30	4	100	0.2
R735Q	25 mM MES pH 6.0	400	2000	4	20	0.2
H679R*	25 mM HEPES pH 7.5	100	2000	4	20	0.2

7.6.2 RapidFire timecourse assays

For the measurement of enzyme activity and the determination of kinetic parameters, assays were set up as reported.[228] A substrate mixture, containing LAA, 2OG, FAS and substrate at 1.2x the final concentration was made up in a well of a 96 2 mL DeepWell plate. A single sample injection was taken using a RapidFire RF 365 high-throughput sampling robot (Agilent) attached to an iFunnel Agilent 6550 accurate mass quadrupole time-of-flight (Q-TOF) mass spectrometer. The enzyme solution was then added to give a final concentration of enzyme and cofactors as stated in Table 7-16. The wells were thoroughly mixed and then the level of hydroxylation was monitored by regular injections into the (Q-TOF) using the RapidFire sampling robot. The time was normalised to the first time point containing enzyme. Data were analysed as described previously. The gradient

of the initial reaction was calculated using non-linear regression fitting in GraphPad Prism 5. The kinetic parameters were then calculated using the GraphPad Prism function for Michaelis Menten kinetics.

7.6.3 AspH variant inhibition assays

Inhibitor solutions were made up in 100% DMSO. The inhibitors were dispensed across 384-well polypropylene assay plates (Greiner) using an ECHO 550 acoustic dispenser (Labcyte) to give an 11-point dilution series with a top concentration of 100 µM. Each reaction was performed in technical duplicates and each assay plate was carried out twice to give independent duplicates.

25 µL of an enzyme mixture comprising H679A or H725A variant in 25 mM HEPES pH 7.5 buffer was dispensed across the inhibitor plates using a multidrop dispenser (ThermoFischer Scientific). The plates were centrifuged (1000 rpm, 10s) and incubated for 15 minutes. 25 µL of a substrate mixture containing LAA (200 µM), 2OG (220 µM for H679A, 430 µM for H725A), FAS (8 µM for H679A, 20 µM for H725A) and hfX-Asp_{103} (8 µM) in 25 mM HEPES pH 7.5 was then dispensed across the plate using the multidrop dispenser. The plates were centrifuged (1000 rpm, 10s) and incubated at room temperature (22 minutes for H679A and 37 minutes for H725A). The reaction was quenched by the addition of 5 µL of 10% (v/v) formic acid per well. The plates were centrifuged (1000 rpm, 20 s) and analysed by MS as described previously.

The mass of the m/z+4 charge state of the substrate and hydroxylated product was used to extract ion chromatogram data. The areas of the substrate and product peaks were integrated using RapidFire Integrator software (Agilent) and data were analysed in Microsoft Excel to calculate the % conversion of hydroxylation. IC50-values were calculated in GraphPad Prism 5 by first obtaining normalized dose-response curves for 2,4-PDBA and DMSO, and then non-linear regression of the raw data.

7.6.4 2OG derivative assays

2OG-derivatives were synthesised by Dr Lennart Brewitz.[258] End-point assays with 2OG derivatives and the R688Q and R735Q variants were performed in a 384 well plate. 30 µL of 600 µM of a 2OG-derivative solution was dispensed across the wells according to a plate map of derivatives. 30 µL of a cofactors mixture comprising 3x the final concentration of LAA, 2OG and FAS was dispensed across the plate using a liquid dispenser (final concentrations and buffers outlined in Table 7-16). Reactions were initiated by addition of 30 µL of a 3x concentration enzyme solution using a liquid dispenser (final concentration in Table 7-16). Reactions were quenched with 10 µL of 10% (v/v) formic acid after 30 minutes or 1 hour. Assays were performed in technical duplicates. Assays were analysed using RapidFire SPE-MS as described previously.

7.6.5 Oxygen isotope studies

$^{18}O_2$ assays

200 µL assays solutions containing LAA (100 µM), 2OG (10 µM for wt AspH, 2000 µM for H679A and H725A variants), FAS (20 µM), hFX-Asp_{103} (5 µM) and enzyme (0.1 µM for WT AspH, 0.2 µm for H679A and H725A variants) were prepared in 500 µL microfuge tubes in an anaerobic chamber. All cofactor solutions were made up using degassed MilliQ and enzyme stocks were incubated in the anaerobic chamber prior to use. The solutions were transferred to a pear-shaped flask with an outlet valve. The flask was sealed with a rubber bung and the outlet valve was closed. The flask was then removed from the anaerobic chamber.

A schematic representation of the experimental-set up is shown in Figure 7-1 and a photograph is shown in Figure 7-2. The outlet of the pear-shaped flask was connected to a custom-made glass manifold (Figure 7-1, blue glass manifold). The custom-made glass manifold was connected to an N_2 line, an $^{18}O_2$ cylinder and a vacuum pump. The flask valve was kept shut and the system was placed under vaccum, then refilled with N_2 using the N_2 gas connection. The process of placing the

system under vaccum and refilling with N_2 was repeated three times to ensure there were no traces of $^{16}O_2$ in any of the lines.

The system was then placed under vacuum and then N_2 was allowed to flow into the system until a pressure of around 700 mbar was reached. The valve to the flask was then opened to connect the pear-shaped flask to the system. The $^{18}O_2$ cylinder was opened until the pressure reached 1000 mbar to mimic atmospheric oxygen concentrations, and the mixture of N_2 and $^{18}O_2$ was allowed to flow into the pear-shaped flask. The flask valve was closed again and the assays were incubated with $^{18}O_2$ for 3 hours. Controls with $^{16}O_2$ were performed in the same method but with a $^{16}O_2$ cylinder attached to the glass gas line in place of the $^{18}O_2$ cylinder. Samples were quenched with 10% (v/v) formic acid in the anaerobic chamber using a needle and were therefore not exposed to atmospheric oxygen until fully quenched. Samples were analysed by LC-MS.

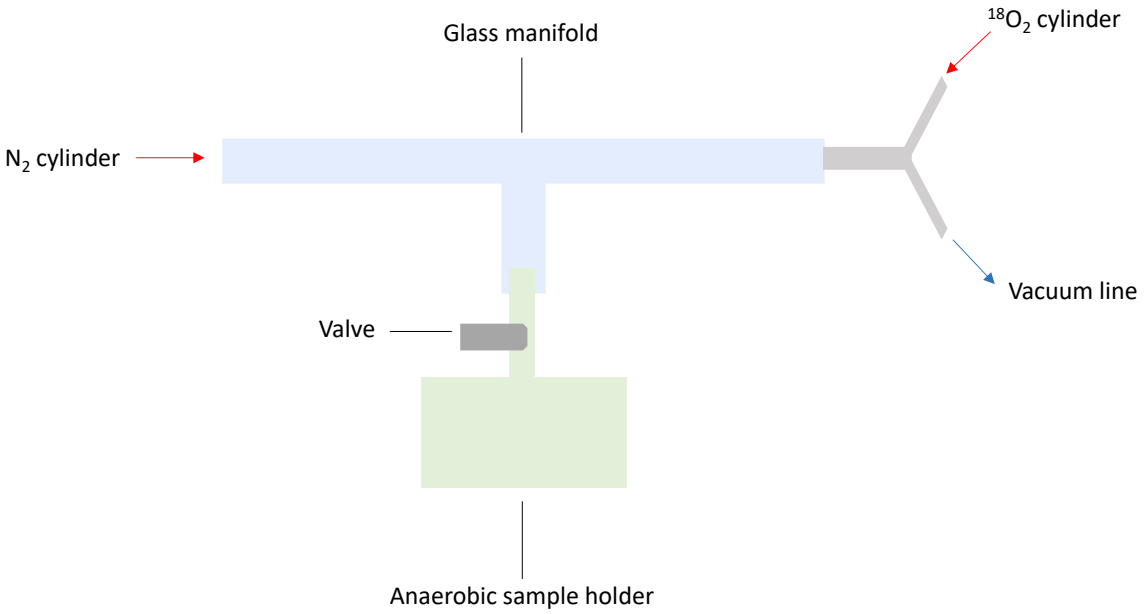


Figure 7-1. A schematic representation of the experimental set-up for performing $^{18}O_2$ labelling experiments. Assays were prepared anaerobically in either eppenderforf tubes or wells of a 96-well plate and sealed in an anaerobica sample holder (green). The anaerobic sample holder contained and outlet tube which could be opened with a valve (valve). The anaerobic chamber outlet tube was connected to the glass manifold (blue), which was connected to N_2 and $^{18}O_2$ gas cylinders, and a vacuum line.

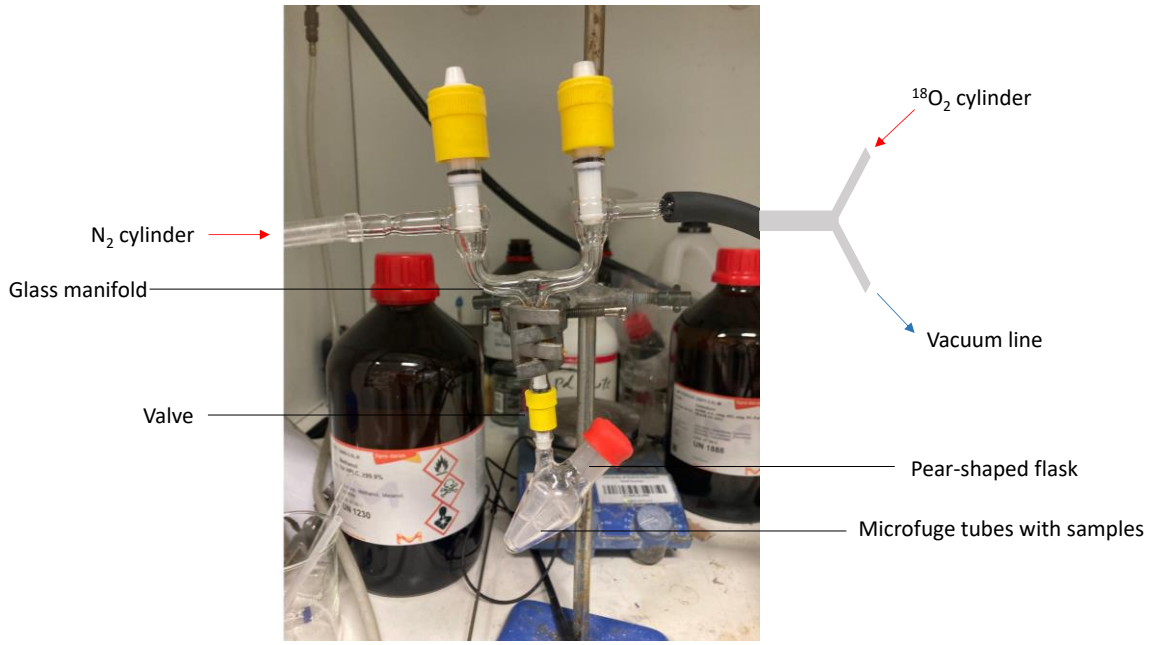


Figure 7-2. Photograph of the actual experimental set up illustrated in Figure 7-1.

pH dependence of $^{18}O_2$ assays

The pH dependence of $^{18}O_2$ assays were carried out using virtually the same experimental set-up as described above, however, the samples were contained in a 96-well plate instead of microfuge tubes. The 96-well plate was placed in a custom made anaerobic plate holder, which has an outlet line. A valve on the outline line controlled gas flow in and out of the anaerobic plate holder box.

$^{18}OH_2$ assays

All cofactor solutions and substrate stocks were prepared using $^{18}OH_2$ (CK Isotopes). Solutions of 50 mM HEPES pH 7.5 and 50 mM HEPES pH 7.5 with 50 mM NaCl in $^{18}OH_2$ were prepared, and the enzyme was buffer exchanged into the desired buffer before use. 200 µL assays solutions containing LAA (100 µM), 2OG (10 µM for WT AspH, 2000 µM for H679A and H725A variants), FAS (20 µM), hfX-Asp_{103} (5 µM) and enzyme (0.1 µM for wt AspH, 0.2 µm for H679A and H725A variants) were set up. The reaction was quenched by addition of 10% (v/v) formic acid after 2.5 hours and samples were analysed by LC-MS.

Chapter 8 | References

1. Goldman, A. D. & Kacar, B. Cofactors are Remnants of Life's Origin and Early Evolution. *J. Mol. Evol.* **89**, 127–133 (2021).
2. Malmstrom, B. G. & Neilands, J. B. Metalloproteins. *Annu. Rev. Biochem.* **33**, 331–354 (1964).
3. Atkinson, J. T. *et al.* Metalloprotein switches that display chemical-dependent electron transfer in cells. *Nat. Chem. Biol.* **15**, 189–195 (2019).
4. Gray, H. B. & Winkler, J. R. Electron flow through metalloproteins. *Biochim. Biophys. Acta - Bioenerg.* **1797**, 1563–1572 (2010).
5. Chim, N., Johnson, P. M. & Goulding, C. W. Insights into redox sensing metalloproteins in Mycobacterium tuberculosis. *Journal of Inorganic Biochemistry* vol. 133 118–126 at (2014).
6. Coates, C. J. & Costa-Paiva, E. M. Multifunctional Roles of Hemocyanins. *Subcell. Biochem.* **94**, 233–250 (2020).
7. Sousa, E. H. S. & Tian, S. Metalloproteins as sensors of gaseous small molecules—From bench to bed and beyond. *Front. Mol. Biosci.* **10**, 1140392 (2023).
8. Aono, S. Novel bacterial gas sensor proteins with transition metal-containing prosthetic groups as active sites. *Antioxidants Redox Signal.* **16**, 678–686 (2012).
9. Woods, J. J. & Wilson, J. J. Bioinorganic Chemistry of Hydrogen Sulfide: Detection, Delivery, and Interactions with Metalloproteins. in *Encyclopedia of Inorganic and Bioinorganic Chemistry* 1–22 (John Wiley & Sons, Ltd, 2021).
10. Williams, M. T. *et al.* Metalloprotein enabled redox signal transduction in microbes. *Curr. Opin. Chem. Biol.* **76**, (2023).
11. Lindskog, S. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.* **74**, 1–20 (1997).
12. Marathe, I. A. *et al.* Protein cofactors and substrate influence Mg^{2+}-dependent structural changes in the catalytic RNA of archaeal RNase P. *Nucleic Acids Res.* **49**, 9444–9458 (2021).
13. Supuran, C. T. Structure and function of carbonic anhydrases. *Biochem. J.* **473**, 2023–2032 (2016).
14. Christianson, D. W. Structural biology of zinc. *Adv. Protein Chem.* **42**, 281–355 (1991).
15. Gomes, C. M. & Wittung-Stafshede, P. *Protein Folding and Metal Ions*. *Protein Folding and Metal Ions* (CRC Press, 2016).
16. Waldron, K. J., Rutherford, J. C., Ford, D. & Robinson, N. J. Metalloproteins and metal sensing. *Nat. 2009 4607257* **460**, 823–830 (2009).
17. Mendel, R. R. The History of the Molybdenum Cofactor—A Personal View. *Molecules* **27**, (2022).
18. Liu, J. *et al.* Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem. Rev.* **114**, 4366–4469 (2014).
19. Dunham, N. P. & Arnold, F. H. Nature's Machinery, Repurposed: Expanding the Repertoire of Iron-Dependent Oxygenases. *ACS Catal.* **10**, 12239–12255 (2020).
20. Frey, P. A. & Reed, G. H. The ubiquity of iron. *ACS Chem. Biol.* **7**, 1477–1481 (2012).
21. Georgiadis, M. M. *et al.* Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii. *Science (80-.).* **257**, 1653–1659 (1992).
22. Morra, S. Fantastic [FeFe]-Hydrogenases and Where to Find Them. *Front. Microbiol.* **13**, 853626 (2022).
23. Rouault, T. A. Mammalian iron–sulphur proteins: novel insights into biogenesis and function. *Nat. Rev. Mol. Cell Biol. 2014 161* **16**, 45–55 (2014).
24. Giardina, B., Messana, I., Scatena, R. & Castagnola, M. The multiple functions of hemoglobin. *Crit. Rev. Biochem. Mol. Biol.* **30**, 165–196 (1995).
25. Solomon, E. I., Decker, A. & Lehnert, N. Non-heme iron enzymes: Contrasts to heme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3589–3594 (2003).
26. Sono, M., Roach, M. P., Coulter, E. D. & Dawson, J. H. Heme-containing oxygenases. *Chem. Rev.* **96**, 2841–2887 (1996).

27. Ferguson-Miller, S. & Babcock, G. T. Heme/copper terminal oxidases. *Chem. Rev.* **96**, 2889–2907 (1996).
28. Nienhaus, K. & Nienhaus, G. U. Probing Heme Protein-Ligand Interactions by UV/Visible Absorption Spectroscopy. in *Methods in Molecular Biology* vol. 305 215–241 (Humana Press Inc., 2005).
29. Hu, S., Spiro, T. G., Morris, I. K., Singh, J. P. & Smith, K. M. Complete Assignment of Cytochrome C Resonance Raman Spectra via Enzymatic Reconstitution with Isotopically Labeled Hemes. *J. Am. Chem. Soc.* **115**, 12446–12458 (1993).
30. Kitagawa, T., Kyogoku, Y., Iizuka, T. & Saito, M. I. Nature of the Iron-Ligand Bond in Ferrous Low Spin Hemoproteins Studied by Resonance Raman Scattering. *J. Am. Chem. Soc.* **98**, 5169–5173 (1976).
31. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662–666 (1958).
32. Ryle, M. J. & Hausinger, R. P. Non-heme iron oxygenases. *Curr. Opin. Chem. Biol.* **6**, 193–201 (2002).
33. Dunwell, J. M. Cupins: A New Superfamily of Functionally Diverse Proteins that Include Germins and Plant Storage Proteins. *Biotechnol. Genet. Eng. Rev.* **15**, 1–32 (1998).
34. Stipanuk, M. H., Simmons, C. R., Karplus, P. A. & Dominy, J. E. Thiol dioxygenases: Unique families of cupin proteins. *Amino Acids* **41**, 91–102 (2011).
35. Fetzner, S. Ring-cleaving dioxygenases with a cupin fold. *Appl. Environ. Microbiol.* **78**, 2505–2514 (2012).
36. Aik, W. S., McDonough, M. A., Thalhammer, A., Chowdhury, R. & Schofield, C. J. Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr. Opin. Struct. Biol.* **22**, 691–700 (2012).
37. Dunwell, J. M., Khuri, S. & Gane, P. J. Microbial Relatives of the Seed Storage Proteins of Higher Plants: Conservation of Structure and Diversification of Function during Evolution of the Cupin Superfamily. *Microbiol. Mol. Biol. Rev.* **64**, 153–179 (2000).
38. Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R. & Goodenough, P. W. Evolution of functional diversity in the cupin superfamily. *Trends Biochem. Sci.* **26**, 740–746 (2001).
39. Agarwal, G., Rajavel, M., Gopal, B. & Srinivasan, N. Structure-based phylogeny as a diagnostic for functional characterization of proteins with a cupin fold. *PLoS One* **4**, e5736 (2009).
40. A., F., Banerjee, A. & Sljivic, B. Modeling the Metal Binding Site in Cupin Proteins. in *On Biomimetics* (IntechOpen, 2011).
41. Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: The most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004).
42. Woo, E. J., Dunwell, J. M., Goodenough, P. W., Marvier, A. C. & Pickersgill, R. W. Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat. Struct. Biol.* **7**, 1036–1040 (2000).
43. Straganz, G. D. & Nidetzky, B. Variations of the 2-His-1-carboxylate theme in mononuclear non-heme Fe II oxygenases. *ChemBioChem* **7**, 1536–1548 (2006).
44. Fujieda, N. *et al.* Cupin Variants as a Macromolecular Ligand Library for Stereoselective Michael Addition of Nitroalkanes. **59**, 7717–7720 (2020).
45. Matsumoto, R. *et al.* An artificial metallolyase with pliable 2-His-1-carboxylate facial triad for stereoselective Michael addition. *Chem. Sci.* **14**, 3932–3937 (2023).
46. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368–371 (2006).
47. Hillringhaus, L. *et al.* Structural and evolutionary basis for the dual substrate selectivity of human KDM4 histone demethylase family. *J. Biol. Chem.* **286**, 41616–41625 (2011).
48. Clifton, I. J., Hsueh, L. C., Baldwin, J. E., Harlos, K. & Schofield, C. J. Structure of proline 3-hydroxylase: Evolution of the family of 2-oxoglutarate dependent oxygenases. *Eur. J.*

Biochem. **268**, 6625–6636 (2001).
49. McCoy, J. G. *et al.* Structure and mechanism of mouse cysteine dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 3084–3089 (2006).
50. Axelrod, H. L. *et al.* Conformational changes associated with the binding of zinc acetate at the putative active site of XcTcmJ, a cupin from Xanthomonas campestris pv. campestris. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1347–1353 (2010).
51. Kloer, D. P., Ruch, S., Al-Babili, S., Beyer, P. & Schulz, G. E. The structure of a retinal-forming carotenoid oxygenase. *Science (80-.).* **308**, 267–269 (2005).
52. Fusetti, F. *et al.* Crystal Structure of the Copper-Containing Quercetin 2,3-Dioxygenase from Aspergillus japonicus. *Structure* **10**, 259–268 (2002).
53. Rajavel, M., Mitra, A. & Gopal, B. Role of Bacillus subtilis BacB in the synthesis of bacilysin. *J. Biol. Chem.* **284**, 31882–31892 (2009).
54. Cleasby, A. *et al.* The X-ray crystal structure of phosphomannose isomerase from Candida albicans at 1.7 Å resolution. *Nat. Struct. Biol.* **3**, 470–479 (1996).
55. Islam, M. S., Leissing, T. M., Chowdhury, R., Hopkinson, R. J. & Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu Rev Biochem* **20**, 585–620 (2018).
56. Lau, W. & Sattely, E. S. Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone. *Science (80-.).* **349**, 1224–1228 (2015).
57. Zhu, Q. & Liu, X. Characterization of non-heme iron aliphatic halogenase WelO5 from Hapalosiphon welwitschii IC-52-3: Identification of a minimal protein sequence motif that confers enzymatic chlorination specificity in the biosynthesis of welwitindolelinones. *Beilstein J. Org. Chem.* **13**, 1168–1173 (2017).
58. Siitonen, V. *et al.* Divergent non-heme iron enzymes in the nogalamycin biosynthetic pathway. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5251–5256 (2016).
59. Mader, S. L., Bräuer, A., Groll, M. & Kaila, V. R. I. Catalytic mechanism and molecular engineering of quinolone biosynthesis in dioxygenase AsqJ. *Nat. Commun.* **9**, 1–8 (2018).
60. Duncan, T. *et al.* Reversal of DNA alkylation damage by two human dioxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16660–16665 (2002).
61. McLuskey, K., Cameron, S., Hammerschmidt, F. & Hunter, W. N. Structure and reactivity of hydroxypropylphosphonic acid epoxidase in fosfomycin biosynthesis by a cation- and flavin-dependent mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 14221–14226 (2005).
62. Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat. Commun.* **7**, 1–12 (2016).
63. Loenarz, C. *et al.* The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO Rep.* **12**, 63–70 (2011).
64. Chowdhury, R. *et al.* Structural basis for oxygen degradation domain selectivity of the HIF prolyl hydroxylases. *Nat. Commun.* **7**, (2016).
65. Hewitson, K. S. *et al.* 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 (2002).
66. Hegg, E. L. & Que, L. The 2-His-1-carboxylate facial triad - An emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur. J. Biochem.* **250**, 625–629 (1997).
67. Markolovic, S. *et al.* Structure-function relationships of human JmjC oxygenases - demethylases versus hydroxylases. *Curr. Opin. Struct. Biol.* **41**, 62–72 (2016).
68. Galonić, D. P., Barr, E. W., Walsh, C. T., Bollinger, J. M. & Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. **3**, 113–116 (2007).
69. Zhao, C. *et al.* An Fe^{2+}- and α-Ketoglutarate-Dependent Halogenase Acts on Nucleotide Substrates. *Angew. Chemie - Int. Ed.* **59**, 9478–9484 (2020).
70. Neugebauer, M. E. *et al.* A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019).
71. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme FeII α-ketoglutarate- and O2-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **102**,

10111–10116 (2005).
72. Eng, J. G. M. *et al.* A Catharanthus roseus Fe(II)/α-ketoglutarate-dependent dioxygenase catalyzes a redox-neutral reaction responsible for vindolinine biosynthesis. *Nat. Commun.* **13**, 1–9 (2022).
73. Van Pée, K. H. & Unversucht, S. Biological dehalogenation and halogenation reactions. *Chemosphere* **52**, 299–312 (2003).
74. Agarwal, V. *et al.* Enzymatic Halogenation and Dehalogenation Reactions: Pervasive and Mechanistically Diverse. *Chem. Rev.* **117**, 5619–5674 (2017).
75. Martinez, S. & Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* **290**, 20702–20711 (2015).
76. Hanauske-Abel, H. M. & Günzler, V. A stereochemical concept for the catalytic mechanism of prolylhydroxylase. Applicability to classification and design of inhibitors. *J. Theor. Biol.* **94**, 421–455 (1982).
77. Zhang, Z. *et al.* Structural and stereoelectronic insights into oxygenase-catalyzed formation of ethylene from 2-oxoglutarate. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4667–4672 (2017).
78. Zhang, Z. *et al.* Crystal structure of a clavaminate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: Evidence for metal centred rearrangements. *FEBS Lett.* **517**, 7–12 (2002).
79. Mantri, M. *et al.* Crystal structure of the 2-Oxoglutarate- and Fe(II)-dependent lysyl hydroxylase JMJD6. *J. Mol. Biol.* **401**, 211–222 (2010).
80. Ryle, M. J., Padmakumar, R. & Hausinger, R. P. Stopped-flow kinetic analysis of Escherichia coli taurine/α- ketoglutarate dioxygenase: Interactions with α ketoglutarate, taurine, and oxygen. *Biochemistry* **38**, 15278–15286 (1999).
81. Sánchez-Fernández, E. M. *et al.* Investigations on the oxygen dependence of a 2-oxoglutarate histone demethylase. *Biochem. J.* **449**, 491–496 (2013).
82. Flashman, E. *et al.* Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen. *FEBS J.* **277**, 4089–4099 (2010).
83. Ho, R. Y. N. *et al.* Resonance Raman studies of the iron(II)-α-keto acid chromophore in model and enzyme complexes. *J. Am. Chem. Soc.* **123**, 5022–5029 (2001).
84. Neidig, M. L. *et al.* CD and MCD of CytC3 and taurine dioxygenase: Role of the facial triad in α-KG-dependent oxygenases. *J. Am. Chem. Soc.* **129**, 14224–14231 (2007).
85. Grzyska, P. K. *et al.* Steady-state and transient kinetic analyses of taurine/α- ketoglutarate dioxygenase: Effects of oxygen concentration, alternative sulfonates, and active-site variants on the Fe(IV)-oxo intermediate. *Biochemistry* **44**, 3845–3855 (2005).
86. Rabe, P. *et al.* X-ray free-electron laser studies reveal correlated motion during isopenicillin N synthase catalysis. *Sci. Adv.* **7**, (2021).
87. Mitchell, A. J. *et al.* Visualizing the reaction cycle in an Iron(II)- and 2-(Oxo)-glutarate-dependent hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017).
88. Copeland, R. A. *et al.* Hybrid radical-polar pathway for excision of ethylene from 2-oxoglutarate by an iron oxygenase. *Science (80-.).* **373**, 1489–1493 (2021).
89. Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. & Krebs, C. The first direct characterization of a high-valent iron intermediate in the reaction of an α-ketoglutarate-dependent dioxygenase: A high-spin Fe(IV) complex in taurine/α-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **42**, 7497–7508 (2003).
90. Riggs-Gelasco, P. J. *et al.* EXAFS spectroscopic evidence for an Fe=O unit in the Fe(IV) intermediate observed during oxygen activation by taurine:α-ketoglutarate dioxygenase. *J. Am. Chem. Soc.* **126**, 8108–8109 (2004).
91. Bassan, A., Borowski, T. & Siegbahn, P. E. M. Quantum chemical studies of dioxygen activation by mononuclear non-heme iron enzymes with the 2-His-1-carboxylate facial triad. *Dalt. Trans.* 3153–3162 (2004).
92. Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M. & Krebs, C. Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl.*

Acad. Sci. U. S. A. **103**, 14738–14743 (2006).
93. Welford, R. W. D. *et al.* Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate dependent oxygenases from bacteria, plants and humans. *FEBS Lett.* **579**, 5170–5174 (2005).
94. Price, J. C., Barr, E. W., Glass, T. E., Krebs, C. & Bollinger, J. M. Evidence for Hydrogen Abstraction from C1 of Taurine by the High-Spin Fe(IV) Intermediate Detected during Oxygen Activation by Taurine:α -Ketoglutarate Dioxygenase (TauD). *J. Am. Chem. Soc.* **125**, 13008–13009 (2003).
95. Price, J. C., Barr, E. W., Glass, T. E., Krebs, C. & Bollinger, J. M. Evidence for Hydrogen Abstraction from C1 of Taurine by the High-Spin Fe(IV) Intermediate Detected during Oxygen Activation by Taurine:α -Ketoglutarate Dioxygenase (TauD). *J. Am. Chem. Soc.* **125**, 13008–13009 (2003).
96. Zhai, G. *et al.* Structural Insight into the Catalytic Mechanism of Non-Heme Iron Halogenase AdaV in 2′-Chloropentostatin Biosynthesis. *ACS Catal.* **12**, 13910–13920 (2022).
97. Wong, S. D. *et al.* Elucidation of the Fe(iv)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320–323 (2013).
98. Ludewig, H. *et al.* Halogenases: structures and functions. *Curr. Opin. Struct. Biol.* **65**, 51–60 (2020).
99. Wilson, R. H., Chatterjee, S., Smithwick, E. R., Dalluge, J. J. & Bhagi-Damodaran, A. Role of Secondary Coordination Sphere Residues in Halogenation Catalysis of Non-heme Iron Enzymes. *ACS Catal.* **12**, 10913–10924 (2022).
100. Matthews, M. L. *et al.* Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17723–17728 (2009).
101. Senn, H. M. Insights into enzymatic halogenation from computational studies. *Front. Chem.* **2**, (2014).
102. Kastner, D. W., Nandy, A., Mehmood, R. & Kulik, H. J. Mechanistic Insights into Substrate Positioning That Distinguish Non-heme Fe(II)/α-Ketoglutarate-Dependent Halogenases and Hydroxylases. *ACS Catal.* **13**, 2489–2501 (2023).
103. Neugebauer, M. E. *et al.* Reaction pathway engineering converts a radical hydroxylase into a halogenase. *Nat. Chem. Biol. 2021 182* **18**, 171–179 (2021).
104. Zhang, J., Wu, P., Zhang, X. & Wang, B. Coordination Dynamics of Iron is a Key Player in the Catalysis of Non-heme Enzymes. *ChemBioChem* **24**, e202300119 (2023).
105. Münch, J., Püllmann, P., Zhang, W. & Weissenborn, M. J. Enzymatic Hydroxylations of sp^3-Carbons. *ACS Catal.* **11**, 9168–9203 (2021).
106. Mitchell, A. J. & Weng, J. K. Unleashing the synthetic power of plant oxygenases: From mechanism to application. *Plant Physiol.* **179**, 813–829 (2019).
107. Hogani, D. A., Smith, S. R., Saari, E. A., McCracken, J. & Hausinger, R. P. Site-directed mutagenesis of 2,4-dichlorophenoxyacetic acid/α- ketoglutarate dioxygenase. Identification of residues involved in metallocenter formation and substrate binding. *J. Biol. Chem.* **275**, 12400–12409 (2000).
108. Li, J. & Fitzpatrick, P. F. Characterization of Metal Ligand Mutants of Phenylalanine Hydroxylase: Insights into the Plasticity of a 2-Histidine-1-Carboxylate Triad †. *Arch Biochem Biophys* **475**, 164–168 (2008).
109. Hewitson, K. S. *et al.* Evidence that two enzyme-derived histidine ligands are sufficient for iron binding and catalysis by factor inhibiting HIF (FIH). *J. Biol. Chem.* **283**, 25971–25978 (2008).
110. Kim, C. Y. *et al.* The chloroalkaloid (−)-acutumine is biosynthesized via a Fe(II)- and 2-oxoglutarate-dependent halogenase in Menispermaceae plants. *Nat. Commun.* **11**, (2020).
111. Mitchell, A. J. *et al.* Structural basis for halogenation by iron-and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016).

112. Grzyska, P. K., Müller, T. A., Campbell, M. G. & Hausinger, R. P. Metal ligand substitution and evidence for quinone formation in taurine/α-ketoglutarate dioxygenase. *J. Inorg. Biochem.* **101**, 797–808 (2007).
113. Gorres, K. L., Pua, K. H. & Raines, R. T. Stringency of the 2-His-1-Asp active-site motif in prolyl 4-hydroxylase. *PLoS One* **4**, e7635 (2009).
114. Mitchell, A. J. *et al.* Structure-Guided Reprogramming of a Hydroxylase to Halogenate Its Small Molecule Substrate. *Biochemistry* **56**, 441–444 (2017).
115. Papadopoulou, A. *et al.* Re-Programming and Optimisation of a L-Proline cis-4-Hydroxylase for the cis-3-Halogenation of its Native Substrate. *ChemCatChem* **13**, 3914–3919 (2021).
116. Chan, N. H. *et al.* Non-Native Anionic Ligand Binding and Reactivity in Engineered Variants of the Fe(II)- and α-Ketoglutarate-Dependent Oxygenase, SadA. *Inorg. Chem.* **61**, 14477–14485 (2022).
117. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* **7**, 715–727 (2006).
118. Jang, Y. K., Wang, L. & Sancar, G. B. RPH1 and GIS1 Are Damage-Responsive Repressors of PHR1. *Mol. Cell. Biol.* **19**, 7630–7638 (1999).
119. Liu, L. *et al.* Hairless is a histone H3K9 demethylase. *FASEB J.* **28**, 1534 (2014).
120. Pappa, S. *et al.* PHF2 histone demethylase prevents DNA damage and genome instability by controlling cell cycle progression of neural progenitors. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19464–19473 (2019).
121. Horton, J. R., Upadhyay, A. K., Hashimoto, H., Zhang, X. & Cheng, X. Structural basis for human PHF2 Jumonji domain interaction with metal ions. *J. Mol. Biol.* **406**, 1–8 (2011).
122. Alonso-De Vega, I. *et al.* PHF2 regulates homology-directed DNA repair by controlling the resection of DNA double strand breaks. *Nucleic Acids Res.* **48**, 4915–4927 (2020).
123. Gronke, R. S. *et al.* Partial purification and characterization of bovine liver aspartyl β-hydroxylase. *J. Biol. Chem.* **265**, 8558–8565 (1990).
124. Pfeffer, I. *et al.* Aspartate/asparagine-β-hydroxylase crystal structures reveal an unexpected epidermal growth factor-like domain substrate disulfide pattern. *Nat. Commun.* **10**, (2019).
125. Pasini, D. *et al.* JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306–310 (2010).
126. Kasinath, V. *et al.* JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science (80-.).* **371**, (2021).
127. Kasinath, V. *et al.* Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science (80-.).* **359**, 940–944 (2018).
128. Sterling, J., Menezes, S. V., Abbassi, R. H. & Munoz, L. Histone lysine demethylases and their functions in cancer. *Int. J. Cancer* **148**, 2375–2388 (2021).
129. Chaturvedi, S. S., Ramanan, R., Waheed, S. O., Karabencheva-Christova, T. G. & Christov, C. Z. Structure-function relationships in KDM7 histone demethylases. *Adv. Protein Chem. Struct. Biol.* **117**, 113–125 (2019).
130. Wen, H. *et al.* Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation. *J. Biol. Chem.* **285**, 9322–9326 (2010).
131. Baba, A. *et al.* PKA-dependent regulation of the histone lysine demethylase complex PHF2–ARID5B. *Nat. Cell Biol. 2011 136* **13**, 668–675 (2011).
132. Horton, J. R. *et al.* A complete methyl-lysine binding aromatic cage constructed by two domains of PHF2. *J. Biol. Chem.* **299**, 102862 (2023).
133. Yoo, J. *et al.* Advances in histone demethylase KDM3A as a cancer therapeutic target. *Cancers (Basel).* **12**, (2020).
134. Panteleyev, A. A., Paus, R., Ahmad, W., Sundberg, J. P. & Christiano, A. M. Molecular and functional aspects of the hairless (hr) gene in laboratory rodents and humans. *Exp. Dermatol.* **7**, 249–267 (1998).
135. Liu, L. *et al.* Hairless regulates heterochromatin maintenance and muscle stem cell

function as a histone demethylase antagonist. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2025281118 (2021).
136. Pace, N. J. & Weerapana, E. Zinc-Binding Cysteines: Diverse Functions and Structural Motifs. *Biomolecules* **4**, 419 (2014).
137. Fernlund, P. & Stenflo, J. Beta-hydroxyaspartic acid in vitamin K-dependent proteins. *J. Biol. Chem.* **258**, 12509–12512 (1983).
138. McMullen, B. A. *et al.* Complete Amino Acid Sequence of the Light Chain of Human Blood Coagulation Factor X: Evidence for Identification of Residue 63 as β-Hydroxyaspartic Acid. *Biochemistry* **22**, 2875–2884 (1983).
139. Przysiecki, C. T. *et al.* Occurrence of beta-hydroxylated asparagine residues in non-vitamin K-dependent proteins containing epidermal growth factor-like domains. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 7856–7860 (1987).
140. Arlaud, G. J. *et al.* Identification of erythro-β-hydroxyasparagine in the EGF-like domain of human C1r. *FEBS Lett.* **222**, 129–134 (1987).
141. Gronke, R. S. *et al.* Aspartyl-beta-hydroxylase: In vitro hydroxylation of a synthetic peptide based on the structure of the first growth factor-like domain of human factor IX (epidermal growth factor-like domain/beta-hydroxyaspartic acid). *Proc. Nadl. Acad. Sci. USA* **86**, 3609–3613 (1989).
142. Stenflo, J. *et al.* Hydroxylation of aspartic acid in domains homologous to the epidermal growth factor precursor is catalyzed by a 2-oxoglutarate-dependent dioxygenase. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 444–447 (1989).
143. Jia, S. *et al.* A fully active catalytic domain of bovine aspartyl (asparaginyl) β- hydroxylase expressed in Escherichia coli: Characterization and evidence for the identification of an active-site region in vertebrate α-ketoglutarate- dependent dioxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7227–7231 (1994).
144. Feriotto, G. *et al.* Myocyte Enhancer Factor 2 Activates Promoter Sequences of the Human AβH-J-J Locus, Encoding Aspartyl-β-Hydroxylase, Junctin, and Junctate. *Mol. Cell. Biol.* **25**, 3261–3275 (2005).
145. Feriotto, G. *et al.* Multiple levels of control of the expression of the human AβH-J-J locus encoding aspartyl-β-hydroxylase, junctin, and junctate. *Ann. N. Y. Acad. Sci.* **1091**, 184–190 (2006).
146. Finotti, A., Treves, S., Zorzato, F., Gambari, R. & Feriotto, G. Upstream stimulatory factors are involved in the P1 promoter directed transcription of the beta-J-J locus. *BMC Mol. Biol.* **9**, (2008).
147. Xue, T., Su, J., Li, H. & Xue, X. Evaluation of HAAH/humbug quantitative detection in the diagnosis of hepatocellular carcinoma. *Oncol. Rep.* **33**, 329–337 (2015).
148. Huyan, T. *et al.* Optimized expression and purification of humbug in Pichia pastoris and its monoclonal antibody preparation. *Iran. J. Public Health* **44**, 1632–1642 (2015).
149. Dinchuk, J. E. *et al.* Aspartyl β-hydroxylase (AspH) and an evolutionarily conserved isoform of Asph missing the catalytic domain share exons with junctin. *J. Biol. Chem.* **275**, 39543–39554 (2000).
150. Treves, S., Feriotto, G., Moccagatta, L., Gambari, R. & Zorzato, F. Molecular cloning, expression, functional characterization, chromosomal localization, and gene structure of junctate, a novel integral calcium binding protein of sarco(endo)plasmic reticulum membrane. *J. Biol. Chem.* **275**, 39555–39568 (2000).
151. Feriotto, G. *et al.* Multiple Levels of Control of the Expression of the Human AβH-J-J Locus Encoding Aspartyl-β-hydroxylase, Junctin, and Junctate. *Ann. N. Y. Acad. Sci.* **1091**, 184–190 (2006).
152. Hong, C. S., Kwon, S. J. & Kim, D. H. Multiple functions of junctin and junctate, two distinct isoforms of aspartyl beta-hydroxylase. *Biochem. Biophys. Res. Commun.* **362**, 1–4 (2007).
153. Greenhalgh, D. G. The role of growth factors in wound healing. *J. Trauma - Inj. Infect. Crit. Care* **41**, 159–167 (1996).

154. Cohen, S. & Carpenter, G. Human epidermal growth factor: isolation and chemical and biological properties. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 1317–1321 (1975).
155. Zeng, F. & Harris, R. C. Epidermal growth factor, from gene organization to bedside. *Semin. Cell Dev. Biol.* **28**, 2–11 (2014).
156. Wing, R., Wong, C. & Guillaud, L. The role of epidermal growth factor and its receptors in mammalian CNS. *Cytokine Growth Factor Rev.* **15**, 147–156 (2004).
157. Carpenter, G. & Cohen, S. Epidermal growth factor. *J. Biol. Chem.* **265**, 7709–7712 (1990).
158. Carpenter, G. & Zendegui, J. G. Epidermal growth factor, its receptor, and related proteins. *Exp. Cell Res.* **164**, 1–10 (1986).
159. Valcarce, C., Björk, I. & Stenflo, J. The epidermal growth factor precursor. A calcium-binding, β- hydroxyasparagine containing modular protein present on the surface of platelets. *Eur. J. Biochem.* **260**, 200–207 (1999).
160. Parries, G., Chen, K., Misono, K. S. & Cohen, S. The human urinary epidermal growth factor (EGF) precursor: Isolation of a biologically active 160-kilodalton heparin-binding pro-EGF with a truncated carboxyl terminus. *J. Biol. Chem.* **270**, 27954–27960 (1995).
161. Cassel, D. & Glaser, L. Proteolytic cleavage of epidermal growth factor receptor. A Ca^{2+}-dependent, sulfhydryl-sensitive proteolytic system in A431 cells. *J. Biol. Chem.* **257**, 9845–9848 (1982).
162. Artavanis-Tsakonas, S., Matsuno, K. & Fortini, M. E. Notch signaling. *Science (80-.).* **268**, 225–232 (1995).
163. Zhang, H. *et al.* Structure and expression of fibrillin-2, a novel microfibrillar component preferentially located in elastic matrices. *J. Cell Biol.* **124**, 855–863 (1994).
164. Banner, D. W. *et al.* The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. *Nature* **380**, 41–46 (1996).
165. Edgington, T. S., Dickinson, C. D. & Ruf, W. The structural basis of function of the TF·VIIa complex in the cellular initiation of coagulation. *Thromb. Haemost.* **78**, 401–405 (1997).
166. Blasi, F., Vassalli, J.-D. & Dan0, K. Urokinase-Type Plasminogen Activator: Proenzyme, Receptor, and Inhibitors. *J. Cell Biol.* **104**, 801–803 (1987).
167. Sakai, L. Y., Keene, D. R., Renard, M. & De Backer, J. FBN1: The Disease-Causing Gene for Marfan Syndrome and Other Genetic Disorders. *Gene* **591**, 279 (2016).
168. Li, T., Miller, C. H., Payne, A. B. & Hooper, W. C. The CDC Hemophilia B mutation project mutation list: a new online resource. *Mol. Genet. Genomic Med.* **1**, 238 (2013).
169. Stenflo, J., Stenberg, Y. & Muranyi, A. Calcium-binding EGF-like modules in coagulation proteinases: Function of the calcium ion in module interactions. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1477**, 51–63 (2000).
170. Downing, A. K. *et al.* Solution structure of a pair of calcium-binding epidermal growth factor-like domains: Implications for the Marfan syndrome and other genetic disorders. *Cell* **85**, 597–605 (1996).
171. Werner, J. M., Knott, V., Handford, P. A., Campbell, I. D. & Downing, A. K. Backbone dynamics of a cbEGF domain pair in the presence of calcium. *J. Mol. Biol.* **296**, 1065–1078 (2000).
172. Reinhard, D. P. *et al.* Calcium determines the shape of fibrillin. *J. Biol. Chem.* **272**, 7368–7373 (1997).
173. Jensen, S. A., Corbett, A. R., Knott, V., Redfield, C. & Handford, P. A. Ca^{2+}-dependent interface formation in fibrillin-1. *J. Biol. Chem.* **280**, 14076–14084 (2005).
174. Hansson, K. & Stenflo, J. Post-translational modifications in proteins involved in blood coagulation. *J. Thromb. Haemost.* **3**, 2633–2648 (2005).
175. Harris, R. J. & Spellman, M. W. O-Linked fucose and other post-translational modifications unique to EGF modules. *Glycobiology* **3**, 219–224 (1993).
176. Saiki, W., Ma, C., Okajima, T. & Takeuchi, H. Current Views on the Roles of O-Glycosylation in Controlling Notch-Ligand Interactions. *Biomol. 2021, Vol. 11, Page 309* **11**, 309 (2021).
177. Grinnell, B. W., Walls, J. D. & Gerlitz, B. Glycosylation of human protein C affects its

secretion, processing, functional activities, and activation by thrombin. *J. Biol. Chem.* **266**, 9778–9785 (1991).
178. Glanville, R. W., Qian, R. Q., McClure, D. W. & Maslen, C. L. Calcium binding, hydroxylation, and glycosylation of the precursor epidermal growth factor-like domains of fibrillin-1, the Marfan gene protein. *J. Biol. Chem.* **269**, 26630–26634 (1994).
179. Brewitz, L., Onisko, B. C. & Schofield, C. J. Combined proteomic and biochemical analyses redefine the consensus sequence requirement for epidermal growth factor-like domain hydroxylation. *J. Biol. Chem.* **298**, (2022).
180. Sunnerhagen, M. S., Persson, E., Dahlqvist, I., Drakenberg, T. & Stenflo, J. The effect of aspartate hydroxylation on calcium binding to epidermal growth factor-like modules in coagulation factors IX and X. *J. Biol. Chem.* **268**, 23339–23344 (1993).
181. Selander-Sunnerhagen, M. *et al.* How an epidermal growth factor (EGF)-like domain binds calcium. High resolution NMR structure of the calcium form of the NH2-terminal EGF-like domain in coagulation factor X. *J. Biol. Chem.* **267**, 19642–19649 (1992).
182. Nelson, R. M., VanDusen, W. J., Friedman, P. A. & Long, G. L. β-Hydroxyaspartic acid and β-hydroxyasparagine residues in recombinant human protein S are not required for anticoagulant cofactor activity or for binding to C4b-binding protein. *J. Biol. Chem.* **266**, 20586–20589 (1991).
183. Silbermann, E., Moskal, P., Bowling, N., Tong, M. & de la Monte, S. M. Role of aspartyl-(asparaginyl)-β-hydroxylase mediated notch signaling in cerebellar development and function. *Behav. Brain Funct.* **6**, 1–13 (2010).
184. Cantarini, M. C. *et al.* Aspartyl-asparagyl β hydroxylase over-expression in human hepatoma is linked to activation of insulin-like growth factor and Notch signaling mechanisms. *Hepatology* **44**, 446–457 (2006).
185. Zheng, W., Wang, X., Hu, J., Bai, B. & Zhu, H. Diverse molecular functions of aspartate β-hydroxylase in cancer (Review). *Oncol. Rep.* **44**, 2364–2372 (2020).
186. Dinchuk, J. E. *et al.* Absence of post-translational aspartyl β-hydroxylation of epidermal growth factor domains in mice leads to developmental defects and an increased incidence of intestinal neoplasia. *J. Biol. Chem.* **277**, 12970–12977 (2002).
187. Dong, X. *et al.* Aspartate β-Hydroxylase expression promotes a malignant pancreatic cellular phenotype. *Oncotarget* **6**, 1231–1248 (2015).
188. Xue. The distribution and expression profiles of human Aspartyl/Asparaginyl beta-hydroxylase in tumor cell lines and human tissues. *Oncol. Rep.* **24**, (2010).
189. Palumbo, K. S. *et al.* Human aspartyl (Asparaginyl) β-hydroxylase monoclonal antibodies: Potential biomarkers for pancreatic carcinoma. *Pancreas* **25**, 39–44 (2002).
190. Amutha, P. & Rajkumar, T. Role of Insulin-like Growth Factor, Insulin-like Growth Factor Receptors, and Insulin-like Growth Factor-binding Proteins in Ovarian Cancer. *Indian J. Med. Paediatr. Oncol.* **38**, 198–206 (2017).
191. Lin, Q. *et al.* ASPH-notch Axis guided Exosomal delivery of Prometastatic Secretome renders breast Cancer multi-organ metastasis. **18**, 1–17 (2019).
192. Kanwal, M., Smahel, M., Olsen, M., Smahelova, J. & Tachezy, R. Aspartate β-hydroxylase as a target for cancer therapy. *J. Exp. Clin. Cancer Res.* **39**, (2020).
193. Monte, S. M. d. la *et al.* Aspartyl-(asparaginyl)-β-hydroxylase regulates hepatocellular carcinoma invasiveness. *J. Hepatol.* **44**, 971–983 (2006).
194. Wang, K. *et al.* Overexpression of aspartyl-(asparaginyl)-β-hydroxylase in hepatocellular carcinoma is associated with worse surgical outcome. *Hepatology* **52**, 164–173 (2010).
195. Barboro, P. *et al.* The prognostic value of aspartate beta-hydroxylase in early breast cancer. *Int. J. Biol. Markers* **37**, 328–335 (2022).
196. Aihara, A. *et al.* A cell-surface β-hydroxylase is a biomarker and therapeutic target for hepatocellular carcinoma. *Hepatology* **60**, 1302–1313 (2014).
197. Huang, C. K. *et al.* Anti-Tumor Effects of Second Generation β-Hydroxylase Inhibitors on Cholangiocarcinoma Development and Progression. **11**, e0150336 (2016).

198. Ogawa, K. *et al.* Aspartate β-hydroxylase promotes pancreatic ductal adenocarcinoma metastasis through activation of SRC signaling pathway. *J. Hematol. Oncol.* **12**, 1–16 (2019).
199. Huyan, T. *et al.* Development of a novel anti-human aspartyl-(Asparaginyl) β-hydroxylase monoclonal antibody with diagnostic and therapeutic potential. *Oncol. Lett.* **13**, 1539–1546 (2017).
200. Xue, T. T. *et al.* Monoclonal antibodies against human aspartyl (asparaginyl) beta-hydroxylase developed by DNA immunization. **28**, 251–257 (2009).
201. Yang, H. *et al.* Single-chain variable fragment antibody against human aspartyl/asparaginyl beta-hydroxylase expressed in recombinant Escherichia coli. *Hybridoma* **30**, 69–79 (2011).
202. Gundogan, F. *et al.* SiRNA inhibition of aspartyl-asparaginyl β-hydroxylase expression impairs cell motility, Notch signaling, and fetal growth. *Pathol. Res. Pract.* **207**, 545 (2011).
203. Noda, T., Shimoda, M., Ortiz, V., Sirica, A. E. & Wands, J. R. Immunization with aspartate-β-hydroxylase-loaded dendritic cells produces antitumor effects in a rat model of intrahepatic cholangiocarcinoma. *Hepatology* **55**, 86–97 (2012).
204. Nordquist, L. T. *et al.* Phase 1 open-label trial to evaluate the safety and immunogenicity of PAN-301-1, a novel nanoparticle therapeutic vaccine, in patients with biochemically relapsed prostate cancer. *J. Clin. Oncol.* **36**, e15166–e15166 (2018).
205. Iwagami, Y. *et al.* Lambda phage-based vaccine induces antitumor immunity in hepatocellular carcinoma. *Heliyon* **3**, (2017).
206. Shawaf, S., Noureddin, B., Khouri, A. & Traboulsi, E. I. A family with a syndrome of ectopia lentis, spontaneous filtering blebs, and craniofacial dysmorphism. *Ophthalmic Genet.* **16**, 163–169 (1995).
207. Chermakani, P. & Sundaresan, P. Traboulsi syndrome: A rare eye disease and its genetic association. *TNOA J. Ophthalmic Sci. Res.* **61**, 41 (2023).
208. Zahavich, L. *et al. P05.04D Aortic dilatation in patients with Traboulsi syndrome*.
209. Siggs, O. M., Souzeau, E. & Craig, J. E. Loss of ciliary zonule protein hydroxylation and lens stability as a predicted consequence of biallelic ASPH variation. *Ophthalmic Genet.* **40**, 12–16 (2019).
210. Patel, N. *et al.* Mutations in ASPH cause facial dysmorphism, lens dislocation, anterior-segment abnormalities, and spontaneous filtering blebs, or Traboulsi syndrome. *Am. J. Hum. Genet.* **94**, 755–759 (2014).
211. Chandran, P. *et al.* A novel 5 bp homozygous deletion mutation in ASPH gene associates with Traboulsi syndrome. *Ophthalmic Genet.* **40**, 185–187 (2019).
212. Jones, G. *et al.* Traboulsi syndrome caused by mutations in ASPH: An autosomal recessive disorder with overlapping features of Marfan syndrome. *Eur. J. Med. Genet.* **65**, 104572 (2022).
213. Pfeffer, I. Interplay between 2-Oxoglutarate Oxygenases and Cancer: Studies on the Aspartyl/Asparaginyl-beta-Hydroxylase. (University of Oxford, 2014).
214. McGinnis, K. *et al.* Site-directed mutagenesis of residues in a conserved region of bovine aspartyl (asparaginyl) β-hydroxylase: Evidence that histidine 675 has a role in binding Fe^{2+}. *Biochemistry* **35**, 3957–3962 (1996).
215. Korioth, F., Gieffers, C. & Frey, J. Cloning and characterization of the human gene encoding aspartyl β-hydroxylase. *Gene* **150**, 395–399 (1994).
216. Huang, C. K. *et al.* Aspartate beta-hydroxylase promotes cholangiocarcinoma progression by modulating RB1 phosphorylation. *Cancer Lett.* **429**, 1–10 (2018).
217. Zou, Q. *et al.* Hydroxylase Activity of ASPH Promotes Hepatocellular Carcinoma Metastasis Through Epithelial-to-Mesenchymal Transition Pathway. *EBioMedicine* **31**, 287–298 (2018).
218. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
219. Guha, S. *et al.* Implication of a rare deletion at distal 16p11.2 in schizophrenia. *JAMA*

Psychiatry **70**, 253–260 (2013).
220. Sun, S. *et al.* Aspartate beta-hydroxylase domain containing 1 as a prognostic marker associated with immune infiltration in skin cutaneous melanoma. *BMC Cancer* **23**, 292 (2023).
221. Seidel, S. *et al.* A hypoxic niche regulates glioblastoma stem cells through hypoxia inducible factor 2α. *Brain* **133**, 983–995 (2010).
222. Hausinger, R. P. Fe(II)/α-ketoglutarate-dependent hydroxylases and related enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
223. Proshlyakov, D. A., Mccracken, J. & Hausinger, R. P. Spectroscopic analyses of 2-oxoglutarate-dependent oxygenases: TauD as a case study. *J. Biol. Inorg. Chem.* **22**, 367–379 (2017).
224. Myllylä, R., Kuutti-Savolainen, E. R. & Kivirikko, K. I. The role of ascorbate in the prolyl hydroxylase reaction. *Biochem. Biophys. Res. Commun.* **83**, 441–448 (1978).
225. Flashman, E., Davies, S. L., Yeoh, K. K. & Schofield, C. J. Investigating the dependence of the hypoxia-inducible factor hydroxylases (factor inhibiting HIF and prolyl hydroxylase domain 2) on ascorbate and other reducing agents. *Biochem. J.* **427**, 135–142 (2010).
226. Nietfeld, J. J. & Kemp, A. The function of ascorbate with respect to prolyl 4-hydroxylase activity. *BBA - Enzymol.* **657**, 159–167 (1981).
227. Khan, A., Schofield, C. J. & Claridge, T. D. W. Reducing Agent-Mediated Nonenzymatic Conversion of 2-Oxoglutarate to Succinate: Implications for Oxygenase Assays. *ChemBioChem* **21**, 2898–2902 (2020).
228. Brewitz, L., Tumber, A. & Schofield, C. J. Kinetic parameters of human aspartate/asparagine–βhydroxylase suggest that it has a possible function in oxygen sensing. *J. Biol. Chem.* **295**, 7826–7838 (2020).
229. Brewitz, L., Tumber, A., Pfeffer, I., Mcdonough, M. A. & Schofield, J. Aspartate/asparagine-β-hydroxylase: a high-throughput mass spectrometric assay for discovery of small molecule inhibitors. doi:10.1038/s41598-020-65123-9.
230. Mantri, M., Zhang, Z., McDonough, M. A. & Schofield, C. J. Autocatalysed oxidative modifications to 2-oxoglutarate dependent oxygenases. *FEBS J.* **279**, 1563–1575 (2012).
231. Horita, S. *et al.* Structure of the ribosomal oxygenase OGFOD1 provides insights into the regio- and stereoselectivity of prolyl hydroxylases. *Structure* **23**, 639–652 (2015).
232. Tumber, A., Salah, E., Brewitz, L., Corner, T. P. & Schofield, C. J. Kinetic and inhibition studies on human Jumonji-C (JmjC) domain-containing protein 5. *RSC Chem. Biol.* **4**, 399–413 (2023).
233. Dao, J. H. *et al.* Kinetic characterization and identification of a novel inhibitor of hypoxia-inducible factor prolyl hydroxylase 2 using a time-resolved fluorescence resonance energy transfer-based assay technology. *Anal. Biochem.* **384**, 213–223 (2009).
234. Rydzik, A. M. *et al.* Development and Application of a Fluoride-Detection-Based Fluorescence Assay for γ-Butyrobetaine Hydroxylase. *ChemBioChem* **13**, 1559–1563 (2012).
235. Nagaoka, K. *et al.* Targeting Aspartate Beta-Hydroxylase with the Small Molecule Inhibitor MO-I-1182 Suppresses Cholangiocarcinoma Metastasis. *Dig. Dis. Sci.* **66**, 1080–1089 (2021).
236. Brasnett, A. *et al.* Human Oxygenase Variants Employing a Single Protein Fe(II) Ligand Are Catalytically Active. *Angew. Chemie - Int. Ed.* **60**, 14657–14663 (2021).
237. Liebschner, D. *et al.* Polder maps: improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. Sect. D, Struct. Biol.* **73**, 148–157 (2017).
238. Prescott, A. G. *et al.* Oxygen Activation by Nonheme Iron(II) Complexes: r-Keto Carboxylate versus Carboxylate. *J. AM. CHEM. SOC* **17**, (2000).
239. Baldwin, J. E. *et al.* Water exchange of intermediates in a non-haem iron, α-ketoglutarate dioxygenase, deacetoxy-/deacetylcephalosporin C synthase. *J. Chem. Soc. Chem. Commun.* 105–108 (1993).

240. Holme, E., Lindstedt, G., Lindstedt, S. & Tofft, M. ^{18}O studies of the 2-ketoglutarate-dependent sequential oxygenation of thymine to 5-carboxyuracil. *J. Biol. Chem.* **246**, 3314–3319 (1971).
241. Lindblad, B., Lindstedt, G., Tofft, M. & Lindstedt, S. The Mechanism of α-Ketoglutarate Oxidation in Coupled Enzymatic Oxygenations. *J. Am. Chem. Soc.* **91**, 4604–4606 (1969).
242. Kikuchi, Y., Suzuki, Y. & Tamiya, N. The source of oxygen in the reaction catalysed by collagen lysyl hydroxylase. *Biochem. J.* **213**, 507–512 (1983).
243. McNeill, L. A. *et al.* The use of dioxygen by HIF prolyl hydroxylase (PHD1). *Bioorganic Med. Chem. Lett.* **12**, 1547–1550 (2002).
244. Baldwin, J. E., Adlington, R. M., Crouch, N. P. & Pereira, I. A. C. Incorporation of ^{18}O-labelled water into oxygenated products produced by the enzyme deacetoxy/deacetylcephalosporin C synthase. *Tetrahedron* **49**, 7499–7518 (1993).
245. Foster, F. Substrate Analogue Studies on Cephalosporin Biosynthesis Enzymes. (University of Oxford, 2022).
246. Huang, J. *et al.* Selective Chlorination of Substrates by the Halogenase SyrB2 Is Controlled by the Protein According to a Combined Quantum Mechanics/Molecular Mechanics and Molecular Dynamics Study. *ACS Catal.* **6**, 2694–2704 (2016).
247. Matthews, M. L. *et al.* Substrate-triggered formation and remarkable stability of the C-H bond-cleaving chloroferryl intermediate in the aliphatic halogenase, SyrB2. *Biochemistry* **48**, 4331–4343 (2009).
248. Lima, F. L. *et al.* Traboulsi syndrome without features of Marfan syndrome caused by a novel homozygous ASPH variant associated with a heterozygous FBN1 variant. *Ophthalmic Genet.* **44**, 366–370 (2023).
249. Abarca Barriga, H. H. *et al.* A novel ASPH variant extends the phenotype of Shawaf-Traboulsi syndrome. *Am. J. Med. Genet. Part A* **176**, 2494–2500 (2018).
250. Vivante, A. *et al.* Exome sequencing discerns syndromes in patients from consanguineous families with congenital anomalies of the kidneys and urinary tract. *J. Am. Soc. Nephrol.* **28**, 69–75 (2017).
251. Geoghegan, K. F. *et al.* Spontaneous α-N-6-phosphogluconoylation of a 'His tag' in Escherichia coli: The cause of extra mass of 258 or 178 Da in fusion proteins. *Anal. Biochem.* **267**, 169–184 (1999).
252. Greve, J. M., Pinkham, A. M., Thompson, Z. & Cowan, J. A. Active site characterization and activity of the human aspartyl (asparaginyl) β-hydroxylase. *Metallomics* **13**, 56 (2021).
253. Islam, K. The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem. Biol.* **25**, 1171–1184 (2018).
254. Sudhamalla, B. *et al.* Complementary Steric Engineering at the Protein-Ligand Interface for Analogue-Sensitive TET Oxygenases. *J. Am. Chem. Soc.* **140**, 10263–10269 (2018).
255. Breski, M., Dey, D., Obringer, S., Sudhamalla, B. & Islam, K. Engineering Biological C-H Functionalization Leads to Allele-Specific Regulation of Histone Demethylases. *J. Am. Chem. Soc.* **138**, 13505–13508 (2016).
256. Jansen, G. A. *et al.* Human phytanoyl-CoA hydroxylase: Resolution of the gene structure and the molecular basis of Refsum's disease. *Hum. Mol. Genet.* **9**, 1195–1200 (2000).
257. Mukherji, M. *et al.* 'Chemical co-substrate rescue' of phytanoyl-CoA 2-hydroxylase mutants causing refsum's disease. *Chem. Commun.* 972–973 (2001) doi:10.1039/b101039p.
258. Brewitz, L., Nakashima, Y. & Schofield, C. J. Synthesis of 2-oxoglutarate derivatives and their evaluation as cosubstrates and inhibitors of human aspartate/asparagine-β-hydroxylase. *Chem. Sci.* **12**, 1327–1342 (2021).
259. Nakashima, Y., Brewitz, L., Tumber, A., Salah, E. & Schofield, C. J. 2-Oxoglutarate derivatives can selectively enhance or inhibit the activity of human oxygenases. *Nat. Commun.* **12**, (2021).
260. Díaz, R. *et al.* Told through the wine: A liquid chromatography-mass spectrometry

interplatform comparison reveals the influence of the global approach on the final annotated metabolites in non-targeted metabolomics. *J. Chromatogr. A* **1433**, 90–97 (2016).
261. Henderson, R. Max Perutz (1914–2002). *Cell* **109**, 13–16 (2002).
262. Perutz, M. F. *et al.* Structure of Hæmoglobin: A three-dimensional fourier synthesis at 5.5-. resolution, obtained by X-ray analysis. *Nature* **185**, 416–422 (1960).
263. Pietzsch, J. (Nobel P. O. Speed read: Deciphering Life's Enigma Code. *The Nobel Prize* https://www.nobelprize.org/prizes/medicine/1962/speedread.
264. Adams, P. D. *et al.* PHENIX: Building new software for automated crystallographic structure determination. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **58**, 1948–1954 (2002).
265. Ogata, H., Lubitz, W. & Higuchi, Y. Structure and function of [NiFe] hydrogenases. *J. Biochem.* **160**, 251–258 (2016).
266. Shaw Stewart, P. & Mueller-Dieckmann, J. Automation in biological crystallisation. *Acta Crystallogr. Sect. F, Struct. Biol. Commun.* **70**, 686 (2014).
267. Zhu, Y. *et al.* Nanoliter-scale protein crystallisation and screening with a microfluidic droplet robot. *Sci. Rep.* **4**, 1–9 (2014).
268. Grabowski, M. *et al.* Synchrotron radiation as a tool for macromolecular X-Ray Crystallography: A XXI century perspective. *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms* **489**, 30–40 (2021).
269. Vonrhein, C. *et al.* Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 293–302 (2011).
270. Winter, G. Xia2: An expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43**, 186–190 (2010).
271. Winter, G. & McAuley, K. E. Automated data collection for macromolecular crystallography. *Methods* **55**, 81–93 (2011).
272. Emma, P. *et al.* First lasing and operation of an ångstrom-wavelength free-electron laser. *Nat. Photonics* **4**, 641–647 (2010).
273. Orville, A. M. Recent results in time resolved serial femtosecond crystallography at XFELs. *Curr. Opin. Struct. Biol.* **65**, 193–208 (2020).
274. Gisriel, C. *et al.* Membrane protein megahertz crystallography at the European XFEL. *Nat. Commun.* **10**, 1–11 (2019).
275. Hunter, M. S. *et al.* X-ray diffraction from Membrane protein nanocrystals. *Biophys. J.* **100**, 198–206 (2011).
276. Subramaniam, S. & Kleywegt, G. J. A paradigm shift in structural biology. *Nat. Methods 2022 191* **19**, 20–23 (2022).
277. Curry, S. Structural biology: A century-long journey into an unseen world. *Interdiscip. Sci. Rev.* **40**, 308–328 (2015).
278. Schaffer, J. E., Kukshal, V., Miller, J. J., Kitainda, V. & Jez, J. M. Beyond X-rays: An overview of emerging structural biology methods. *Emerg. Top. Life Sci.* **5**, 221–230 (2021).
279. Passmore, L. A. & Russo, C. J. Specimen Preparation for High-Resolution Cryo-EM. in *Methods in Enzymology* vol. 579 51–86 (Europe PMC Funders, 2016).
280. Weissenberger, G., Henderikx, R. J. M. & Peters, P. J. Understanding the invisible hands of sample preparation for cryo-EM. *Nat. Methods* **18**, 463–471 (2021).
281. Grishaev, A. Sample preparation, data collection, and preliminary data analysis in biomolecular solution X-ray scattering. *Curr. Protoc. Protein Sci.* **1**, 17.14 (2012).
282. Keedy, D. A. *et al.* Crystal cryocooling distorts conformational heterogeneity in a model michaelis complex of DHFR. *Structure* **22**, 899–910 (2014).
283. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16247–16252 (2011).
284. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* **357**, 423–424 (1992).
285. Tilton, R. F., Dewan, J. C. & Petsko, G. A. Effects of Temperature on Protein Structure and

Dynamics: X-ray Crystallographic Studies of the Protein Ribonuclease-A at Nine Different Temperatures from 98 to 320 K. *Biochemistry* **31**, 2469–2481 (1992).
286. Eisenmesser, E. Z. *et al.* Intrinsic dynamics of an enzyme underlies catalysis. *Nature* **438**, 117–121 (2005).
287. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science (80-.).* **295**, 1520–1523 (2002).
288. Fraser, J. S. *et al.* Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
289. Skaist Mehlman, T. *et al.* Room-temperature crystallography reveals altered binding of small-molecule fragments to PTP1B. *Elife* **12**, (2023).
290. Wallach, I., Dzamba, M. & Heifets, A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. (2015).
291. Cox, O. B. *et al.* A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**, 2322–2330 (2016).
292. Cerofolini, L. *et al.* Integrative approaches in structural biology: A more complete picture from the combination of individual techniques. *Biomolecules* vol. 9 at https://doi.org/10.3390/biom9080370 (2019).
293. Nitta, R., Imasaki, T. & Nitta, E. Recent progress in structural biology: lessons from our research history. *Microscopy* **67**, 187–195 (2018).
294. Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. **470**, (2011).
295. Schulz, E. C., Yorke, B. A., Pearson, A. R. & Mehrabi, P. Best practices for time-resolved serial synchrotron crystallography. *Acta Crystallogr. Sect. D, Struct. Biol.* **78**, 14–29 (2022).
296. Barends, T. R. M. *et al.* Anomalous signal from S atoms in protein crystallographic data from an X-ray free-electron laser. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 838–842 (2013).
297. Johansson, L. C. *et al.* Lipidic phase membrane protein serial femtosecond crystallography. *Nat. Methods* **9**, 263–265 (2012).
298. Weierstall, U., Spence, J. C. H. & Doak, R. B. Injector for scattering measurements on fully solvated biospecies. *Rev. Sci. Instrum.* **83**, 35108 (2012).
299. Barends, T. R. M. *et al.* De novo protein crystal structure determination from X-ray free-electron laser data. *Nature* **505**, 244–247 (2014).
300. Redecke, L. *et al.* Natively inhibited trypanosoma brucei cathepsin B structure determined by using an X-ray laser. *Science (80-.).* **339**, 227–230 (2013).
301. Boutet, S. *et al.* High-Resolution Protein Structure Determination by Serial Femtosecond Crystallography. *Science (80-.).* **337**, 362–364 (2012).
302. Martiel, I., Müller-Werkmeister, H. M. & Cohen, A. E. Strategies for sample delivery for femtosecond crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **75**, 160–177 (2019).
303. Weierstall, U. Liquid sample delivery techniques for serial femtosecond crystallography. *Philos. Trans. R. Soc. B Biol. Sci.* **369**, (2014).
304. Echelmeier, A. *et al.* 3D printed droplet generation devices for serial femtosecond crystallography enabled by surface coating. *J. Appl. Crystallogr.* **52**, 997–1008 (2019).
305. Hutchison, C. D. M. *et al.* X-ray free electron laser determination of crystal structures of dark and light states of a reversibly photoswitching fluorescent protein at room temperature. *Int. J. Mol. Sci.* **18**, 1918 (2017).
306. Roessler, C. G. *et al.* Acoustic Injectors for Drop-On-Demand Serial Femtosecond Crystallography. *Structure* **24**, 631–640 (2016).
307. Ishigami, I. *et al.* Snapshot of an oxygen intermediate in the catalytic reaction of cytochrome c oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 3572–3577 (2019).
308. Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat. Methods* **14**, 443–449 (2017).
309. Yun, J. H. *et al.* Early-stage dynamics of chloride ion–pumping rhodopsin revealed by a

femtosecond X-ray laser. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
310. Ibrahim, M. *et al.* Untangling the sequence of events during the S2 → S3 transition in photosystem II and implications for the water oxidation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 12624–12635 (2020).
311. Knoška, J. *et al.* Ultracompact 3D microfluidics for time-resolved structural biology. *Nat. Commun.* **11**, 1–12 (2020).
312. Dasgupta, M. *et al.* Mix-and-inject XFEL crystallography reveals gated conformational dynamics during enzyme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25634–25640 (2019).
313. Leissing, T. M. Studies on FIH Interactions. (University of Oxford, 2017).
314. Gati, C. *et al.* Serial crystallography on in vivo grown microcrystals using synchrotron radiation. *IUCrJ* **1**, 87–94 (2014).
315. Stellato, F. *et al.* Room-temperature macromolecular serial crystallography using synchrotron radiation. *IUCrJ* **1**, 204–212 (2014).
316. Coquelle, N. *et al.* Raster-scanning serial protein crystallography using micro- and nano-focused synchrotron beams. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 1184–1196 (2015).
317. Murray, T. D. *et al.* A high-transparency, micro-patternable chip for X-ray diffraction analysis of microcrystals under native growth conditions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **71**, 1987–1997 (2015).
318. Roedig, P. *et al.* Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *urn:issn:1600-5767* **49**, 968–975 (2016).
319. Oghbaey, S. *et al.* Fixed target combined with spectral mapping: Approaching 100% hit rates for serial crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **72**, 944–955 (2016).
320. Owen, R. L. *et al.* Low-dose fixed-target serial synchrotron crystallography. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 373–378 (2017).
321. Mehrabi, P. *et al.* The HARE chip for efficient time-resolved serial synchrotron crystallography. *J. Synchrotron Radiat.* **27**, 360–370 (2020).
322. Sherrell, D. A. *et al.* Fixed-target serial crystallography at the Structural Biology Center. *J. Synchrotron Radiat.* **29**, 1141–1151 (2022).
323. Doak, R. B. *et al.* Crystallography on a chip - Without the chip: Sheet-on-sheet sandwich. *Acta Crystallogr. Sect. D Struct. Biol.* **74**, 1000–1007 (2018).
324. Rabe, P. *et al.* Anaerobic fixed-target serial crystallography. *IUCrJ* **7**, 1–12 (2020).
325. Horrell, S. *et al.* Fixed target serial data collection at diamond light source. *J. Vis. Exp.* **2021**, 62200 (2021).
326. Thorne, R. E. Determining biomolecular structures near room temperature using X-ray crystallography: concepts, methods and future optimisation. *Acta Crystallogr. Sect. D Struct. Biol.* **79**, 78–94 (2022).
327. Weinert, T. *et al.* Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nat. Commun.* **8**, 1–11 (2017).
328. Kamps, J. J. A. G. Biophysical studies on enzymes related to antibiotics. (University of Oxford, 2020).
329. Wranik, M. *et al.* Watching the release of a photopharmacological drug from tubulin using time-resolved serial crystallography. *Nat. Commun.* **14**, 1–12 (2023).
330. Herr, C. Q. & Hausinger, R. P. Amazing Diversity in Biochemical Roles of Fe(II)/2-Oxoglutarate Oxygenases. *Trends Biochem. Sci.* **43**, 517–532 (2018).
331. Mitchell, A. J. *et al.* Structural basis for halogenation by iron-and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016).
332. Miller, S. W., Movsesyan, A., Zhang, S., Fernández, R. & Posakony, J. W. Evolutionary emergence of hairless as a novel component of the notch signaling pathway. *Elife* **8**, e48115 (2019).
333. Di Costanzo, L., Flores, L. V. & Christianson, D. W. Stereochemistry of guanidine–metal

interactions: Implications for L-arginine–metal interactions in protein structure and function. *Proteins Struct. Funct. Bioinforma.* **65**, 637–642 (2006).
334. Cosper, M. M. *et al.* Characterization of the Cofactor Composition of Escherichia coli Biotin Synthase. *Biochemistry* **43**, 2007–2021 (2004).
335. Berkovitch, F., Nicolet, Y., Wan, J. T., Jarrett, J. T. & Drennan, C. L. Crystal Structure of Biotin Synthase, an S-Adenosylmethionine-Dependent Radical Enzyme. *Science (80-.).* **303**, 76–79 (2004).
336. Ferraroni, M. *et al.* Crystal structure of a zinc-activated variant of human carbonic anhydrase I, CA I Michigan 1: Evidence for a second zinc binding site involving arginine coordination. *Biochemistry* **41**, 6237–6244 (2002).
337. Bewley, M. C., Jeffrey, P. D., Patchett, M. L., Kanyo, Z. F. & Baker, E. N. Crystal structures of Bacillus caldovelox arginase in complex with substrate and inhibitors reveal new insights into activation, inhibition and catalysis in the arginase superfamily. *Structure* **7**, 435–448 (1999).
338. Rosano, G. L. & Ceccarelli, E. A. Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain. *Microb. Cell Fact.* **8**, 1–9 (2009).
339. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
340. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta - Proteins Proteomics* **1804**, 996–1010 (2010).
341. Gutiérrez-González, M. *et al.* Optimisation of culture conditions for the expression of three different insoluble proteins in Escherichia coli. *Sci. Rep.* **9**, 1–11 (2019).
342. Bhatwa, A. *et al.* Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Front. Bioeng. Biotechnol.* **9**, 630551 (2021).
343. De Bernardez-Clark, E. & Georgiou, G. Inclusion Bodies and Recovery of Proteins from the Aggregated State. in 1–20 (1991). doi:10.1021/bk-1991-0470.ch001.
344. Lobstein, J. *et al.* SHuffle, a novel Escherichia coli protein expression strain capable of correctly folding disulfide bonded proteins in its cytoplasm. *Microb. Cell Fact.* **11**, 1–16 (2012).
345. Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 14371–14376 (2008).
346. Bernier, S. C., Cantin, L. & Salesse, C. Systematic analysis of the expression, solubility and purification of a passenger protein in fusion with different tags. *Protein Expr. Purif.* **152**, 92–106 (2018).
347. Baynes, B. M., Wang, D. I. C. & Trout, B. L. Role of arginine in the stabilization of proteins against aggregation. *Biochemistry* **44**, 4919–4925 (2005).
348. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L. Y. A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.* **126**, 8933–8939 (2004).
349. Vagenende, V., Yap, M. G. S. & Trout, B. L. Mechanisms of protein stabilization and prevention of protein aggregation by glycerol. *Biochemistry* **48**, 11084–11096 (2009).
350. Pacheco, B., Crombet, L., Loppnau, P. & Cossar, D. A screening strategy for heterologous protein expression in Escherichia coli with the highest return of investment. *Protein Expr. Purif.* **81**, 33–41 (2012).
351. Myllyharju, J. & Kivirikko, K. I. Characterization of the iron- and 2-oxoglutavate binding sites of human prolyl 4-hydroxylase. *EMBO J.* **16**, 1173–1180 (1997).
352. Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **10**, 921–923 (2014).
353. Salgado, E. N., Faraone-Mennella, J. & Tezcan, F. A. Controlling protein-protein interactions through metal coordination: Assembly of a 16-helix bundle protein. *J. Am.*

Chem. Soc. **129**, 13374–13375 (2007).
354. Salgado, E. N., Lewis, R. A., Faraone-Mennella, J. & Tezcan, F. A. Metal-mediated self-assembly of protein superstructures: Influence of secondary interactions on protein oligomerization and aggregation. *J. Am. Chem. Soc.* **130**, 6082–6084 (2008).
355. Salgado, E. N., Lewis, R. A., Mossin, S., Rheingold, A. L. & Tezcan, F. A. Control of protein oligomerization symmetry by metal coordination: C 2 and C 3 symmetrical assemblies through Cu(II) and Ni(II) coordination. *Inorg. Chem.* **48**, 2726–2728 (2009).
356. Brodin, J. D. *et al.* Evolution of metal selectivity in templated protein interfaces. *J. Am. Chem. Soc.* **132**, 8610–8617 (2010).
357. McCoy, A. J. Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 32 (2007).
358. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).

Chapter 9 | Appendix

9.1 Protein sequences

AspH$_{315-758}$
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEK
LRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLKLSLKR
RSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFILK
AQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLK
AQPWWTPKETGYTELVKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENAC
KGAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETKTWEEG
KVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI

AspHD1$_{1-390}$
MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGSMKEGRGSFSVERGPRKERETAQSGMWKGNSPAG
SQGAAMEGTGGELGGQGNWGPEDAPGLLARASLIMLPWPLPLASSALTLLFGALTSLFLWYCYRLGSQDMQ
ALGAGSRAGGVRGGPVGCSEAGGPSPGGPGDPGEGPRTEGLVSRRLRAYARRYSWAGMGRVRRAAQGGP
GPGRGPGVLGIQRPGLLFLPDLPSAPFVPRDAQRHDVELLESSFPAILRDFGAVSWDFSGTTPPPRGWSPPLAP
GCYQLLLYQAGRCQPSNCRRCPGAYRALRGLRSFMSANTFGNAGFSVLLPGARLEGRCGPTNARVRCHLGLKI
PPGCELVVGGEPQCWAEGHCLLVDDSFLHTVAHNGSPEDGPRVVFIVDLWHPNVAGAERQALDFVFAPDP

AspHD1$_{96-390}$
MGSSHHHHHHSLEVLFQGPYRLGSQDMQALGAGSRAGGVRGGPVGCSEAGGPSPGGPGDPGEGPRTEGLV
SRRLRAYARRYSWAGMGRVRRAAQGGPGPGRGPGVLGIQRPGLLFLPDLPSAPFVPRDAQRHDVELLESSFP
AILRDFGAVSWDFSGTTPPPRGWSPPLAPGCYQLLLYQAGRCQPSNCRRCPGAYRALRGLRSFMSANTFGNA
GFSVLLPGARLEGRCGPTNARVRCHLGLKIPPGCELVVGGEPQCWAEGHCLLVDDSFLHTVAHNGSPEDGPRV
VFIVDLWHPNVAGAERQALDFVFAPDP

AspHD1$_{206-390}$
MGSSHHHHHHSLEVLFQGPRHDVELLESSFPAILRDFGAVSWDFSGTTPPPRGWSPPLAPGCYQLLLYQAGRC
QPSNCRRCPGAYRALRGLRSFMSANTFGNAGFSVLLPGARLEGRCGPTNARVRCHLGLKIPPGCELVVGGEPQ
CWAEGHCLLVDDSFLHTVAHNGSPEDGPRVVFIVDLWHPNVAGAERQALDFVFAPDP

His$_6$-SUMO-AspHD1$_{206-390}$
MGSSHHHHHHMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSL
RFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGLEVLFQGPRHDVELLESSFPAILRDFGAVSWDFSGTTPPP
RGWSPPLAPGCYQLLLYQAGRCQPSNCRRCPGAYRALRGLRSFMSANTFGNAGFSVLLPGARLEGRCGPTNA
RVRCHLGLKIPPGCELVVGGEPQCWAEGHCLLVDDSFLHTVAHNGSPEDGPRVVFIVDLWHPNVAGAERQAL
DFVFAPDP

His$_6$-MBP-AspHD1$_{206-390}$
MGSSHHHHHHSGSSMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEKFPQVAATGDGP
DIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIP
ALDKELKAKGKSALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMN
ADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKE
FLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTAVINAA
SGRQTVDEALKDAQTNGILEVLFQGPRHDVELLESSFPAILRDFGAVSWDFSGTTPPPRGWSPPLAPGCYQLLL
YQAGRCQPSNCRRCPGAYRALRGLRSFMSANTFGNAGFSVLLPGARLEGRCGPTNARVRCHLGLKIPPGCELV
VGGEPQCWAEGHCLLVDDSFLHTVAHNGSPEDGPRVVFIVDLWHPNVAGAERQALDFVFAPDP

9.2 Additional data

9.2.1 The pH dependence of $^{18}O_2$ reactions

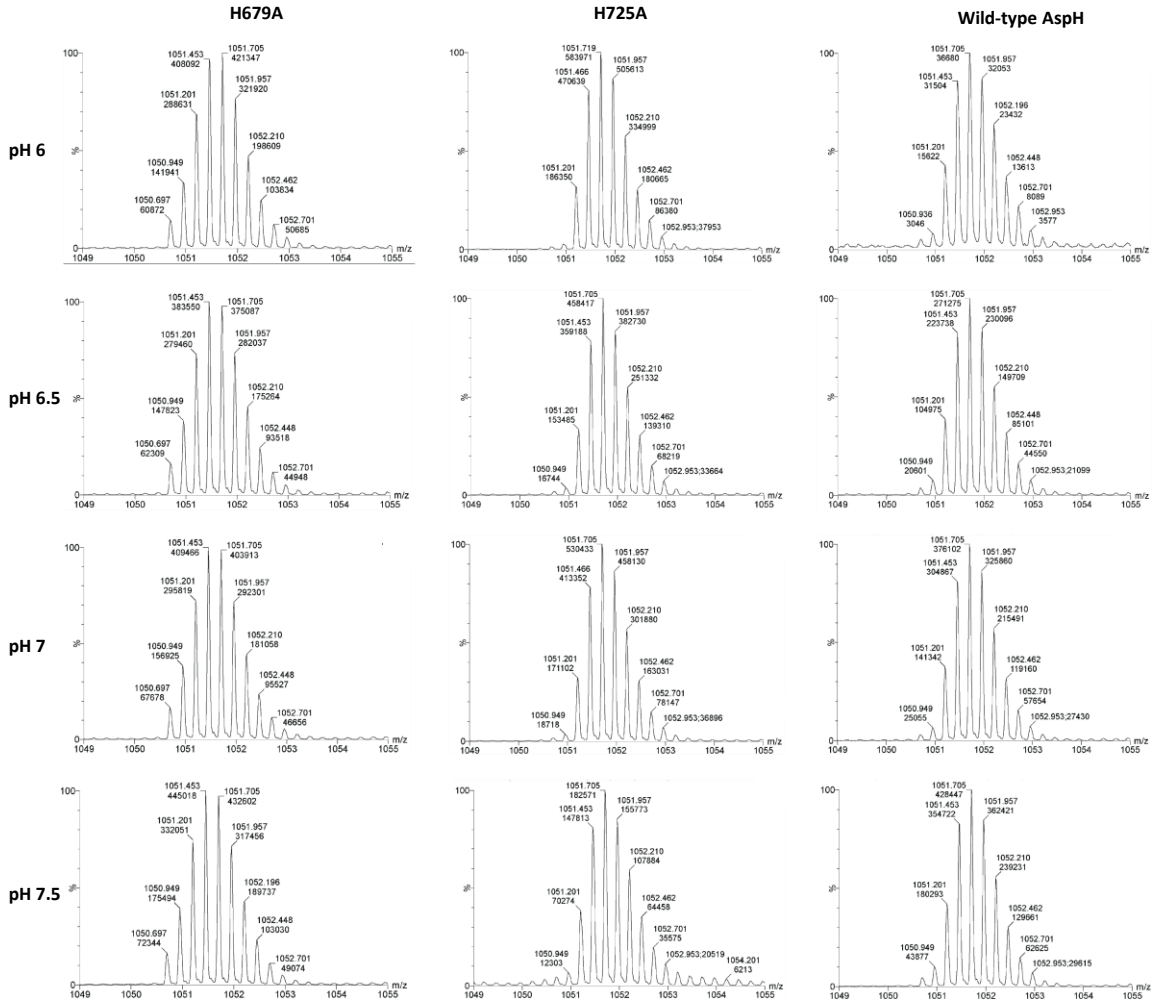


Figure 9-1. LC-MS spectra of the $[M+4H]^{4+}$ fragment of the hydroxylated hFX-Asp_{103} substrate, resulting from hydroxylation reactions catalysed by the H679A and H725A variants, and wt AspH under an $^{18}O_2$ atmopshere at pH 6, 6.5, 7 and 7.5. Data are discussed in the main text in section 2.3.7.1.

9.2.2 Time-course data for Chapter 3

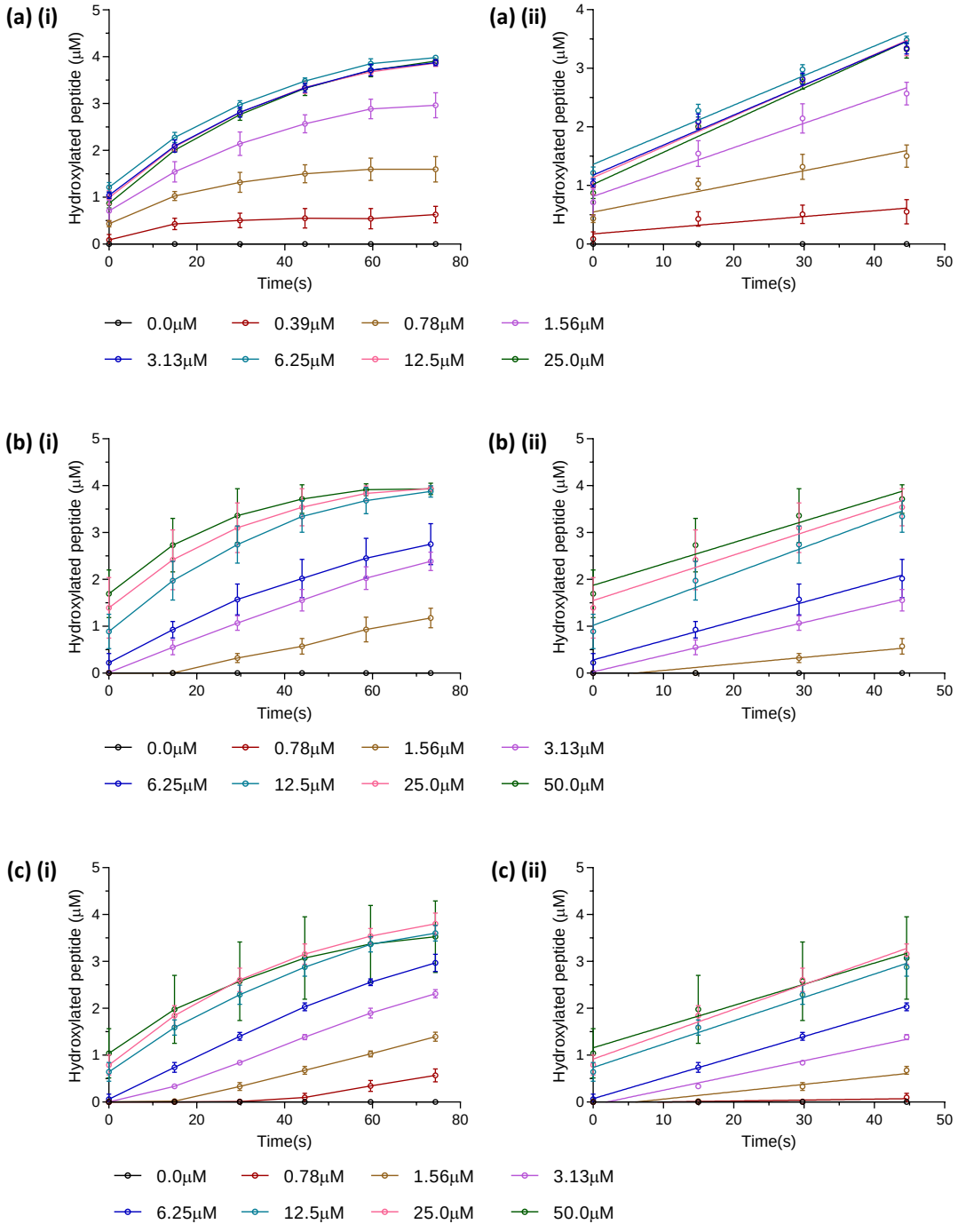


Figure 9-2. Time course data used to calculate Michaelis-Menten kinetics for the G434V AspH variant. (a) (i) Time course data with respect to 2OG concentration and (ii) points used to calculate the initial rates of reaction. (b) (i) Time course data with respect to Fe(II) concentration and (ii) points used to calculate the initial rates of reaction. (c) (i) Time course data with respect to Fe(II) concentration in the presence of LAA and (ii) points used to calculate the initial rates of reaction.

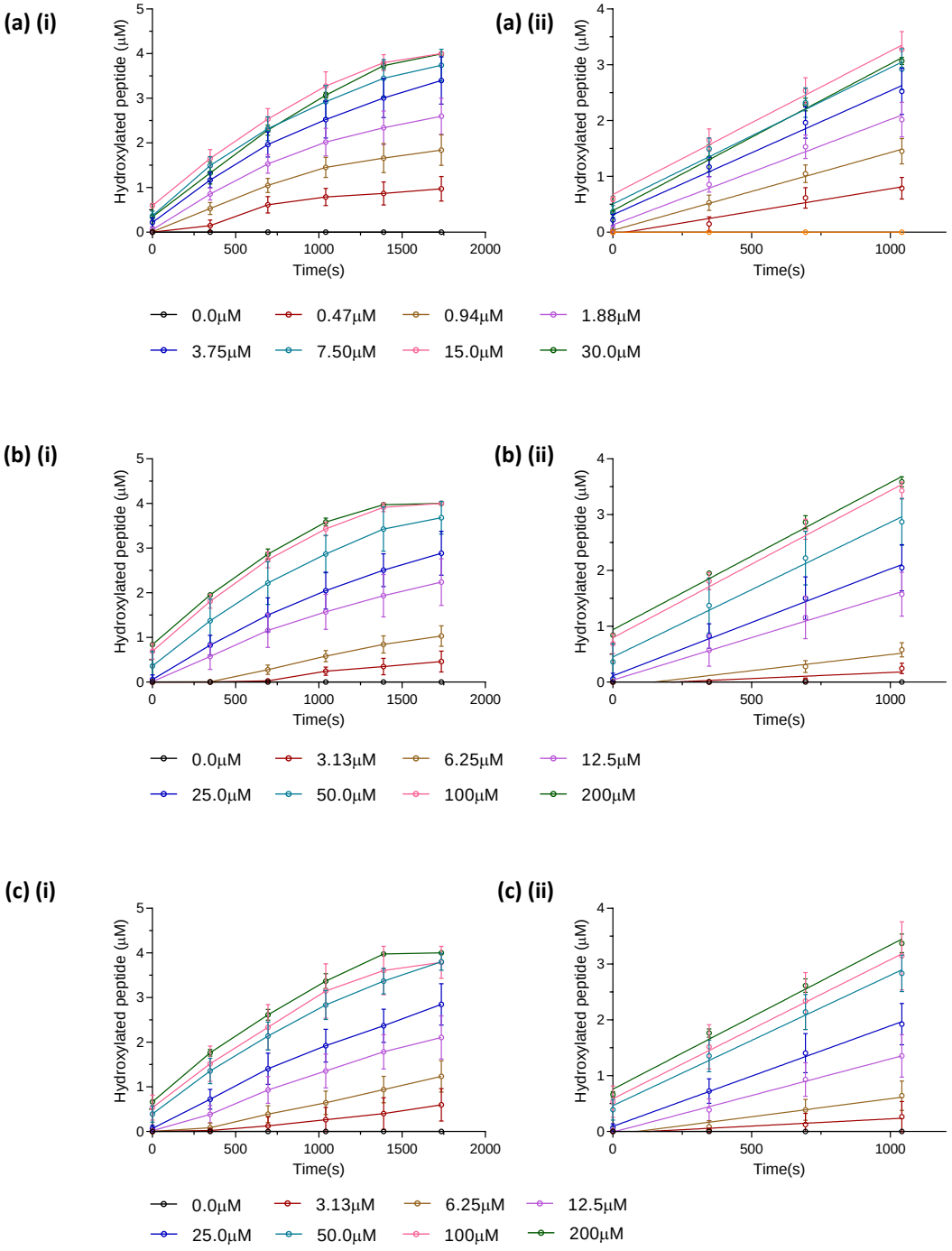


Figure 9-3. Time course data used to calculate Michaelis-Menten kinetics for the R688Q AspH variant. (a) (i) Time course data with respect to 2OG concentration and (ii) points used to calculate the initial rates of reaction. (b) (i) Time course data with respect to Fe(II) concentration and (ii) points used to calculate the initial rates of reaction. (c) (i) Time course data with respect to Fe(II) concentration in the presence of LAA and (ii) points used to calculate the initial rates of reaction.

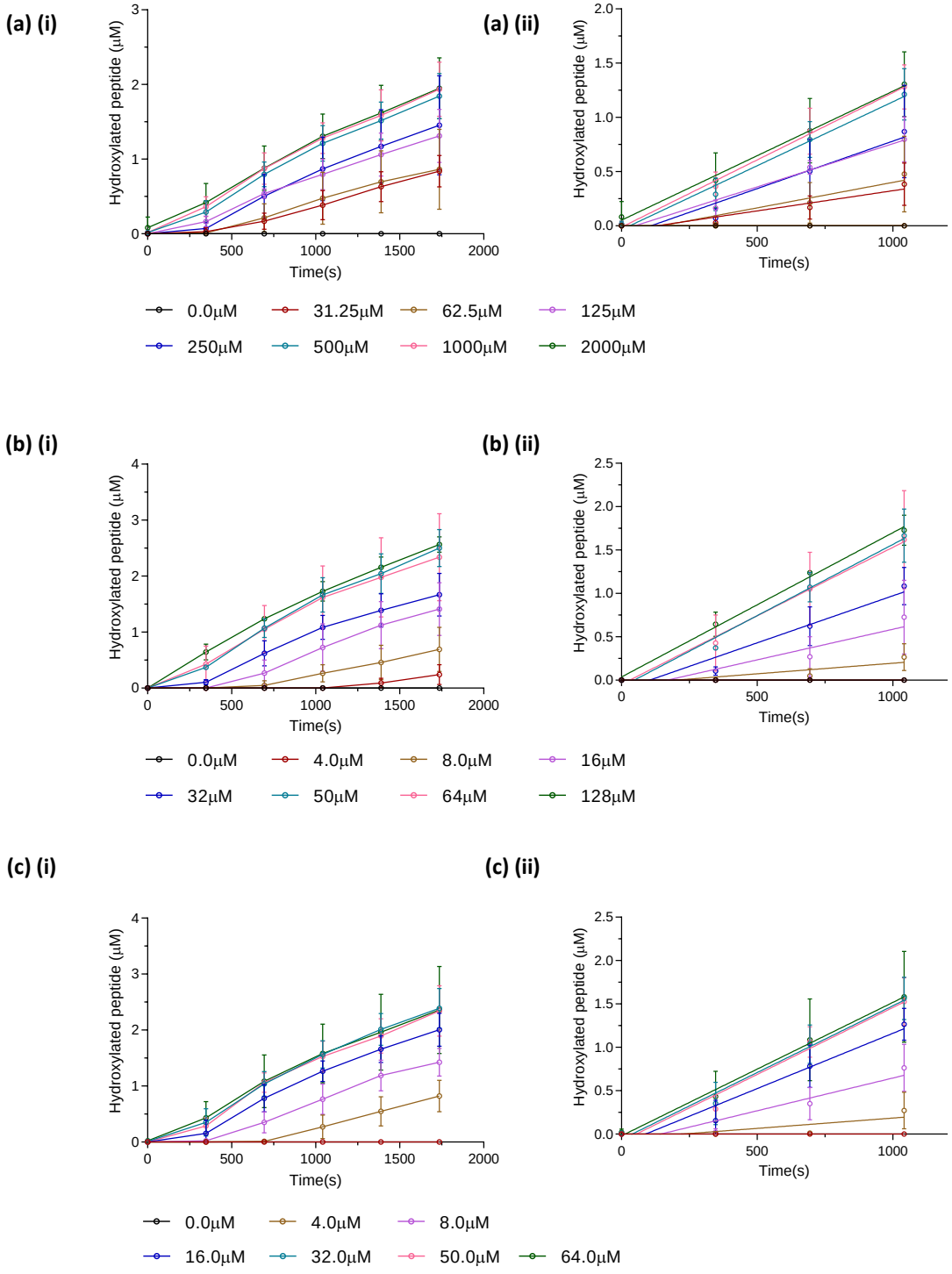


Figure 9-4. Time course data used to calculate Michaelis-Menten kinetics for the R735Q AspH variant. (a) (i) Time course data with respect to 2OG concentration and (ii) points used to calculate the initial rates of reaction. (b) (i) Time course data with respect to Fe(II) concentration and (ii) points used to calculate the initial rates of reaction. (c) (i) Time course data with respect to Fe(II) concentration in the presence of LAA and (ii) points used to calculate the initial rates of reaction.

9.3 ACIE publication

GDCh Research Articles Angewandte International Edition Chemie

Check for updates

Biomimetic Catalysts

How to cite: Angew. Chem. Int. Ed. 2021, 60, 14657–14663
International Edition: doi.org/10.1002/anie.202103711
German Edition: doi.org/10.1002/ange.202103711

Human Oxygenase Variants Employing a Single Protein Fe^{II} Ligand Are Catalytically Active

*Amelia Brasnett[+], Inga Pfeffer[+], Lennart Brewitz[+], Rasheduzzaman Chowdhury, Yu Nakashima, Anthony Tumber, Michael A. McDonough, and Christopher J. Schofield**

Abstract: *Aspartate/asparagine-β-hydroxylase (AspH) is a human 2-oxoglutarate (2OG) and Fe^{II} oxygenase that catalyses C3 hydroxylations of aspartate/asparagine residues of epidermal growth factor-like domains (EGFDs). Unusually, AspH employs two histidine residues to chelate Fe^{II} rather than the typical triad of two histidine and one glutamate/aspartate residue. We report kinetic, inhibition, and crystallographic studies concerning human AspH variants in which either of its Fe^{II} binding histidine residues are substituted for alanine. Both the H725A and, in particular, the H679A AspH variants retain substantial catalytic activity. Crystal structures clearly reveal metal-ligation by only a single protein histidine ligand. The results have implications for the functional assignment of 2OG oxygenases and for the design of non-protein biomimetic catalysts.*

Introduction

2-Oxoglutarate (2OG) and Fe^{II} dependent oxygenase superfamily enzymes are widely distributed and catalyse a range of oxidations, typically hydroxylations, though they catalyse other reactions including hydrogenations, desaturations, fragmentations, and ring formations.[1] In humans, they have important physiological roles, for example in the hypoxic response and epigenetics.[2] Most, but not all, of their reactions comprise two electron substrate oxidations coupled to conversion of 2OG to succinate and CO_2. They have a conserved distorted double-stranded β-helix core fold which supports an active site with a single Fe^{II}. The consensus mechanism for 2OG oxygenases involves an ordered sequential process in which 2OG, then substrate, then O_2 bind at the active site.[3] Structural studies have shown many 2OG oxygenases employ a triad of conserved His, Asp or Glu, His residues (the HXD/E…H motif) to ligate the active site Fe^{II}, leaving three coordination sites for 2OG, then O_2 to bind to the Fe^{II}, the former via its oxalyl group.[1a,4]

Aspartate/asparagine-β-hydroxylase (AspH) is a human 2OG oxygenase that catalyses C3 hydroxylations of Asp/Asn residues in epidermal growth factor like domains (EGFDs; Figure 1a)[5]—AspH is a proposed target for cancer treatment.[6] AspH is highly unusual amongst structurally characterised 2OG dependent hydroxylases in employing only two residues, H679 and H725, to ligate its Fe^{II} (Figure 1b).[7] In this regard, it is related to the 2OG dependent halogenases, where the absence of the Asp/Glu residue of the HXD/E…H motif is used to enable Fe^{II} halide binding.[1b,4] Here we report the unexpected observation that AspH variants with a single Fe^{II} histidine ligand retain substantial catalytic activity. The results have implications for studying the bioinformatical functional assignment of 2OG oxygenases, which have

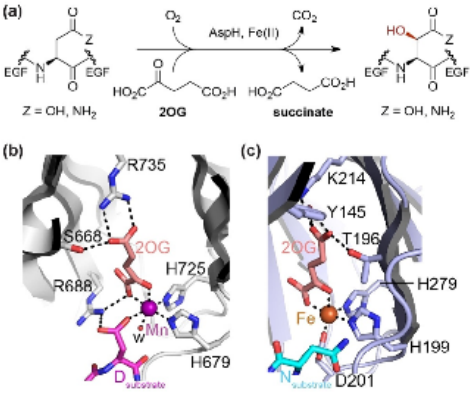


Figure 1. AspH catalyses a typical 2OG oxygenase reaction but has an atypical Fe^{II} binding mode, employing two rather than the typical three protein residues. a) The AspH reaction; b) view of the AspH Fe^{II} binding site (H679 and H725 complex the active site metal ion with Mn substituting for Fe; PDB: 6YYW)[8] compared with that of c) a human 2OG oxygenase with an Fe^{II} binding triad (i.e. H199, H279, and D201; PDB: 1H2L),[9] that is, the asparagine residue hydroxylase, factor inhibiting hypoxia-inducible transcription factor HIF-α (FIH).

[*] A. Brasnett,[+] Dr. I. Pfeffer,[+] Dr. L. Brewitz,[+] Dr. R. Chowdhury, Dr. Y. Nakashima, Dr. A. Tumber, Dr. M. A. McDonough, Prof. Dr. C. J. Schofield
Chemistry Research Laboratory and the Ineos Oxford Institute for Antimicrobial Research, University of Oxford
12 Mansfield Road, Oxford, OX1 3TA (UK)
E-mail: christopher.schofield@chem.ox.ac.uk
Dr. Y. Nakashima
Present address: Institute of Natural Medicine, University of Toyama 2630-Sugitani, 930-0194, Toyama (Japan)
[+] These authors contributed equally to this work.
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202103711.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

15213773, 2021, 26, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202103711 by Oxford University, Wiley Online Library on [18/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

extensively relied on the HXD/E…H triad, and the design of non-protein biomimetic catalysts.

Results and Discussion

Catalytically inactive variants of Fe^{II} ligating residues of the 2OG oxygenases are commonly used in controls during functional assignment work involving cellular studies, including for AspH.[10] As part of our work on AspH, we therefore made the H679A and H725A variants of our previously used human AspH construct (His_6-$AspH_{315-758}$).[7] The requisite mutations were made by standard methods and the variants were purified to near homogeneity by Ni^{II}-affinity and size-exclusion chromatography (>95% pure by SDS-PAGE and MS analysis; Supporting Figure S1). Initially, we conducted turnover assays for the AspH variants alongside wildtype (wt) AspH under reported assay conditions, analysing hydroxylation of a synthetic peptide substrate, hFX-EGFD1$_{86-124}$-4Ser (Supporting Figure S2a),[7] which mimics the requisite disulfide isomer of the EGFD1 of the reported AspH substrate human coagulation factor X (hFX).[11] Unexpectedly, the results revealed that, whilst the variants were less active than wt AspH, they both retained substantial activity: in particular the H679A variant manifested ≈50% substrate turnover under conditions where wt AspH gave >95% turnover (30 minutes incubation); the H725A variant was less active (<5%), but clear evidence for substrate hydroxylation was observed by mass spectrometry (MS) (Supporting Figure S2b). We then carried out more detailed kinetic analyses using our previously reported solid phase extraction coupled to MS (SPE-MS) assay, monitoring substrate depletion/product formation (+16 Da),[12] under optimized buffer conditions (Supporting Figure S3).

Apparent Michaelis constants (K_m^{app}) and turnover numbers (k_{cat}^{app}) of wt AspH for Fe^{II} and 2OG are in good agreement with reported data, which were determined using a different synthetic cyclic peptide substrate.[12] The k_{cat}^{app}-values for the H679A AspH variant are marginally lower than those of wt AspH (Table 1), assuming approximately full enzyme activity as suggested by active site titration of wt AspH.[12] By contrast, the k_{cat}^{app}-values for the H725A AspH variant are approximately half those for H679A AspH (Table 1), in agreement with the observed lower substrate hydroxylation in our initial turnover assays (Supporting Figure S2b).

Interestingly, the H679A AspH K_m^{app}-values for Fe^{II}, either in the presence or absence of L-ascorbate (LAA), and the hFX-EGFD1$_{86-124}$-4Ser substrate peptide are, within experimental error, in the range of those of wt AspH (Table 1) and those reported for other 2OG oxygenases.[13] As for wt AspH,[12] the presence of LAA reduces the K_m^{app}-value for Fe^{II}, but LAA is not a requirement for productive catalysis by the AspH variants (Supporting Figure S7). Notably, the H679A and wt AspH K_m^{app}-values for 2OG differ substantially, the former is about two orders of magnitude higher than the latter (Table 1, entry 3). The slightly different assay conditions (0.2 μM H679A AspH in the absence of NaCl vs. 0.1 μM wt AspH and 50 mM NaCl in the reaction buffer; Supporting Figures S4–S6) likely do not account for this substantial discrepancy. Though higher than the 2OG K_m^{app}-values for wt AspH and other human 2OG oxygenases,[14] the 2OG K_m^{app} concentration for H679A is in the range of 2OG concentrations reported in cells (up to >1 mM[15]) and lower than the reported 2OG K_m^{app}-value for γ-butyrobetaine hydroxylase (≈150–470 μM).[16]

Despite the difference in the K_m^{app}-values for 2OG, the results with H679A AspH imply that the presence of only a single polar Fe^{II}-coordinating residue does not preclude efficient hydroxylation—at least in the case of this variant under conditions when 2OG is not limiting. They also reveal the importance of the Fe^{II} chelating histidines in productive 2OG binding.

The H725A AspH substitution appears to have a more pronounced effect than the H679A substitution on the K_m^{app}-values for Fe^{II}, which are approximately twofold and threefold higher than those of wt and H679A AspH, respectively (Table 1, entries 1 and 2). Strikingly, for H725A AspH, the K_m^{app}-value for 2OG is ≈200 times greater than that of wt AspH and about double that of H679A AspH (Table 1, entry 3). By contrast, the H725A AspH K_m-value for the hFX-EGFD1$_{86-124}$-4Ser substrate peptide is in the range of that of H679A AspH (Table 1, entry 4).

The k_{cat}/K_m-values (specificity constants) highlight the large discrepancies between wt AspH and the two variants with respect to 2OG (Table 1). The H679A and H725A AspH

Table 1: Steady-state kinetic parameters for the H679A and H725A AspH variants. Apparent turnover numbers (k_{cat}^{app}), apparent Michaelis constants (K_m^{app}), and specificity constants (k_{cat}/K_m) of wt, H679A, and H725A His_6-$AspH_{315-758}$ determined for Fe^{II}, 2OG, and the hFX-EGFD1$_{86-124}$-4Ser substrate peptide.[a,b]

Entry		AspH variant	k_{cat}^{app} [s^{-1}]	K_m^{app} [μM]	k_{cat}/K_m [$mM^{-1}s^{-1}$]
1	Fe^{II}[c]	wt	0.24±0.04	7.2±2.6	33.3±13.2
		H679A	0.18±0.01	4.4±0.8	40.9±7.8
		H725A	0.09±0.01	12.3±3.0	7.3±2.0
2	Fe^{II}[d]	wt	0.27±0.04	4.3±1.6	62.8±25.2
		H679A	0.19±0.01	2.7±0.5	70.4±13.6
		H725A	0.10±0.01	8.1±1.8	12.3±3.0
3	2OG	wt	0.31±0.03	1.1±0.4	282±106
		H679A	0.18±0.01	109±20	1.7±0.3
		H725A	0.08±0.01	211±65	0.38±0.13
4	hFX-EGFD1$_{86-124}$-4Ser[e]	wt[f]	n.d.	n.d.	n.d.
		H679A	0.18±0.01	1.5±0.3	120±24.9
		H725A	0.10±0.02	1.7±1.1	58.8±39.8

[a] Determined using 0.2 μM AspH variant or 0.1 μM wt AspH by SPE-MS. Detailed analyses are given in Supporting Figures S4–S6. [b] Mean of three independent runs ($n=3$; mean ± standard deviation, SD). [c] Without L-ascorbate (LAA). [d] With LAA. [e] K_m and k_{cat} values were determined monitoring the hydroxylation of the hFX-EGFD1$_{86-124}$-4Ser substrate peptide (Supporting Figure S2a). [f] Apparent substrate inhibition combined with low detectability of the peptide at low concentrations prevented determination.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2021, 26, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202103711 by Oxford University, Wiley Online Library on [18/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

k_{cat}/K_m-values for 2OG are substantially smaller than that of wt AspH, whereas the k_{cat}/K_m-values for Fe^{II} are, within error, similar. NMR turnover assays in the absence of substrate show that H679A and H725A AspH-catalysed oxidative decarboxylation of 2OG is highly coupled to substrate oxidation, demonstrating that the discrepancy is not due to substrate-uncoupled 2OG oxidation (Supporting Figure S8).

We further investigated the effects of the His to Ala substitutions on the interactions of Fe^{II} and 2OG with AspH by determining half-maximum inhibitory concentrations (IC_{50}-values) of reported 2OG oxygenase inhibitors comprising metal ions and 2OG analogues (Supporting Information Section 2). Mn, Co, Ni, and Zn ions are reported wt AspH inhibitors[17] and also inhibit both AspH variants (Table 2, entries 1–4). In agreement with results for wt AspH, most efficient AspH variant inhibition by the investigated metal ions was observed for Zn^{II}, with Mn^{II} being the least potent. The reduced potency of the metal ions for AspH variant over wt AspH inhibition in part reflects the higher Fe^{II} concentration in the AspH variant inhibition assays (4 μM for H679A AspH and 10 μM for H725A AspH vs. 2 μM for wt AspH; Supporting Information Section 2). By contrast and consistent with the kinetic analyses, the inhibition results for the 2OG analogues reveal clear differences for the variants compared to wt AspH. The 2OG-competitive broad-spectrum 2OG oxygenase inhibitor *N*-oxalylglycine (NOG) potently inhibits wt AspH,[7,17] but does not apparently inhibit either AspH variant (Table 2, entry 5). This discrepancy possibly in part reflects the higher 2OG concentration used with the AspH variant, compared to the wt AspH inhibition assays (110 μM for H679A AspH and 215 μM for H725A AspH vs. 3 μM for wt AspH; Supporting Information Section 2). Another broad-spectrum 2OG oxygenase inhibitor, pyridine-2,4-dicarboxylic acid (2,4-PDCA), inhibits the AspH variants, albeit less efficiently compared to wt AspH (Table 2, entry 6). The potency of 2,4-PDCA appears to decrease with increasing 2OG assay concentration; however, for wt AspH, it has been reported that 2,4-PDCA maintains its potency at high 2OG concentration ($IC_{50} \approx 0.1$ μM at 200 μM 2OG vs. $IC_{50} \approx 0.03$ μM at 3 μM 2OG),[18] suggesting that varying the AspH Fe^{II} binding ligands has a pronounced effect on the stability of the AspH:Fe^{II}:2OG/2OG analogue complex.

Table 2: Inhibition of AspH variants by 2OG oxygenase inhibitors.

Entry	Inhibitor	AspH variant	IC_{50} [μM][a]
1	$MnCl_2$	wt[17]	5.1 ± 1.1
		H679A	23.1 ± 8.6
		H725A	> 40
2	$CoCl_2$	wt[17]	0.4 ± 0.2
		H679A	1.5 ± 0.2
		H725A	3.0 ± 0.5
3	$NiSO_4$	wt[17]	0.17 ± 0.08
		H679A	2.7 ± 0.4
		H725A	1.8 ± 0.4
4	$Zn(OAc)_2$	wt[17]	0.05 ± 0.01
		H679A	0.16 ± 0.01
		H725A	0.16 ± 0.01
5	*N*-oxalyl-glycine (NOG)	wt[17]	1.1 ± 0.3
		H679A	inactive
		H725A	inactive
6	Pyridine-2,4-dicarboxylic acid (2,4-PDCA)	wt[17]	0.03 ± 0.01
		H679A	0.13 ± 0.01
		H725A	11.2 ± 1.4

[a] Mean of two independent runs ($n = 2$; mean ± SD). H679A/H725A AspH inhibition assays were performed as described in the Supporting Information Section 2 using 0.1 μM AspH variant and 4.0 μM hFX-$EGFD1_{86\text{-}124}$-4Ser (Supporting Figure S2a); the assays were of good quality as high S/N ratios and Z′-factors[19] (> 0.5 for each plate) manifest (Supporting Figure S9).

To investigate the structural basis of the kinetic differences between wt AspH and its variants, the latter were crystallised in the presence of Mn^{II} and 2OG or NOG, with or without the synthetic hFX-$EGFD1_{86\text{-}124}$-4Ser substrate peptide. The structures were solved by molecular replacement using reported structures (PDB ID: 5JTC[17] or 5JQY[7]) as search models. The AspH variants crystallised in one of two forms in the $P2_12_12_1$ space group (1.6–2.7 Å resolution; Supporting Figures S10–S20 and Supporting Table S1), in accord with reported wt AspH structures.[7,8,17]

The crystallisation conditions were found to determine the nature of the AspH variant structures. When H679A AspH was crystallised with Mn^{II} and NOG, but without the hFX-$EGFD1_{86\text{-}124}$-4Ser substrate, a metal ion complex was not obtained. Comparison with a wt AspH structure reveals that the substitution of H679 by alanine neither substantially changes the overall AspH fold nor the conformation of key active site residues engaged in Fe^{II}/2OG binding in wt AspH (Supporting Figure S11). Instead of the anticipated NOG, an acetate from the buffer was present at the active site, with its carboxylate positioned to interact with R735 (2.6 and 3.0 Å) and S668 (2.4 Å)—these residues normally interact with the 2OG C5 carboxylate (Figure 1b). Several active site water molecules, but no metal ions, were observed, suggesting impaired metal ion binding capacity in the absence of the H679 imidazole (Figure 2a).

To obtain a metal-bound H679A AspH structure, we added the hFX-$EGFD1_{86\text{-}124}$-4Ser substrate to the Mn^{II}/NOG containing crystallisation mixture aiming to stabilise metal ion binding, potentially by additional metal ion coordination with the substrate aspartyl side chain carboxylate, as observed in some cases for wt AspH:Mn:ligand:hFX-$EGFD1_{86\text{-}124}$-4Ser structures,[7,8,17] or by sterically hindering the release of the metal ion and/or 2OG/NOG. Clear electron density for $EGFD1_{86\text{-}124}$-4Ser was observed in the resultant H679A AspH structure, with the AspH and substrate conformations being similar to those reported (Supporting Figures S14–S20); however, neither a Mn ion nor NOG were bound in the active site (Figure 2b). Instead, as observed before with acetate, a formate ion from the buffer was positioned to interact with R735 (2.7 and 2.9 Å) and S668 (2.5 Å) (Figure 2b). Similar results were obtained when H725A AspH was crystallised in the presence of Mn^{II}, 2OG, and hFX-$EGFD1_{86\text{-}124}$-Ser, with a propionate ion being positioned to interact with R735 and S668, though the propionate was observed in two conformations (Figure 2c).

These crystallographic observations correlate with the increased 2OG K_m-values of the AspH variants compared to

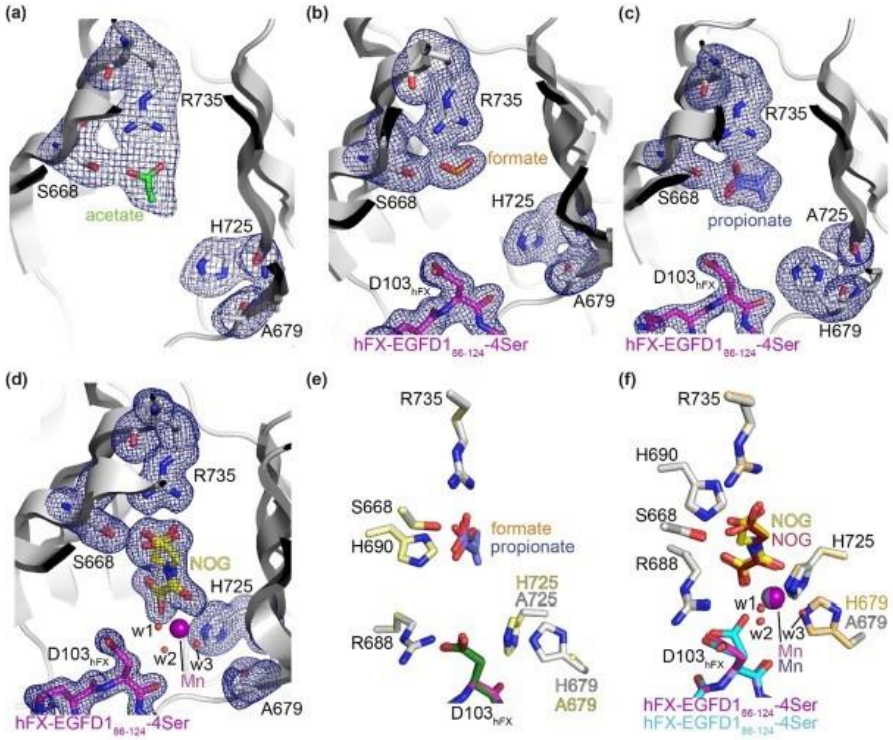


Figure 2. Crystal structure views of the H679A and H725A AspH variants. Colors: grey: His_6-$AspH_{115-258}$ variants; yellow: carbon-backbone of *N*-oxalylglycine (NOG); green: carbon-backbone of acetate; orange: carbon-backbone of formate; slate blue: carbon-backbone of propionate; violet: Mn; magenta: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen. w: water. a–d) Representative OMIT electron density maps (mF_o-DF_c) contoured to 3.0σ around key residues and the a) acetate in the H679A AspH:acetate structure, b) formate in the H679A AspH:hFX-$EGFD1_{86-124}$-4Ser structure, c) propionate in the H725A AspH:hFX-$EGFD1_{86-124}$-4Ser structure (note: propionate adopts two conformations and the orientation of its carboxylate differs from the orientations of acetate and formate in a,b), and d) NOG in the H679A AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure; e) superimposition of views of the H679A AspH:hFX-$EGFD1_{86-124}$-4Ser and H725A AspH:hFX-$EGFD1_{86-124}$-4Ser structures (colours: pale yellow: H725A AspH; dark green: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide); f) superimposition of views of the H679A AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure and the reported wt AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure (colours: bronze: wt AspH; lavender blue: Mn; maroon: carbon-backbone of NOG; cyan: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide; PDB ID: 5JQY).[7]

wt AspH (Table 1) and lack of inhibition by NOG (Table 2, entry 5); thus, under conditions where wt AspH crystallises with metal and 2OG/NOG (and, in some cases, hFX-$EGFD1_{86-124}$-Ser),[7,8,17] the latter are outcompeted for binding by abundant carboxylic acids in the crystallisation buffer. The crystallographic studies also unambiguously reveal the presence of only one potential Fe^{II}-binding histidine residue for the H679A and H725A AspH variants (Figure 2a–c).

We worked to obtain crystallisation conditions in the absence of carboxylic acids in the crystallisation buffer and obtained a H679A AspH structure in complex with Mn^{II}, NOG, and hFX-$EGFD1_{86-124}$-4Ser (Supporting Figure S18). The structure revealed clear density for the side chains of H725 and A679 (Figure 2d). Superimposition of views of analogous H679A and wt AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structures (PDB ID: 5JQY)[7] reveals no substantial alterations in the conformations of H725 or of residues directly engaged in NOG binding (i.e. S668, R688, H690, and R735), or of the Mn ion position (Figure 2f).

Electron density was observed for a Mn ion, which is ligated by the NOG oxalyl group (2.1 and 2.3 Å) and three water molecules (w1: 2.2 Å, w2: 2.1 Å, and w3: 2.2 Å; Figure 2d). As for wt AspH, the NOG C5 carboxylate group is positioned to interact with the side chains of S668 (2.5/2.9 Å) and R735 (2.2 and 2.7/3.1 Å). In addition to ligating the metal, the NOG oxalyl carboxylate is positioned to interact with the side chains of R688 (2.8 and 3.1 Å) and H690 (2.8 Å), as observed in wt AspH:Mn:NOG/2OG structures;[7,8] how-

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2021, 26, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202103711 by Oxford University, Wiley Online Library on [18/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ever, two conformations for the NOG glycine unit are observed, one of which corresponds to the single conformation of NOG in the reported wt AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure[7] (Figure 2 f). In reported wt AspH:Mn:2OG structures, 2OG is present as a single conformer which occupies the same conformation as NOG in wt AspH structures.[8] Note that wt AspH:Mn:2OG and wt AspH:Mn:2OG:substrate structures reveal that substituting NOG for 2OG has no substantial effects on the overall fold of wt AspH.[8] Thus, the H679A substitution appears to disfavour NOG binding as evidenced by the two conformations observed with H679A AspH compared to the one with wt AspH, consistent with the lack of potent inhibition of the variants compared to wt AspH (Table 2, entry 5).

In the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser complex, the Mn ion is clearly ligated by only one protein-bound ligand, that is, H725 with the *N*τ-imidazole-Mn distance being 2.2 Å, the same as in wt AspH, and the A679 methyl carbon-Mn distance being 5.8 Å (compared to 2.3 Å for the wt AspH H679 *N*τ-imidazole-Mn distance; Figure 2 d and 3).[7] Water w2 occupies the same coordination site as the single coordinating water in the analogous wt AspH structure (Figure 3). Strikingly, water w3 occupies the coordination site of the *N*τ-imidazole nitrogen of H679 in the wt AspH structure (Mn-w3 distance: 2.2 Å; Figure 3), apparently compensating for the H679 to alanine mutation. Water w1 coordinates to Mn at a position *trans* to the *N*τ-imidazole nitrogen of H725, that is, at the position where O_2 is predicted to bind,[8] in a manner which does not change the relative alignment of the substrate backbone with respect to the reported wt AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure. A single conformation for the substrate D103$_{hFX}$ side chain is observed in the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure, apparently positioning its C3 methylene for stereoselective oxidation during catalysis. However, two conformations are observed for the D103$_{hFX}$ side chain carboxylate in the wt AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure as well as in other AspH:Mn:ligand:substrate structures, including those complexed with 2OG (Figure 3).[7,8] In one conformation, one of the D103$_{hFX}$ carboxylate oxygen atoms is positioned to coordinate the Mn ion (2.9 Å), that is, it occupies a position close to the position where water w1 complexes the Mn ion in the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure. In the wt AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure, the other D103$_{hFX}$ side chain carboxylate conformer adopts a similar conformation as exclusively observed for the D103$_{hFX}$ side chain carboxylate in the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure.

Conclusion

The combined solution studies presented here, clearly demonstrate that the H679A and H725A AspH variants retain substantial catalytic activity compared to wt AspH (Table 1 and Supporting Figure S2). The crystallographic analyses provide clear evidence for the presence of a single protein (histidine) ligand in each variant. Previous mutation studies with bovine AspH reported no activity for the H675A bovine AspH variant (corresponding to human H679A AspH by sequence alignment[7]), with reduced activity for the H675D/E variants, which can coordinate metals via their side chains.[20]

Surprisingly, the kinetic analyses revealed comparatively small effects of the H679A and H725A mutations on the

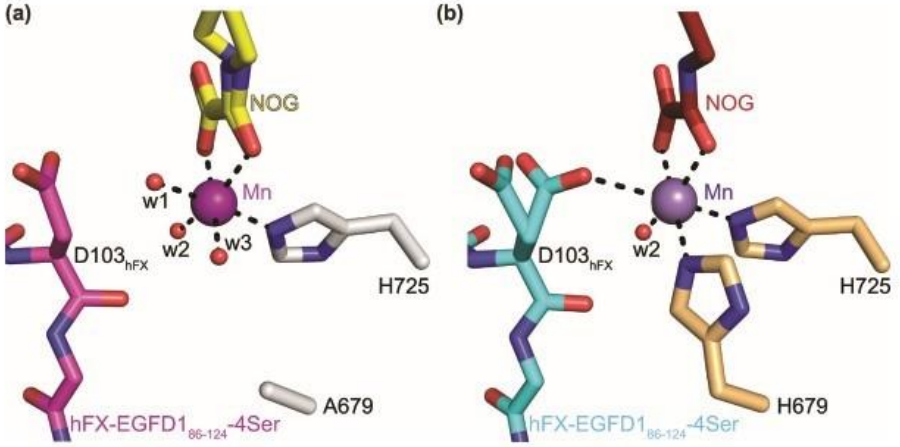


Figure 3. A single protein-bound ligand (i.e. H725) supports the metal ion in the H679A AspH variant. Colors: red: oxygen; blue: nitrogen. w: water. a) Active site view of the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure (colors: grey: H679A AspH; violet: Mn; yellow: carbon-backbone of NOG; magenta: carbon-backbone of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide); b) active site view of the reported wt AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure (colours: bronze: wt AspH; lavender blue: Mn; maroon: carbon-backbone of NOG; cyan: carbon-backbone of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide; PDB ID: 5JQY).[7]

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2021, 26, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202103711 by Oxford University, Wiley Online Library on [18/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

15213773, 2021, 26, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202103711 by Oxford University, Wiley Online Library on [18/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

K_m^{app}-value for Fe^{II}, that is, the values for wt and H679A AspH are similar and approximately three times lower than the value for H725A AspH (Table 1). The involvement of other (second sphere) active site residues in Fe^{II} binding is likely, but there is no evidence for direct Fe^{II} chelation by other residues, though this and/or the involvement of the substrate in Fe^{II} binding cannot be entirely ruled out during catalysis. By contrast, the effects of the histidine substitutions on the K_m^{app}-value for 2OG are much more substantial, that is, the values for H679A AspH (≈109 μM) and H725A AspH (≈211 μM) are more than two orders of magnitude higher than the value for wt AspH (≈1.1 μM; Table 1, entry 3). In wt AspH, the H679 imidazole coordinates Fe^{II} *trans* to the 2OG C1 carboxylate (Figure 1b and Supporting Figure S21).[8] The replacement of this interaction by a water (or hydroxide) in H679A AspH may alter backdonation of electron density from the Fe ion to the ligand during 2OG binding or in intermediates during catalysis, resulting in altered kinetic parameters, including an increased H679A AspH K_m^{app}-value for 2OG. The H725 imidazole that coordinates Fe^{II} opposite to the proposed O_2-coordination site (Supporting Figure S21) may similarly regulate the stability of O_2 (and 2OG) derived intermediates, including the Fe^{III}-2OG-superoxide species in catalysis.

The observation of substantial catalytic activity for a 2OG oxygenase with a single protein-based Fe^{II} ligand has implications for functional assignment work on the superfamily. The results show care should be taken in the use of Fe^{II} ligand variants in negative controls in cellular and in vivo work. Studies with isolated enzyme/substrate(s) should be carried out (note that the K_m^{app}-values for the AspH variants investigated are in the range of cellular 2OG concentrations[15]). Work with FIH has shown that substitution of the Asp-residue of the HXD…H motif for Gly, but not Ala or Glu, enables retention of some activity (note that AspH also has an HXG…H motif), further supporting the need for experimental determination of a lack of activity.[21] Secondly, bioinformatics searches for 2OG oxygenases commonly look for DSBH fold proteins with the consensus HXD/E…H motif (or an HXA/G…H motif in the case of 2OG dependent halogenases[4]). The knowledge that they can operate with two histidines as in the case of AspH or only one histidine ligand, substantially expands the set of potential 2OG oxygenases. Indeed, a number of potential 2OG oxygenases have atypical Fe^{II} binding residues, including, in humans, PHD finger protein 2 (PHF2)[22] and hairless;[23] of particular interest is the naturally occurring human AspH variant of unknown function, aspartate β-hydroxylase domain-containing protein 1 (AspHD1), which sequence analysis predict has only the distal His of the typical HXD/E…H motif as Fe^{II} ligand, the proximal His (i.e. corresponding to H679 in wt AspH) is substituted for an Arg residue (Supporting Figure S22). These enzymes are the subject of ongoing studies.

Many attempts to make non-protein biomimetic oxidation catalysts of non-heme Fe^{II} oxygenases have employed 3 or more Lewis basic metal ion ligands.[24] The results presented here suggest that catalysts with 2 ligands, or even 1 ligand, should be viable—we appreciate holding the metal ion on the catalyst will be a challenge—this might be achieved by use of suitable cosubstrates (the use of cosubstrates other than 2OG has been reported to retain the catalytic activity of human 2OG oxygenases[18,25]) or by use of metal trapping by a hydrophobic barrier as shown for other types of metal catalysts.

Acknowledgements

We thank the Wellcome Trust (106244/Z/14/Z), Cancer Research UK (C8717/A18245), and the Biotechnology and Biological Sciences Research Council (BB/J003018/1 and BB/R000344/1) for funding. A.B. acknowledges funding from the Biotechnology and Biological Sciences Research Council (grant number BB/M011224/1). L.B. thanks the Deutsche Forschungsgemeinschaft for a fellowship (BR 5486/2-1). Y.N. thanks the JSPS for an Overseas Research Fellowship (2020060219) and the Daiichi Sankyo Foundation of Life Science. We thank the Diamond Light Source and staff for allocation of beam time and support.

Conflict of interest

The authors declare no conflict of interest.

Keywords: aspartate/asparagine-β-hydroxylase · biomimetic catalysis · facial triad · metallo-enzymes · 2-oxoglutarate dependent oxygenase

[1] a) *2-Oxoglutarate-Dependent Oxygenases* (Eds.: C. J. Schofield, R. P. Hausinger), The Royal Society of Chemistry, Cambridge, **2015**; b) S.-S. Gao, N. Naowarojna, R. Cheng, X. Liu, P. Liu, *Nat. Prod. Rep.* **2018**, *35*, 792–837.
[2] a) W. G. Kaelin, P. J. Ratcliffe, *Mol. Cell* **2008**, *30*, 393–402; b) C. J. Schofield, P. J. Ratcliffe, *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 343–354; c) K. D. Rasmussen, K. Helin, *Genes Dev.* **2016**, *30*, 733–750; d) E. L. Greer, Y. Shi, *Nat. Rev. Genet.* **2012**, *13*, 343–357.
[3] a) H. M. Hanauske-Abel, V. Günzler, *J. Theor. Biol.* **1982**, *94*, 421–455; b) S. Martinez, R. P. Hausinger, *J. Biol. Chem.* **2015**, *290*, 20702–20711.
[4] S. Kal, L. Que, *J. Biol. Inorg. Chem.* **2017**, *22*, 339–365.
[5] a) J. Stenflo, E. Holme, S. Lindstedt, N. Chandramouli, L. H. Tsai Huang, J. P. Tam, R. B. Merrifield, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 444–447; b) R. S. Gronke, W. J. VanDusen, V. M. Garsky, J. W. Jacobs, M. K. Sardana, A. M. Stern, P. A. Friedman, *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 3609–3613; c) F. Korioth, C. Gieffers, J. Frey, *Gene* **1994**, *150*, 395–399.
[6] a) L. Lavaissiere, S. Jia, M. Nishiyama, S. de la Monte, A. M. Stern, J. R. Wands, P. A. Friedman, *J. Clin. Invest.* **1996**, *98*, 1313–1323; b) M. Kanwal, M. Smahel, M. Olsen, J. Smahelova, R. Tachezy, *J. Exp. Clin. Cancer Res.* **2020**, *39*, 163; c) W. Zheng, X. Wang, J. Hu, B. Bai, H. Zhu, *Oncol. Rep.* **2020**, *44*, 2364–2372.
[7] I. Pfeffer, L. Brewitz, T. Krojer, S. A. Jensen, G. T. Kochan, N. J. Kershaw, K. S. Hewitson, L. A. McNeill, H. Kramer, M. Münzel, R. J. Hopkinson, U. Oppermann, P. A. Handford, M. A. McDonough, C. J. Schofield, *Nat. Commun.* **2019**, *10*, 4910.
[8] L. Brewitz, Y. Nakashima, C. J. Schofield, *Chem. Sci.* **2021**, *12*, 1327–1342.
[9] J. M. Elkins, K. S. Hewitson, L. A. McNeill, J. F. Seibel, I. Schlemminger, C. W. Pugh, P. J. Ratcliffe, C. J. Schofield, *J. Biol. Chem.* **2003**, *278*, 1802–1806.

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

[10] a) Q. Zou, Y. Hou, H. Wang, K. Wang, X. Xing, Y. Xia, X. Wan, J. Li, B. Jiao, J. Liu, A. Huang, D. Wu, H. Xiang, T. M. Pawlik, H. Wang, W. Y. Lau, Y. Wang, F. Shen, *EBioMedicine* **2018**, *31*, 287–298; b) X. Dong, Q. Lin, A. Aihara, Y. Li, C.-K. Huang, W. Chung, Q. Tang, X. Chen, R. Carlson, C. Nadolny, G. Gabriel, M. Olsen, J. R. Wands, *Oncotarget* **2015**, *6*, 1231–1248; c) N. Ince, S. M. de la Monte, J. R. Wands, *Cancer Res.* **2000**, *60*, 1261–1266.
[11] a) P. Fernlund, J. Stenflo, *J. Biol. Chem.* **1983**, *258*, 12509–12512; b) B. A. McMullen, K. Fujikawa, W. Kisiel, T. Sasagawa, W. N. Howald, E. Y. Kwa, B. Weinstein, *Biochemistry* **1983**, *22*, 2875–2884.
[12] L. Brewitz, A. Tumber, C. J. Schofield, *J. Biol. Chem.* **2020**, *295*, 7826–7838.
[13] a) P. Koivunen, M. Hirsilä, V. Günzler, K. I. Kivirikko, J. Myllyharju, *J. Biol. Chem.* **2004**, *279*, 9899–9904; b) R. S. Gronke, D. J. Welsch, W. J. VanDusen, V. M. Garsky, M. K. Sardana, A. M. Stern, P. A. Friedman, *J. Biol. Chem.* **1990**, *265*, 8558–8565.
[14] P. Koivunen, M. Hirsilä, A. M. Remes, I. E. Hassinen, K. I. Kivirikko, J. Myllyharju, *J. Biol. Chem.* **2007**, *282*, 4524–4532.
[15] a) E. A. Siess, D. G. Brocks, H. K. Lattke, O. H. Wieland, *Biochem. J.* **1977**, *166*, 225–235; b) K. Thirstrup, S. Christensen, H. A. Møller, A. Ritzén, A.-L. Bergström, T. N. Sager, H. S. Jensen, *Pharmacol. Res.* **2011**, *64*, 268–273.
[16] A. M. Rydzik, I. K. H. Leung, G. T. Kochan, A. Thalhammer, U. Oppermann, T. D. W. Claridge, C. J. Schofield, *ChemBioChem* **2012**, *13*, 1559–1563.
[17] L. Brewitz, A. Tumber, I. Pfeffer, M. A. McDonough, C. J. Schofield, *Sci. Rep.* **2020**, *10*, 8650.
[18] L. Brewitz, A. Tumber, X. Zhang, C. J. Schofield, *Bioorg. Med. Chem.* **2020**, *28*, 115675.
[19] J.-H. Zhang, T. D. Y. Chung, K. R. Oldenburg, *J. Biomol. Screening* **1999**, *4*, 67–73.
[20] a) K. McGinnis, G. M. Ku, W. J. VanDusen, J. Fu, V. Garsky, A. M. Stern, P. A. Friedman, *Biochemistry* **1996**, *35*, 3957–3962; b) S. Jia, K. McGinnis, W. J. VanDusen, C. J. Burke, A. Kuo, P. R. Griffin, M. K. Sardana, K. O. Elliston, A. M. Stern, P. A. Friedman, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7227–7231.
[21] a) K. S. Hewitson, S. L. Holmes, D. Ehrismann, A. P. Hardy, R. Chowdhury, C. J. Schofield, M. A. McDonough, *J. Biol. Chem.* **2008**, *283*, 25971–25978; b) J. A. Hangasky, C. Y. Taabazuing, C. B. Martin, S. J. Eron, M. J. Knapp, *J. Inorg. Biochem.* **2017**, *166*, 26–33.
[22] J. R. Horton, A. K. Upadhyay, H. Hashimoto, X. Zhang, X. Cheng, *J. Mol. Biol.* **2011**, *406*, 1–8.
[23] L. Liu, H. Kim, A. Casta, Y. Kobayashi, L. S. Shapiro, A. M. Christiano, *FASEB J.* **2014**, *28*, 1534–1542.
[24] a) L. Que, W. B. Tolman, *Nature* **2008**, *455*, 333–340; b) V. A. Larson, B. Battistella, K. Ray, N. Lehnert, W. Nam, *Nat. Rev. Chem.* **2020**, *4*, 404–419; c) K. Gopalaiah, *Chem. Rev.* **2013**, *113*, 3248–3296; d) G. Olivo, O. Cussó, M. Borrell, M. Costas, *J. Biol. Inorg. Chem.* **2017**, *22*, 425–452.
[25] a) K. Majamaa, H. M. Hanauske-Abel, V. Günzler, K. I. Kivirikko, *Eur. J. Biochem.* **1984**, *138*, 239–245; b) M. Mukherji, N. J. Kershaw, C. H. MacKinnon, I. J. Clifton, A. S. Wierzbicki, C. J. Schofield, M. D. Lloyd, *Chem. Commun.* **2001**, 972–973; c) S. Martinez, M. Fellner, C. Q. Herr, A. Ritchie, J. Hu, R. P. Hausinger, *J. Am. Chem. Soc.* **2017**, *139*, 11980–11988.

Manuscript received: March 15, 2021
Accepted manuscript online: April 22, 2021
Version of record online: May 19, 2021

 © 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

15213773, 2021, 26, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/anie.202103711 by Oxford University, Wiley Online Library on [18/12/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

Human Oxygenase Variants Employing a Single Protein Fe^{II} Ligand Are Catalytically Active

*Amelia Brasnett[+], Inga Pfeffer[+], Lennart Brewitz[+], Rasheduzzaman Chowdhury, Yu Nakashima, Anthony Tumber, Michael A. McDonough, and Christopher J. Schofield**

anie_202103711_sm_miscellaneous_information.pdf

Supporting Information

Table of contents

1. Supporting figures and tables S2-S29
2. AspH variant inhibition assays S30
3. Crystallography S31
4. References S31-S32

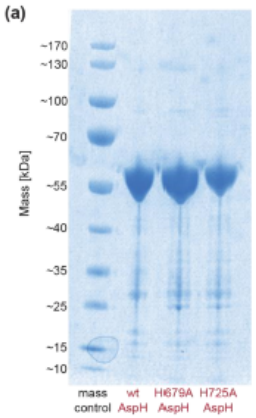
1. Supporting figures and tables

Supporting Figure S1. The H679A and H725A variants of human *N*-terminally His$_6$-tagged AspH$_{315-758}$ were obtained in highly purified form as analysed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) with Coomassie staining and MS analysis (continues on the following page).

Site-directed mutagenesis to obtain plasmid DNA coding for the H679A and H725A variants of the reported human wildtype (wt) *N*-terminal His$_6$-AspH$_{315-758}$ (His$_6$-AspH$_{315-758}$) construct[1] were carried out using the QuikChange (Stratagene) site-directed mutagenesis kit according to the manufacturer's instructions. For PCR experiments, high fidelity PfuTurbo DNA polymerase (Stratagene) was used, primers were designed for the H679A AspH variant (forward primer: 5′-CTCACGTGTGGCCGGCCACAGGGCCC-3′, reversed primer: 5′-GGGCCCTGTGGCCGGCCACACGTGAG-3′) and for the H725A AspH variant (forward primer: 5′-GATGACTCCTTTGAGGCCGAGGTATGGCAGG-3′, reversed primer: 5′-CCTGCCATACCTCGGCCTCAAAGGAGTCATC-3′).

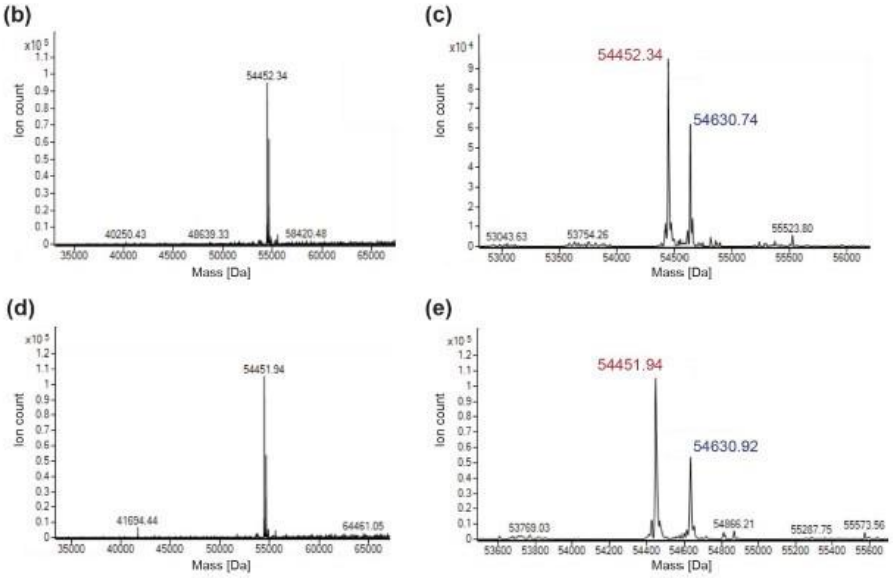
AspH variants were produced according to our reported procedure for wt His$_6$-AspH$_{315-758}$.[1] A pET-28a(+) vector encoding for the H679A or H725A His$_6$-AspH$_{315-758}$ variants was transformed into *E. coli* BL21 (DE3) cells. The resultant cells were grown in 2TY media supplemented with kanamycin (0.05 mM) at 37 °C with shaking (180 rpm). AspH production was induced at an OD_{600} of ~1.2 at 18 °C by adding isopropyl ß-D-thiogalactopyranoside (1 M) to a final concentration of 0.1 mM. Cells were shaken for 16 h at 18 °C (180 rpm) and harvested by centrifugation (8000 rpm, 8 min, 4 °C); the resultant cell pellets were stored at -80 °C. Frozen cells were resuspended (30 g / 100 mL) in ice-cold 50 mM HEPES buffer (pH 7.5, 500 mM NaCl, 5 mM imidazole) containing EDTA-free protease inhibitor cocktail tablets (1 tablet/50 mL; Roche Diagnostics or Sigma-Aldrich) and DNAse I (bovine pancreas, grade II, Roche Diagnostics). The cells were lysed by sonication on ice (8x30 sec bursts; Sonics Vibra-Cell VCX500, amplitude: 60%), the lysates were then centrifuged (20000 rpm, 30 min, 4 °C). The supernatant containing AspH was purified at 4 °C by Ni(II)-affinity chromatography (HisTrap HP column, GE Healthcare; 1 mL/min flow rate) using an ÄKTA Pure machine (GE Healthcare) with a gradient of wash (50 mM HEPES, pH 7.5, 500 mM NaCl, 40 mM imidazole) and elution buffers (50 mM HEPES, pH 7.5, 500 mM NaCl, 500 mM imidazole). Eluted fractions containing AspH were pooled, then concentrated using an Amicon Ultra centrifugal filter (4000 rpm, 4 °C), and further purified by size-exclusion chromatography using a HiLoad 26/60 Superdex 75 pg 300 mL column with a flow rate of 1 mL/min and 50 mM HEPES (pH 7.5, 150 mM NaCl) as elution buffer. AspH variants were >95% pure by SDS-PAGE analysis and had the anticipated mass. The H679A and H725A variants of His$_6$-AspH$_{315-758}$ were stored in 50 mM HEPES buffer (pH 7.5, 150 mM NaCl) at a concentration of approximately 100 µM at -78 °C; fresh aliquots were used for all biochemical experiments.

(**a**) 10-12% SDS-PAGE analysis of the H679A and H725A His$_6$-AspH$_{315-758}$ variants after Ni(II)-affinity chromatography (HisTrap column) and subsequent size-exclusion chromatography (Superdex 75).



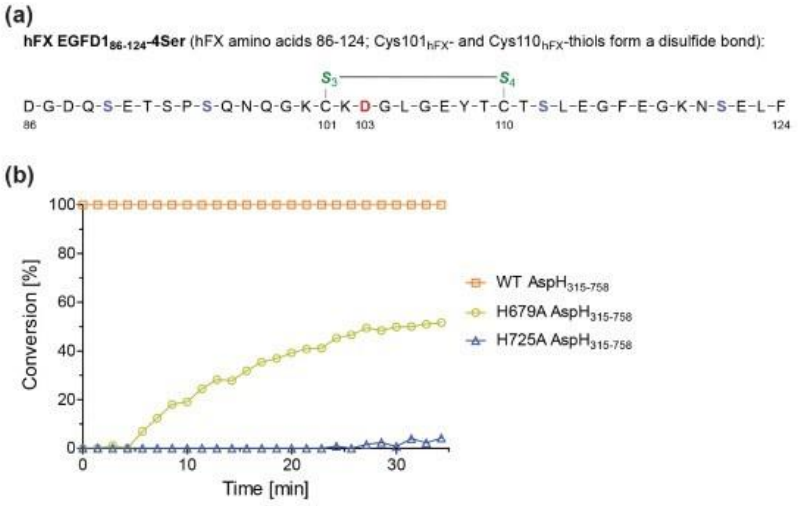
2

Deconvoluted mass spectra of (**b**, and a close-up in **c**) the purified H679A His_6-$AspH_{315-758}$ variant and (**d**, and a close-up in **e**) the purified H725A His_6-$AspH_{315-758}$ variant after size-exclusion chromatography show the expected calculated mass (in red) of the His_6-$AspH_{315-758}$ variants (H679A His_6-$AspH_{315-758}$: 54.452 kDa, calculated 54.453 kDa; H725A His_6-$AspH_{315-758}$: 54.452 kDa, calculated 54.453 kDa) and an additional species (+ 178 Da, in blue) corresponding to His_6-$AspH_{315-758}$ variants containing a gluconoylated His_6-tag, a common covalent modification of recombinant *N*-terminal His_6-tagged proteins produced in *E. coli*[2], which does not impact substantially (if at all) on AspH-catalysis.



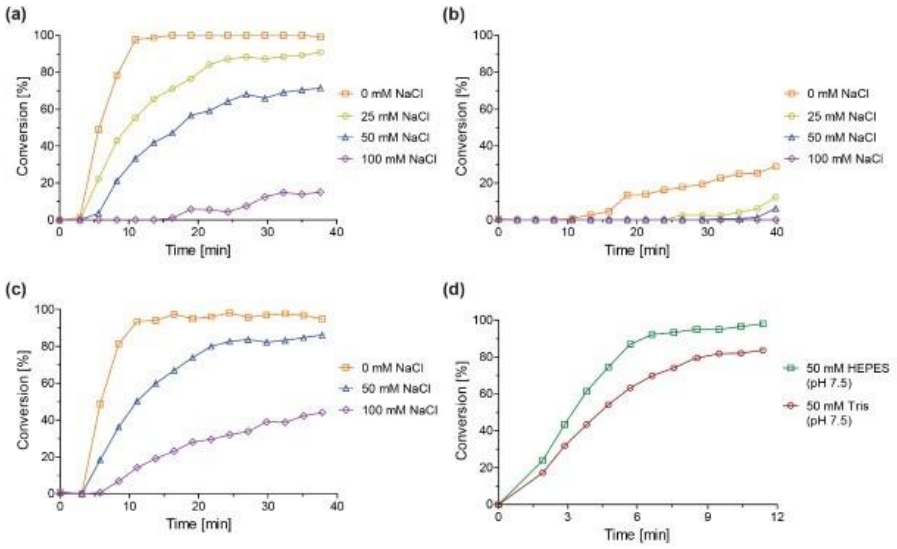
3

Supporting Figure S2. Initial turnover assays of wt AspH and its H679A and H725A variants reveal catalytic activity. (**a**) The hFX-EGFD1$_{86\text{-}124}$-4Ser peptide[1] used for kinetic and crystallographic experiments is based on the sequence of EGFD1 of the confirmed AspH substrate human coagulation Factor X (hFX).[3] The peptide macrocycle formed by the disulphide between Cys101$_{hFX}$ and Cys110$_{hFX}$ mimics a non-canonical (*i.e.* Cys 1–2, 3–4, 5–6) EGFD disulfide pattern, which has been previously identified as a substrate requirement for AspH catalysis.[1] The peptide was synthesized and purified by GL Biochem (Shanghai) Ltd. The AspH hydroxylation site (Asp103$_{hFX}$) is in red; cystine sulfurs are in green; serine residues substituting for hFX cysteine residues are in light blue; numbering is according to the sequence of hFX;[3] (**b**) under the reaction conditions, wt His$_6$-AspH$_{315\text{-}758}$ catalyses the hydroxylation of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substantially faster than the H679A variant (>95% conversion after 30 min vs ~50% conversion for H679A AspH), which is much more active than the H725A variant (<5% conversion after 30 min). Reactions were initiated by addition of the His$_6$-AspH$_{315\text{-}758}$ variants (final concentration: 0.3 μM) to the reaction mixture containing the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide (4.0 μM), 2OG (200 μM), LAA (400 μM), and FAS (10 μM) in reaction buffer (50 mM Tris, 50 mM NaCl, pH 7.5) in separate wells of a 96 deep well plate (Greiner). The AspH-catalysed hydroxylation of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide was monitored by SPE-MS.



Supporting Figure S3. The catalytic activity of the AspH variants was highest in HEPES buffer in the absence of NaCl. H679A His_6-$AspH_{315-758}$ activity decreases with increasing NaCl concentration and increasing pH, in agreement with a study on wt AspH.[4] The H679A AspH variant was more active in 50 mM HEPES buffer (pH 7.5) than in 50 mM Tris buffer (pH 7.5). The HEPES concentration was reduced to 25 mM in order to improve peptide detection by MS. Optimized buffer conditions for all subsequent kinetic and inhibition SPE-MS assays of the H679A and H725A AspH variants were: 25 mM HEPES, pH 7.5. Reactions were initiated by addition of the H679A His_6-$AspH_{315-758}$ variant (final concentration: 0.3 μM) to the reaction mixture containing the hFX-$EGFD1_{86-124}$-4Ser peptide (4.0 μM), 2OG (200 μM), LAA (400 μM), and FAS (10 μM) in reaction buffer. AspH-catalysed hydroxylation of the hFX-$EGFD1_{86-124}$-4Ser peptide (Supporting Figure S2a) was monitored by SPE-MS.

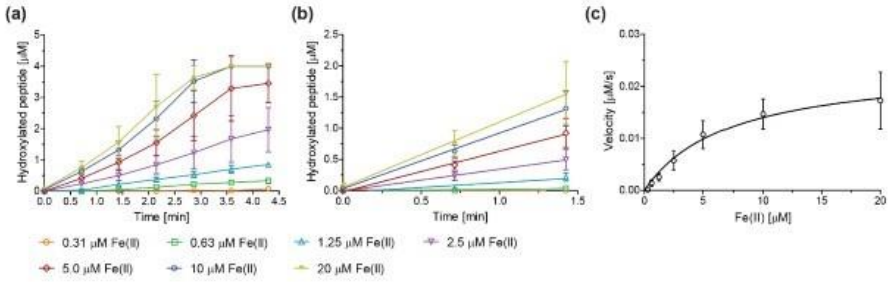
(**a**) Buffer: 50 mM Tris, pH 7.5. Colour code: orange boxes: 0 mM NaCl, yellow circles: 25 mM NaCl, blue triangle: 50 mM NaCl, violet diamond: 100 mM NaCl; (**b**) buffer: 50 mM Tris, pH 8.0. Colour code: orange boxes: 0 mM NaCl, yellow circles: 25 mM NaCl, blue triangle: 50 mM NaCl, violet diamond: 100 mM NaCl; (**c**) buffer: 50 mM HEPES, pH 7.5. Colour code: orange boxes: 0 mM NaCl, blue triangle: 50 mM NaCl, violet diamond: 100 mM NaCl; (**d**) direct comparison of HEPES and Tris reaction buffers using 0.3 μM H679A AspH (enzyme/substrate ratio = 1/20). Colour code: green boxes: 50 mM HEPES (pH 7.5, 0 mM NaCl), red circles: 50 mM Tris (pH 7.5, 0 mM NaCl).



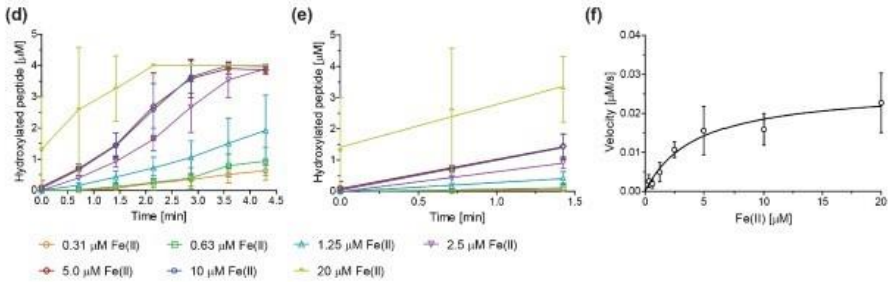
5

Supporting Figure S4. Kinetic parameters for wt AspH (continues on the following page). Turnover numbers (k_{cat}^{app}) and Michaelis constants (K_m^{app}) of wt AspH were determined in independent triplicates for 2OG and ammonium iron(II) sulfate hexahydrate (FAS, $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$), both in the presence and absence of L-ascorbic acid (LAA) monitoring the AspH-catalysed hydroxylation of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide (Supporting Figure S2a) by SPE-MS (SPE-MS parameters are given in Section 2 of the Supporting Information). Note that the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate cannot be detected by SPE-MS at low concentrations, hampering an accurate determination of its wt AspH K_m-value. Conditions: 0.1 µM wt His$_6$-AspH$_{315\text{-}758}$ using variable concentrations of 2OG, FAS, and hFX-EGFD1$_{86\text{-}124}$-4Ser, respectively, in 25 mM HEPES (pH 7.5, 20° C; note, to increase enzyme activity, 50 mM NaCl was added to the reaction buffer). Measurement times were normalized to the first sample injection analysed after the addition of the Enzyme Mixture to the Substrate Mixture (t = 0 s), by which time low levels of hydroxylation were manifest. Data are shown as the mean average of three independent runs (n = 3; mean ± standard deviation, SD). Turnover numbers (k_{cat}^{app}) were calculated from the maximum velocities (v_{max}^{app}) values obtained, assuming ideal active site availability as supported by a reported active site titration of wt AspH.[4]

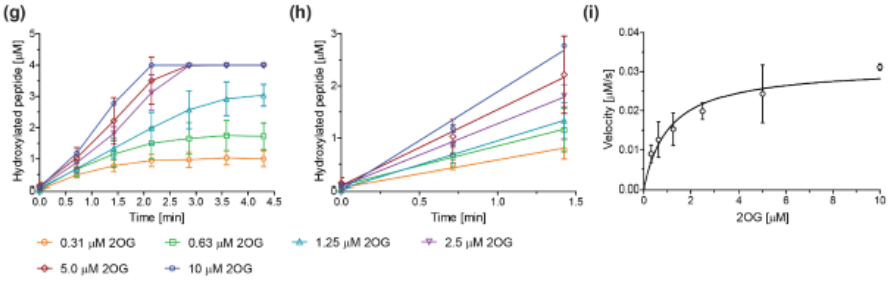
(**a**) Time course of the wt AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate in the absence of LAA for the shown Fe(II) concentrations (conditions: 0.1 µM wt His$_6$-AspH$_{315\text{-}758}$, 20 µM 2OG, and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**b**) hydroxylation rates used to determine kinetic parameters of wt AspH for Fe(II) in the absence of LAA; (**c**) K_m^{app} of wt AspH for Fe(II) in the absence of LAA.



(**d**) Time course of the wt AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate in the presence of LAA for the shown Fe(II) concentrations (conditions: 0.1 µM wt His$_6$-AspH$_{315\text{-}758}$, 100 µM LAA, 20 µM 2OG, and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**e**) hydroxylation rates used to determine kinetic parameters of wt AspH for Fe(II) in the presence of LAA; (**f**) K_m^{app} of wt AspH for Fe(II) in the presence of LAA.



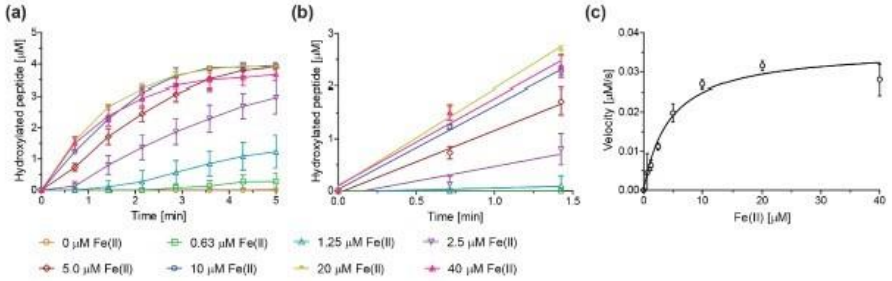
(**g**) Time course of the wt AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate for the shown 2OG concentrations (conditions: 0.1 µM wt His$_6$-AspH$_{315\text{-}758}$, 100 µM LAA, 20 µM Fe(II), and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**h**) hydroxylation rates used to determine kinetic parameters of wt AspH for 2OG; (**i**) K_m^{app} of wt AspH for 2OG.



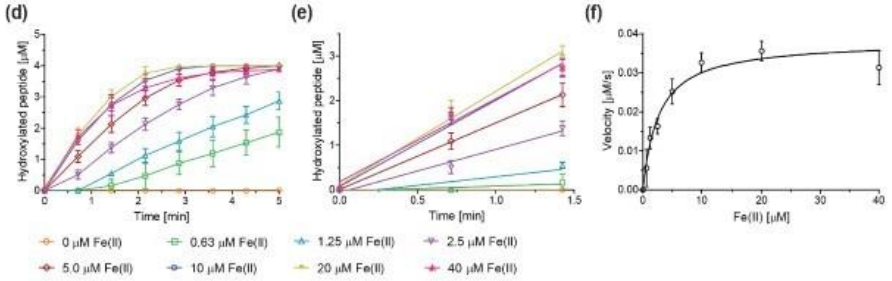
7

Supporting Figure S5. Kinetic parameters for H679A AspH (continues on the following page). Turnover numbers (k_{cat}^{app} or k_{cat}) and Michaelis constants (K_m^{app} or K_m) of the H679A AspH variant were determined in independent triplicates for 2OG, ammonium iron(II) sulfate hexahydrate (FAS, $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$), both in the presence and absence of L-ascorbic acid (LAA), and the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate (Supporting Figure S2a), monitoring the H679A AspH-catalysed hydroxylation of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide by SPE-MS (SPE-MS parameters are given in Section 2 of the Supporting Information). Conditions: 0.2 µM H679A His$_6$-AspH$_{315\text{-}758}$ using variable concentrations of 2OG, FAS, and hFX-EGFD1$_{86\text{-}124}$-4Ser, respectively, in 25 mM HEPES (pH 7.5, 20° C). Measurement times were normalized to the first sample injection analysed after the addition of the Enzyme Mixture to the Substrate Mixture (t = 0 s), by which time low levels of hydroxylation were manifest. Data are shown as the mean average of three independent runs (n = 3; mean ± standard deviation, SD). Turnover numbers (k_{cat}^{app} or k_{cat}) were calculated from the maximum velocities (v_{max}^{app} or v_{max}) values obtained, assuming ideal active site availability as supported by a reported active site titration of wt AspH.[4]

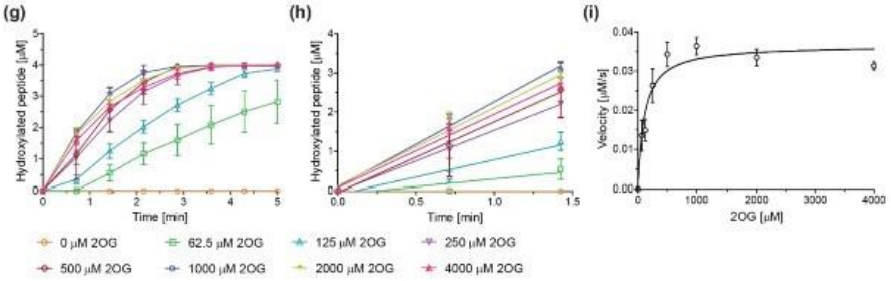
(**a**) Time course of the H679A AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate in the absence of LAA for the shown Fe(II) concentrations (conditions: 0.2 µM H679A His$_6$-AspH$_{315\text{-}758}$, 2000 µM 2OG, and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**b**) hydroxylation rates used to determine kinetic parameters of H679A AspH for Fe(II) in the absence of LAA; (**c**) K_m^{app} of H679A AspH for Fe(II) in the absence of LAA.



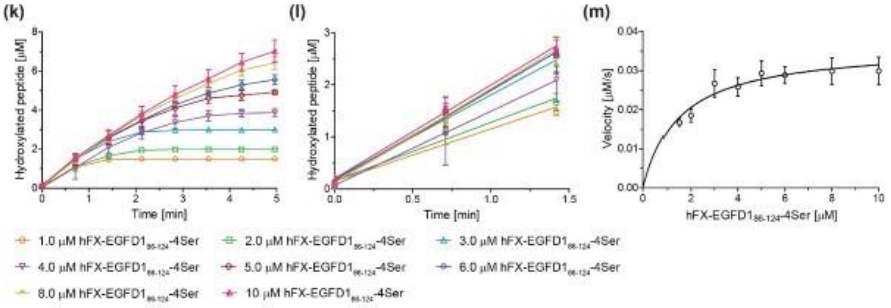
(**d**) Time course of the H679A AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate in the presence of LAA for the shown Fe(II) concentrations (conditions: 0.2 µM H679A His$_6$-AspH$_{315\text{-}758}$, 400 µM LAA, 2000 µM 2OG, and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**e**) hydroxylation rates used to determine kinetic parameters of H679A AspH for Fe(II) in the presence of LAA; (**f**) K_m^{app} of H679A AspH for Fe(II) in the presence of LAA.



(**g**) Time course of the H679A AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate for the shown 2OG concentrations (conditions: 0.2 µM H679A His$_6$-AspH$_{315\text{-}758}$, 400 µM LAA, 20 µM Fe(II), and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**h**) hydroxylation rates used to determine kinetic parameters of H679A AspH for 2OG; (**i**) K_m^{app} of H679A AspH for 2OG.



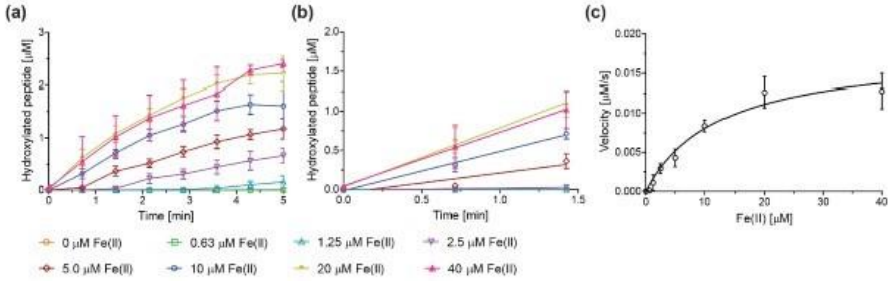
(**j**) Time course of the H679A AspH variant-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate for the shown hFX-EGFD1$_{86\text{-}124}$-4Ser concentrations (conditions: 0.2 µM H679A His$_6$-AspH$_{315\text{-}758}$, 400 µM LAA, 2000 µM 2OG, and 20 µM FAS); (**k**) hydroxylation rates used to determine kinetic parameters of H679A AspH for hFX-EGFD1$_{86\text{-}124}$-4Ser. Note that the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate cannot be detected by SPE-MS at concentrations below 1.0 µM, hampering a more accurate determination of the K_m-value for hFX-EGFD1$_{86\text{-}124}$-4Ser; (**l**) K_m of H679A AspH for the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate.



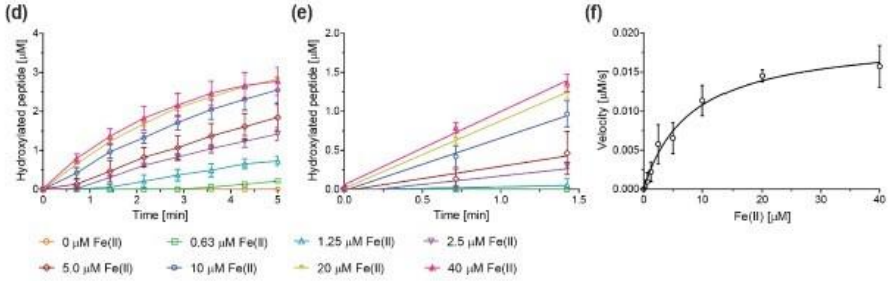
9

Supporting Figure S6. Kinetic parameters for H725A AspH (continues on the following page). Turnover numbers (k_{cat}^{app} or k_{cat}) and Michaelis constants (K_m^{app} or K_m) of the H725A AspH variant were determined in independent triplicates for 2OG, ammonium iron(II) sulfate hexahydrate (FAS, $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$), both in the presence and absence of L-ascorbic acid (LAA), and the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate (Supporting Figure S2a), monitoring the H725A AspH-catalysed hydroxylation of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide by SPE-MS (SPE-MS parameters are given in Section 2 of the Supporting Information). Conditions: 0.2 µM H725A His$_6$-AspH$_{315\text{-}758}$ using variable concentrations of 2OG, FAS, and hFX-EGFD1$_{86\text{-}124}$-4Ser, respectively, in 25 mM HEPES (pH 7.5, 20° C). Measurement times were normalized to the first sample injection analysed after the addition of the Enzyme Mixture to the Substrate Mixture (t = 0 s), by which time low levels of hydroxylation were manifest. Data are shown as the mean average of three independent runs (n = 3; mean ± standard deviation, SD). Turnover numbers (k_{cat}^{app} or k_{cat}) were calculated from the maximum velocities (v_{max}^{app} or v_{max}) values obtained, assuming ideal active site availability as supported by a reported active site titration of wt AspH.[4]

(**a**) Time course of the H725A AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate in the absence of LAA for the shown Fe(II) concentrations (conditions: 0.2 µM H725A His$_6$-AspH$_{315\text{-}758}$, 2000 µM 2OG, and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**b**) hydroxylation rates used to determine kinetic parameters of H725A AspH for Fe(II) in the absence of LAA; (**c**) K_m^{app} of H725A AspH for Fe(II) in the absence of LAA.

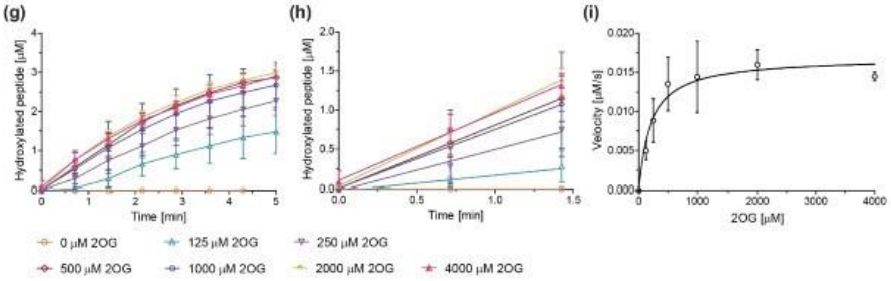


(**d**) Time course of the H725A AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate in the presence of LAA for the shown Fe(II) concentrations (conditions: 0.2 µM H725A His$_6$-AspH$_{315\text{-}758}$, 400 µM LAA, 2000 µM 2OG, and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**e**) hydroxylation rates used to determine kinetic parameters of H725A AspH for Fe(II) in the presence of LAA; (**f**) K_m^{app} of H725A AspH for Fe(II) in the presence of LAA.

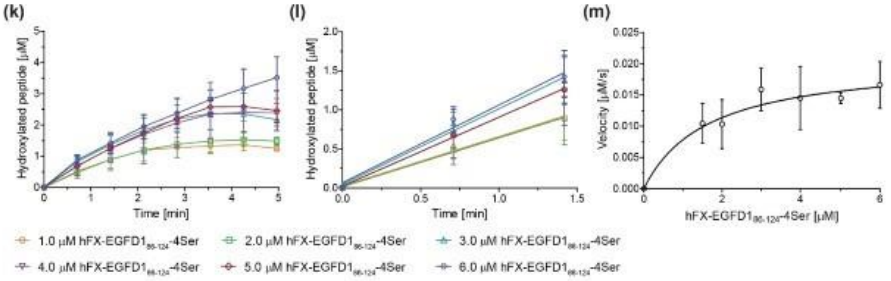


10

(**g**) Time course of the H725A AspH-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate for the shown 2OG concentrations (conditions: 0.2 µM H725A His$_6$-AspH$_{315\text{-}758}$, 400 µM LAA, 20 µM Fe(II), and 4.0 µM hFX-EGFD1$_{86\text{-}124}$-4Ser); (**h**) hydroxylation rates used to determine kinetic parameters of H725A AspH for 2OG; (**i**) K_m^{app} of H725A AspH for 2OG.



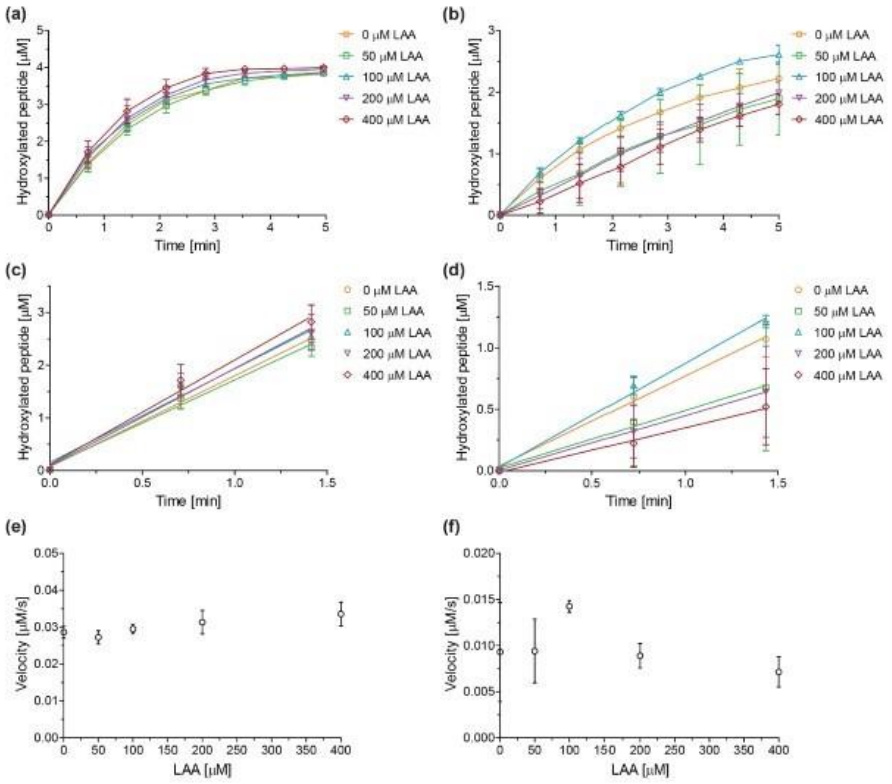
(**j**) Time course of the H725A AspH variant-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate for the shown hFX-EGFD1$_{86\text{-}124}$-4Ser concentrations (conditions: 0.2 µM H725A His$_6$-AspH$_{315\text{-}758}$, 400 µM LAA, 2000 µM 2OG, and 20 µM FAS); (**k**) hydroxylation rates used to determine kinetic parameters of H725A AspH for hFX-EGFD1$_{86\text{-}124}$-4Ser. Note that the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate cannot be detected by SPE-MS at concentrations below 1.0 µM, hampering a more accurate determination of the K_m-value for hFX-EGFD1$_{86\text{-}124}$-4Ser; (**l**) K_m of H725A AspH for the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide substrate.



11

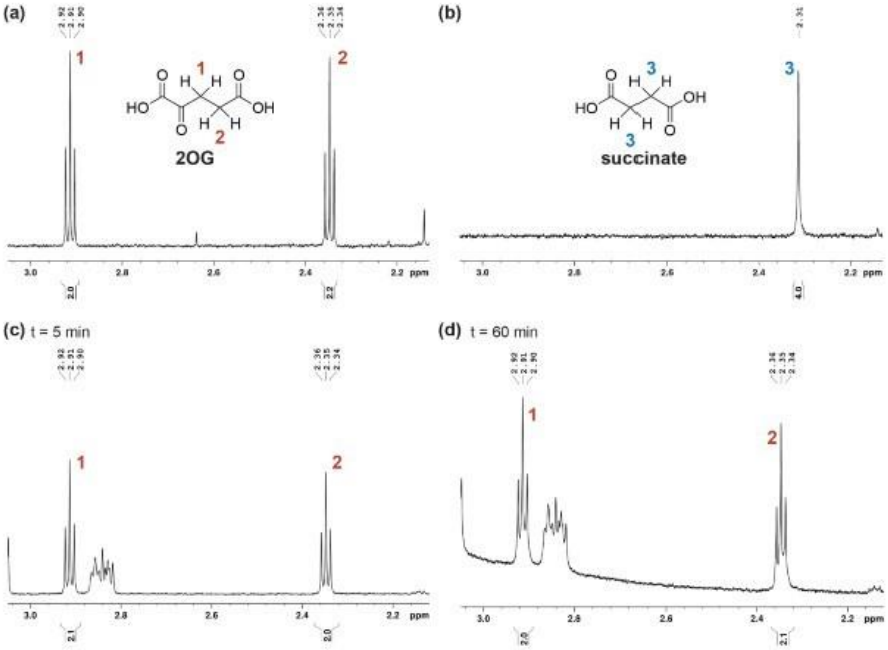
Supporting Figure S7. The H679A and H725A AspH variants do not require L-ascorbic acid for activity. The kinetic parameters of the H679A and H725A AspH variants were determined in independent triplicates for L-ascorbic acid (LAA), monitoring the AspH variant-catalysed hydroxylation of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide (Supporting Figure S2a) by SPE-MS (SPE-MS parameters are given in Section 2 of the Supporting Information). Conditions: 0.2 μM H679A or H725A His$_6$-AspH$_{315\text{-}758}$ in 25 mM HEPES (pH 7.5, 20° C), 4.0 μM hFX-EGFD1$_{86\text{-}124}$-4Ser, 2000 μM 2OG, 20 μM ammonium iron(II) sulfate hexahydrate (FAS), and variable concentrations of LAA. Measurement times were normalized to the first sample injection analysed after the addition of the Enzyme Mixture to the Substrate Mixture (t = 0 s), by which time low levels of hydroxylation were manifest. Data are shown as the mean average of three independent runs (n = 3; mean ± standard deviation, SD). From the initial maximum velocities (v_{max}^{app}) values obtained, turnover numbers (k_{cat}^{app}) were calculated assuming ideal active site availability as supported by an active site titration of wt AspH.[4]

(**a** and **b**) Time courses of the (**a**) H679A and (**b**) H725A AspH variant-catalysed hydroxylation reaction of the hFX-EGFD1$_{86\text{-}124}$-4Ser substrate peptide for the shown concentrations of LAA; (**c** and **d**) hydroxylation rates used to determine kinetic parameters of the (**c**) H679A and (**d**) H725A AspH variant for LAA; (**e** and **f**) effect of LAA on the (**e**) H679A and (**f**) H725A AspH variant activity, the AspH variant k_{cat}^{app}-values for LAA are constant throughout the concentration range investigated, including in the absence of LAA indicating that LAA is not required for the AspH variant activity. The AspH variant k_{cat}^{app}-values for LAA are in the range of those determined when varying the 2OG, Fe(II), or substrate peptide concentrations.



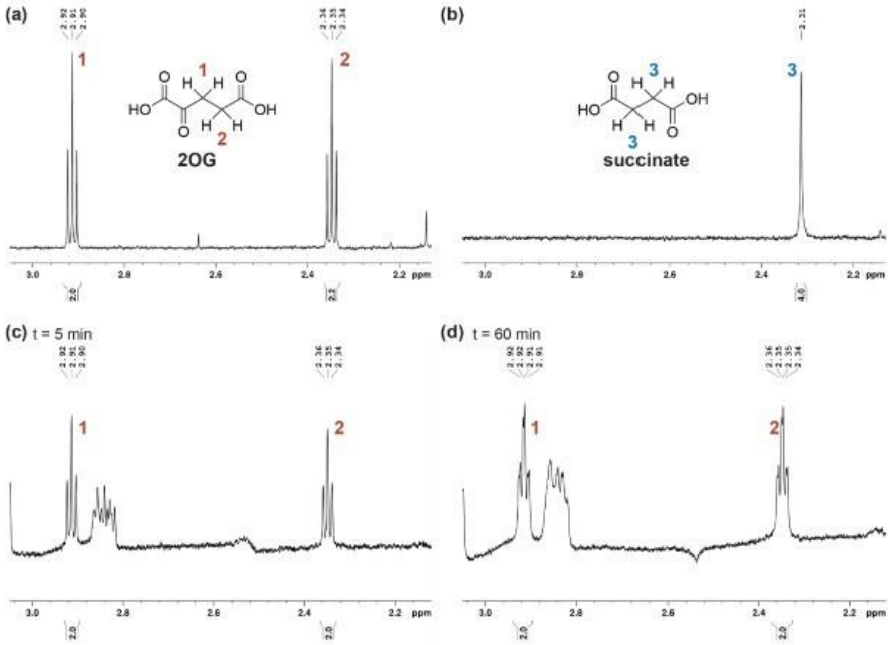
Supporting Figure S8. The H679A and H725A AspH variants do not efficiently catalyse the oxidative decarboxylation of 2OG to succinate and CO_2 in the absence of substrate (continues on the following page). The H679A and H725A AspH variant-catalysed oxidative decarboxylation of 2OG uncoupled from substrate oxidation was monitored using a Bruker AVIII 700 NMR machine equipped with a 5mm $^{1}H(^{13}C/^{15}N)$ inverse cryoprobe. Assay conditions: 1 µM H679A or H725A His_6-$AspH_{315-758}$ variant, 200 µM 2OG, and 20 µM ammonium iron(II) sulfate hexahydrate (FAS, $(NH_4)_2Fe(SO_4)_2{\cdot}6H_2O$) in 50 mM aqueous Tris-$d_{11}$ containing 10%$_{v/v}$ D_2O (pH 7.5, 25° C).

(I) Using H679A His_6-$AspH_{315-758}$: (**a**) Close-up (3.0-2.2 ppm) of the ^{1}H NMR spectrum of commercially sourced 2OG in 50 mM aqueous Tris-d_{11} containing 10%$_{v/v}$ D_2O (pH 7.5, 25° C); (**b**) close-up (3.0-2.2 ppm) of the ^{1}H NMR spectrum of commercially sourced succinate in 50 mM aqueous Tris-d_{11} containing 10%$_{v/v}$ D_2O (pH 7.5, 25° C); (**c** and **d**) a timecourse NMR experiment (t = 5 min, **c**; t = 60 min, **d**) reveals that the H679A AspH does not efficiently catalyse oxidative decarboxylation of 2OG to succinate and CO_2 in the absence of substrate. The results are in agreement NMR conversion assays using wt AspH.[1]



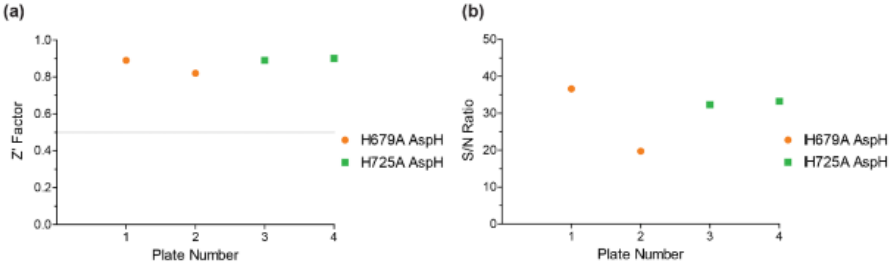
13

(II) Using H725A His_6-$AspH_{315-758}$: (**a**) Close-up (3.0-2.2 ppm) of the ^1H NMR spectrum of commercially sourced 2OG in 50 mM aqueous Tris-d_{11} containing 10%$_{v/v}$ D_2O (pH 7.5, 25° C); (**b**) close-up (3.0-2.2 ppm) of the ^1H NMR spectrum of commercially sourced succinate in 50 mM aqueous Tris-d_{11} containing 10%$_{v/v}$ D_2O (pH 7.5, 25° C); (**c** and **d**) a timecourse NMR experiment (t = 5 min, **c**; t = 60 min, **d**) reveals that the H725A AspH does not efficiently catalyse oxidative decarboxylation of 2OG to succinate and CO_2 in the absence of substrate. The results are in agreement NMR conversion assays using wt AspH.[1]



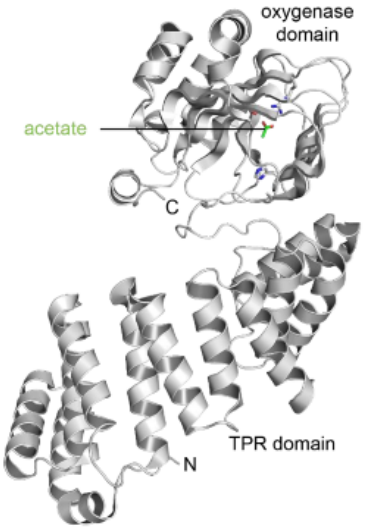
14

Supporting Figure S9. Robustness of the H679A and H725A AspH variant SPE-MS inhibition assays. SPE-MS inhibition assays were performed as described in Section 2 of the Supporting Information. (**a**) Z'-factors[5] and (**b**) signal-to-noise ratios (S/N) for the inhibition assay plates analysed to determine IC_{50}-values. The Z'-factors >0.5 (grey line) indicate a stable and robust assay of high quality.[5] Z'-factors and S/N-ratios were determined according to the literature using Microsoft Excel.[5] Colour code: H679A AspH variant: orange circles; H725A AspH variant: green boxes.



15

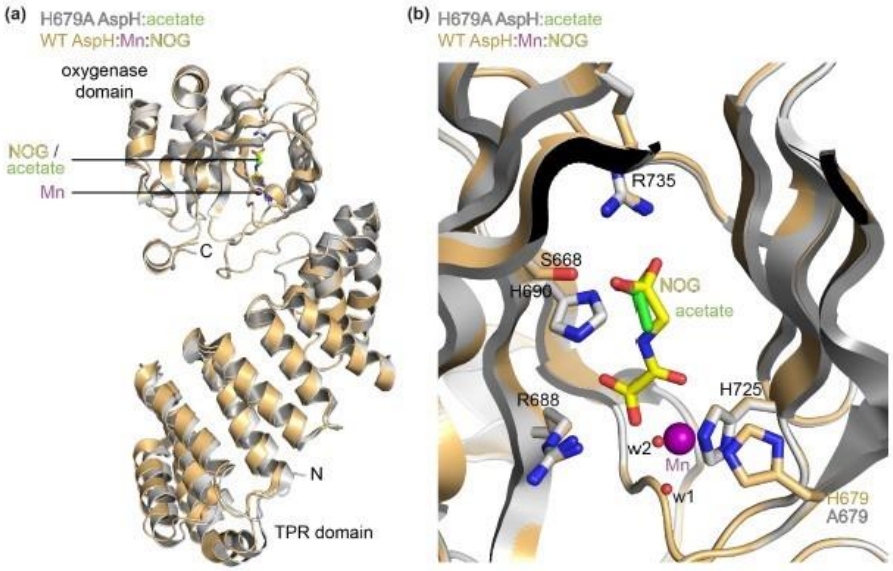
Supporting Figure S10. View of the protein fold from a crystal structure of H679A AspH with acetate present at the active site (H679A AspH:acetate; PDB ID: 6QA5). Overview of the H679A AspH:acetate crystal structure (2.7 Å resolution); Note, electron density for a metal ion bound to the H679A AspH active site was not observed. Colours: grey: H679A His_6-$AspH_{315-758}$; green: carbon-backbone of acetate; red: oxygen; blue: nitrogen. Note, electron density for a metal ion bound to the H679A AspH active site was not observed.



16

Supporting Figure S11. The substitution of H679 by alanine neither substantially alters the overall AspH fold nor the conformations of key active site residues. Colours: grey: H679A His_6-$AspH_{315-758}$; green: carbon-backbone of acetate; red: oxygen; blue: nitrogen. w: water.

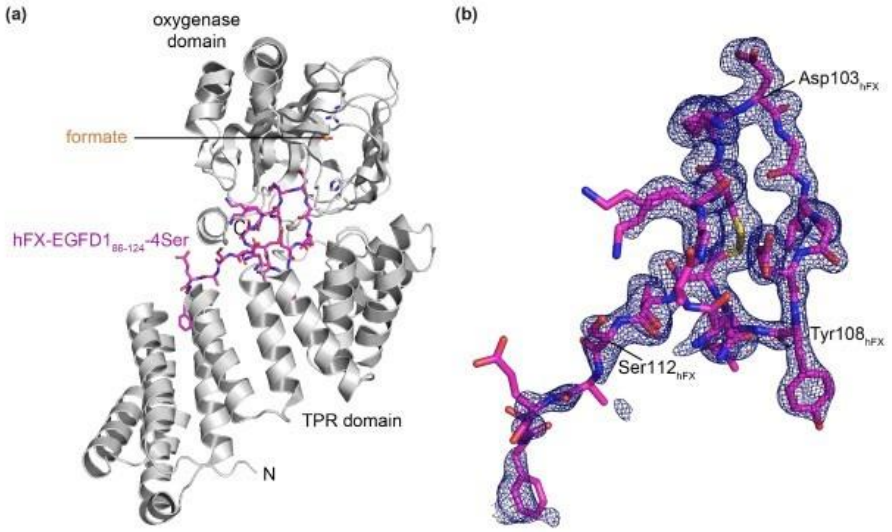
(**a**) Superimposition of a view from the H679A AspH:acetate structure (Supporting Figure S10) with one from the reported AspH:Mn:NOG structure (bronze: wt AspH; yellow: carbon-backbone of NOG; violet: Mn; PDB ID: 5JZA)[1] reveals similar AspH conformations (Cα RMSD = 0.276 Å); (**b**) superimposition of a view from the active site of the H679A AspH:acetate structure with one from the reported AspH:Mn:NOG structure (bronze: wt AspH; yellow: carbon-backbone of NOG; violet: Mn; PDB ID: 5JZA)[1] reveals that the conformations of those wt AspH residues directly engaged in NOG/2OG binding (*i.e.* S668, R688, H690, and R735)[1,6] do not alter substantially upon substituting H679 for alanine, even though electron density for a metal ion is not observed. The conformation of H725, whose side chain imidazole engages in metal binding, does not alter substantially upon substituting H679 by alanine. Note, that no substantial conformational changes in the wt AspH structure have been reported to occur when substituting the broad-spectrum 2OG oxygenase inhibitor NOG by the natural cosubstrate 2OG.[6]



17

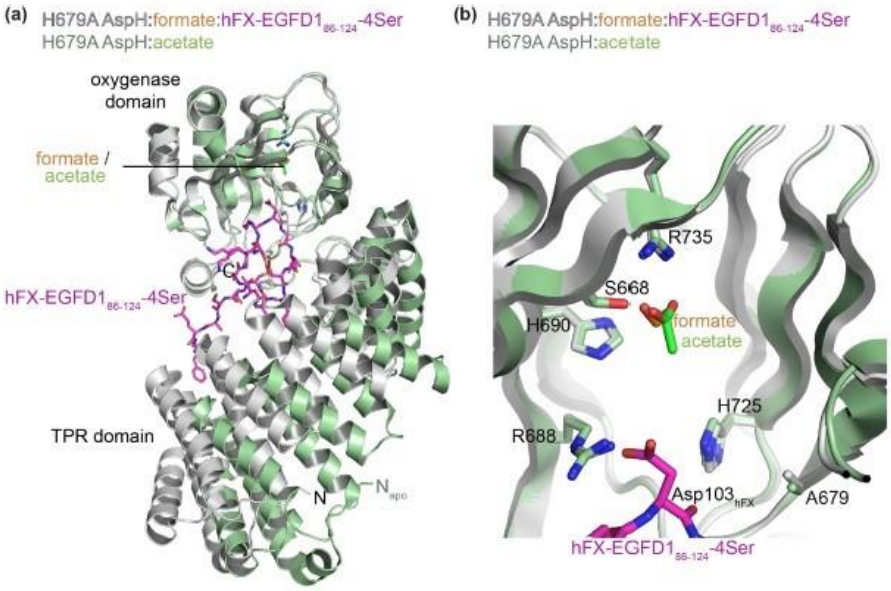
Supporting Figure S12. Views from a crystal structure of H679A AspH complexed with formate at the active site and a synthetic EGFD substrate peptide (H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser; PDB ID: 6Q9I). Colours: grey: H679A His$_6$-AspH$_{315\text{-}758}$; orange: carbon-backbone of formate; magenta: carbon-backbone of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen.

(**a**) Overview of the H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser crystal structure (1.9 Å resolution). Note, electron density for a metal ion bound to the H679A AspH active site was not observed; (**b**) OMIT electron density map (mF_o–DF_c) contoured to 3σ around the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide from the H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser structure reveals electron density for residues Gly99$_{hFX}$ to Ser112$_{hFX}$ including for the disulfide bridged (Cys101$_{hFX}$ and Cys110$_{hFX}$) ten-membered non-canonical EGFD macrocycle. Clear electron density is observed for the side chain of Tyr108$_{hFX}$, an EGFD residue which is part of the AspH-substrate consensus sequence requirement; the Tyr108$_{hFX}$ side chain interacts with the TPR domain of AspH as revealed in reported wt AspH:substrate complex structures.[1]



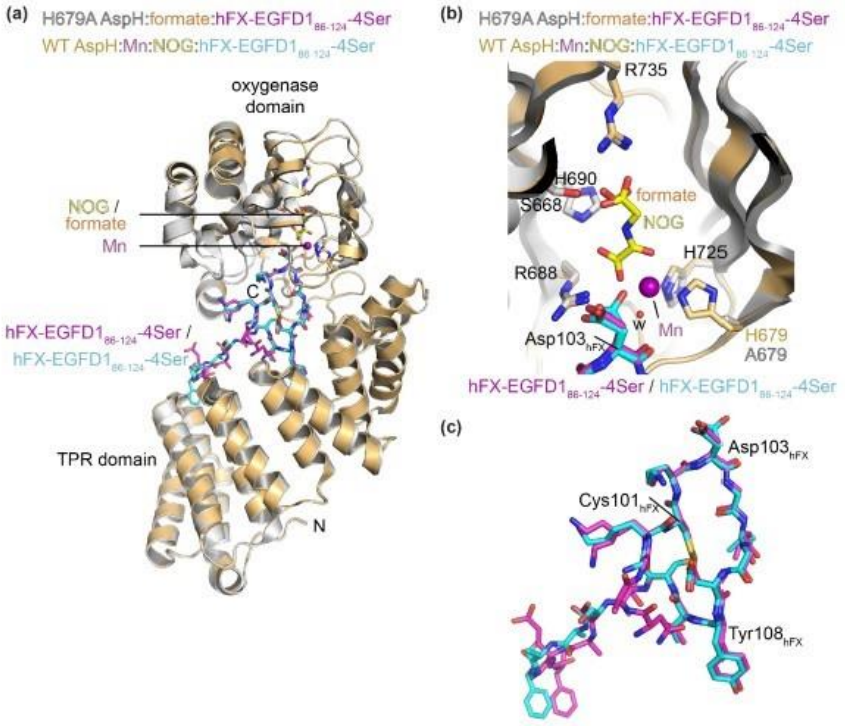
Supporting Figure S13. H679A AspH binds EGFD substrates through an induced fit mechanism in the absence of 2OG and metal ions. Colours: grey: H679A His_6-$AspH_{315-758}$; orange: carbon-backbone of formate; magenta: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen.

(**a**) Superimposition of a view from the H679A AspH:hFX-$EGFD1_{86-124}$-4Ser structure (Supporting Figure S12) with one from the H679A AspH:acetate structure (H679A AspH: pale green; carbon-backbone of acetate: green; Supporting Figure S10) reveals that the distance between the Cα atoms of Leu433 on TPR repeat α6 and Pro756 in the H679A AspH C-terminal region decreases from ~18 Å to ~13 Å upon substrate binding (Cα RMSD = 2.755 Å). Evidence for an induced fit mechanism involving this conformational change has been described for wt AspH crystallized in the presence/absence of EGFD peptide substrates using NOG as an inhibitor or 2OG as a cosubstrate and Mn as metal ion substituting for Fe(II);[1,6] (**b**) superimposition of a view from the active site of the H679A AspH:hFX-$EGFD1_{86-124}$-4Ser structure (Supporting Figure S12) with one from the H679A AspH:acetate structure (H679A AspH: pale green; carbon-backbone of acetate: green; Supporting Figure S10) reveals that the conformations of those H679A AspH residues directly engaged in NOG/2OG binding in wt AspH (*i.e.* S668, R688, H690, and R735)[1,6] do not alter substantially upon substrate binding, even though electron density for a metal ion is not observed. The conformation of H725, whose side chain imidazole engages in metal binding in wt AspH, does not alter substantially upon substrate binding. The conformations of several (*i.e.* Glu615, Asp616, Glu617, Arg620, Gln627, and Gln632), but not all (*i.e.* Arg686 and Arg688), H679A AspH residues engaged in EGFD substrate binding change upon substrate binding in a similar manner as reported for wt AspH structures in the presence/absence of EGFD peptide substrates when using NOG as an inhibitor or 2OG as a cosubstrate and Mn as metal ion substituting for Fe(II).[1,6]



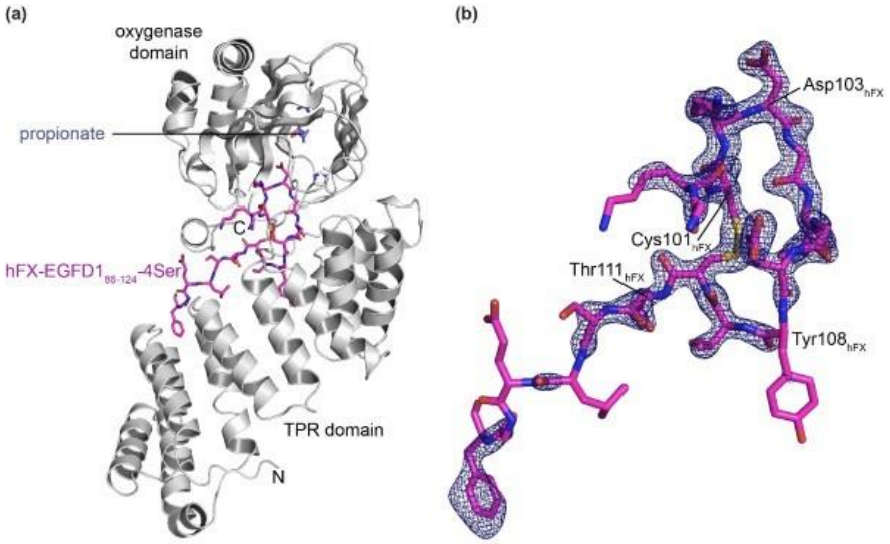
Supporting Figure S14. The substitution of H679 by alanine neither substantially alters the overall AspH fold nor the conformation of key active site residues in the presence of a substrate peptide. Colours: grey: H679A His_6-$AspH_{315-758}$; orange: carbon-backbone of formate; magenta: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen. w: water.

(**a**) Superimposition of a view from the H679A AspH:hFX-EGFD1$_{86-124}$-4Ser structure (Supporting Figure S12) with one from the reported AspH:Mn:NOG:hFX-EGFD1$_{86-124}$-4Ser structure (bronze: wt AspH; yellow: carbon-backbone of NOG; cyan: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide; violet: Mn; PDB ID: 5JQY)[1] reveals similar AspH conformations (Cα RMSD = 0.300 Å); (**b**) superimposition of a view from the active site of the H679A AspH:hFX-EGFD1$_{86-124}$-4Ser structure with one from the reported AspH:Mn:NOG:hFX-EGFD1$_{86-124}$-4Ser structure (bronze: wt AspH; yellow: carbon-backbone of NOG; cyan: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide; violet: Mn; PDB ID: 5JQY)[1] reveals that the conformations of those wt AspH residues directly engaged in NOG/2OG binding (*i.e.* S668, R688, H690, and R735)[1,6] do not alter substantially upon substituting H679 for alanine, even though electron density for a metal ion is not observed. The conformation of H725, whose side chain imidazole engages in metal binding, does not alter substantially upon substituting H679 by alanine; (**c**) superimposition of the hFX-EGFD1$_{86-124}$-4Ser substrate peptide from the H679A AspH:hFX-EGFD1$_{86-124}$-4Ser structure with the one from the reported AspH:Mn:NOG:hFX-EGFD1$_{86-124}$-4Ser structure (cyan: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide; PDB ID: 5JQY)[1] reveals similar peptide conformations (Cα RMSD = 0.700 Å), in particular of the amino acids forming the disulfide bridged ($Cys101_{hFX}$ and $Cys110_{hFX}$) ten-membered non-canonical EGFD macrocycle. Note that the $Asp103_{hFX}$ side-chain carboxylate from the reported AspH:Mn:NOG:hFX-EGFD1$_{86-124}$-4Ser structure adopts two conformations.[1]



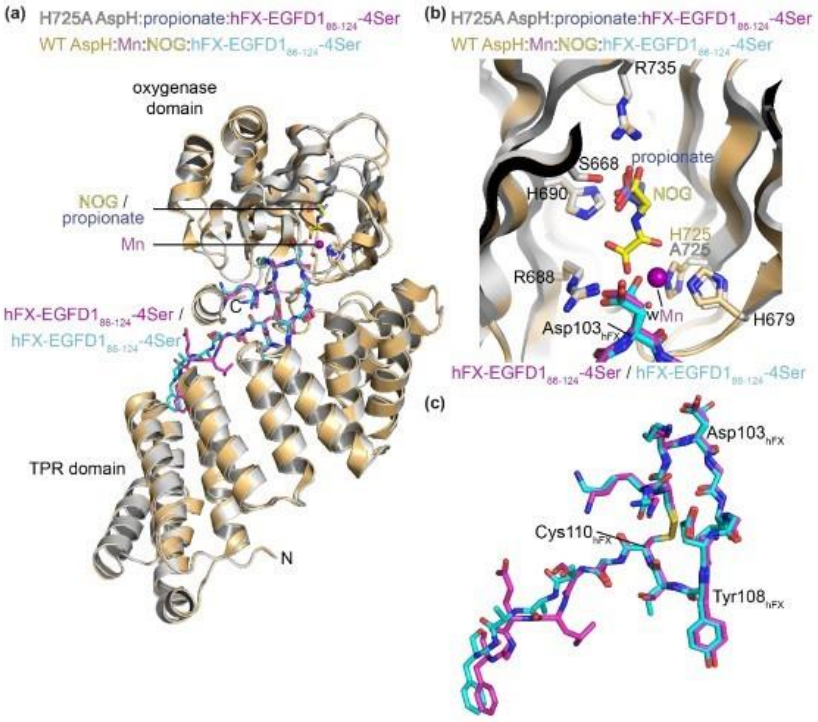
Supporting Figure S15. Views from a crystal structure of H725A AspH complexed with propionate at the active site and a synthetic EGFD substrate peptide (H725A AspH:hFX-EGFD1$_{86-124}$-4Ser; PDB ID: 7E6J). Colours: grey: H725A His$_6$-AspH$_{315-758}$; slate blue: carbon-backbone of propionate; magenta: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen.

(**a**) Overview of the H725A AspH:hFX-EGFD1$_{86-124}$-4Ser crystal structure (1.9 Å resolution). Note, electron density for a metal ion bound to the H725A AspH active site was not observed; (**b**) OMIT electron density map (mF_o-DF_c) contoured to 3σ around the hFX-EGFD1$_{86-124}$-4Ser peptide from the H725A AspH:hFX-EGFD1$_{86-124}$-4Ser structure reveals electron density for residues Lys100$_{hFX}$ to Thr111$_{hFX}$ including for the disulfide bridged (Cys101$_{hFX}$ and Cys110$_{hFX}$) ten-membered non-canonical EGFD macrocycle. However, electron density is not observed for the side chain of Tyr108$_{hFX}$, an EGFD residue which is part of the AspH-substrate consensus sequence requirement and which has been reported to interact with the TPR domain of wt AspH.[1]



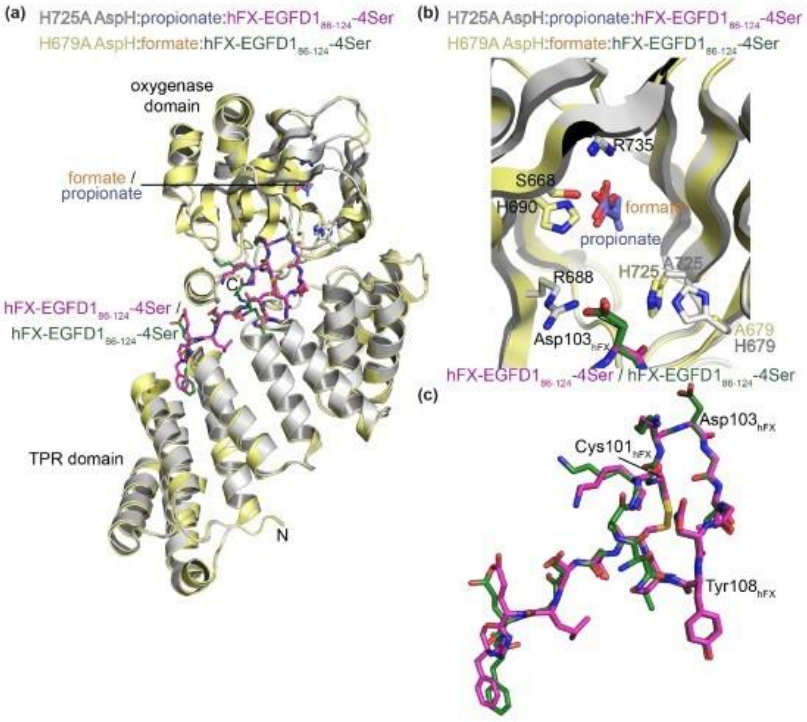
Supporting Figure S16. The substitution of H725 by alanine neither substantially alters the overall AspH fold nor the conformation of key active site residues in the presence of a substrate peptide. Colours: grey: H725A His_6-$AspH_{315-758}$; slate blue: carbon-backbone of propionate; magenta: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen. w: water.

(**a**) Superimposition of a view from the H725A AspH:hFX-$EGFD1_{86-124}$-4Ser structure (Supporting Figure S15) with one from the reported AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure (bronze: wt AspH; yellow: carbon-backbone of NOG; cyan: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide; violet: Mn; PDB ID: 5JQY)[1] reveals similar AspH conformations (Cα RMSD = 0.207 Å); (**b**) superimposition of a view from the active site of the H725A AspH:hFX-$EGFD1_{86-124}$-4Ser structure with one from the reported AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure (bronze: wt AspH; yellow: carbon-backbone of NOG; cyan: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide; violet: Mn; PDB ID: 5JQY)[1] reveals that the conformations of those wt AspH residues directly engaged in NOG/2OG binding (*i.e.* S668, R688, H690, and R735)[1,6] do not alter substantially upon substituting H725 by alanine, even though electron density for a metal ion is not observed. The conformation of H679, whose side chain imidazole engages in metal binding, does not alter substantially upon substituting H725 for alanine; (**c**) superimposition of the hFX-$EGFD1_{86-124}$-4Ser substrate peptide from the H725A AspH:hFX-$EGFD1_{86-124}$-4Ser structure with the one from the reported AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure (cyan: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide; PDB ID: 5JQY)[1] reveals similar peptide conformations (Cα RMSD = 0.266 Å), in particular of the amino acids forming the disulfide bridged ($Cys101_{hFX}$ and $Cys110_{hFX}$) ten-membered non-canonical EGFD macrocycle. Note that the $Asp103_{hFX}$ side-chain carboxylate from the reported AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure adopts two conformations.[1]



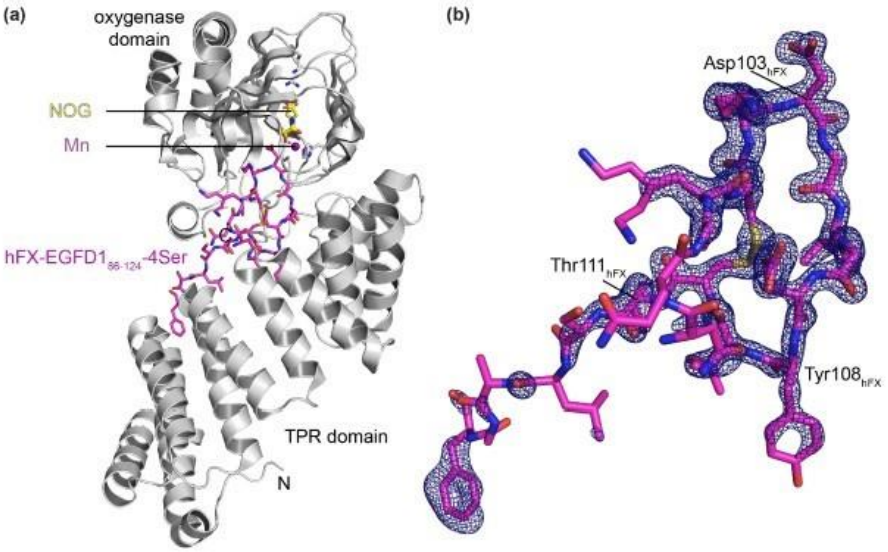
22

Supporting Figure S17. The overall folds and conformations of the H725 and H679A AspH variants are similar. Colours: grey: H725A His_6-$AspH_{315-758}$; slate blue: carbon-backbone of propionate; magenta: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen.
(**a**) Superimposition of a view from the H725A AspH:hFX-EGFD1$_{86-124}$-4Ser structure (Supporting Figure S15) with one from the H679A AspH:hFX-EGFD1$_{86-124}$-4Ser structure (pale yellow: H679A His_6-$AspH_{315-758}$; orange: carbon-backbone of formate; dark green: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide; Supporting Figure S12) reveals similar AspH conformations (Cα RMSD = 0.268 Å); (**b**) superimposition of a view from the active site of the H725A AspH:hFX-EGFD1$_{86-124}$-4Ser structure with one from the H679A AspH:hFX-EGFD1$_{86-124}$-4Ser structure (pale yellow: H679A His_6-$AspH_{315-758}$; orange: carbon-backbone of formate; dark green: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide; Supporting Figure S12) reveals that the conformations of those wt AspH residues directly engaged in NOG/2OG binding (*i.e.* S668, R688, H690, and R735)[1,6] do not alter substantially if either H679 or H725 were substituted for alanine, even though electron density for a metal ion is not observed; (**c**) superimposition of the hFX-EGFD1$_{86-124}$-4Ser substrate peptide from the H725A AspH:hFX-EGFD1$_{86-124}$-4Ser structure with the one from the H679A AspH:hFX-EGFD1$_{86-124}$-4Ser structure (dark green: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide; Supporting Figure S12) reveals similar peptide conformations (Cα RMSD = 0.076 Å), in particular of the amino acids forming the disulfide bridged ($Cys101_{hFX}$ and $Cys110_{hFX}$) ten-membered non-canonical EGFD macrocycle.



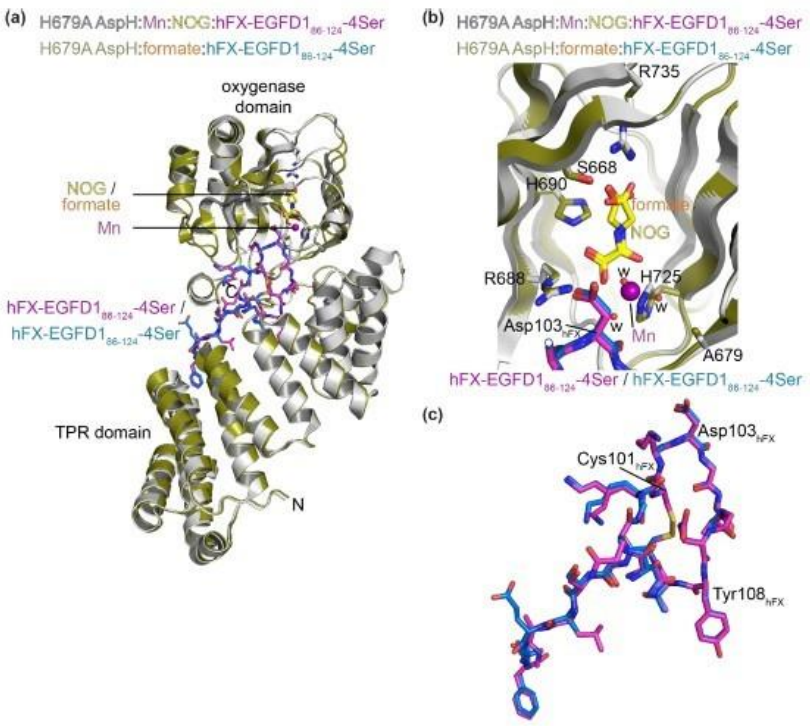
Supporting Figure S18. Views from a crystal structure of H679A AspH complexed with Mn, NOG, and a synthetic EGFD substrate peptide (H679A AspH:Mn:NOG:hFX-EGFD1$_{86-124}$-4Ser; PDB ID: 6Q9F). Colours: grey: H679A His$_6$-AspH$_{315-758}$; yellow: carbon-backbone of *N*-oxalylglycine (NOG); violet: Mn; magenta: carbon-backbone of the hFX-EGFD1$_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen.

(**a**) Overview of the H679A AspH:NOG:hFX-EGFD1$_{86-124}$-4Ser crystal structure (1.6 Å resolution); (**b**) OMIT electron density map (mF_o–DF_c) contoured to 3σ around the hFX-EGFD1$_{86-124}$-4Ser peptide from the H679A AspH:Mn:NOG:hFX-EGFD1$_{86-124}$-4Ser structure reveals electron density for residues Lys100$_{hFX}$ to Thr111$_{hFX}$ including for the disulfide bridged (Cys101$_{hFX}$ and Cys110$_{hFX}$) ten-membered non-canonical EGFD macrocycle. Incomplete electron density is observed for the side chain of Tyr108$_{hFX}$, an EGFD residue which is part of the AspH-substrate consensus sequence requirement. In reported AspH:substrate complex structures, the side chain of Tyr108$_{hFX}$ has been shown to interact with the TPR domain of AspH.[1]



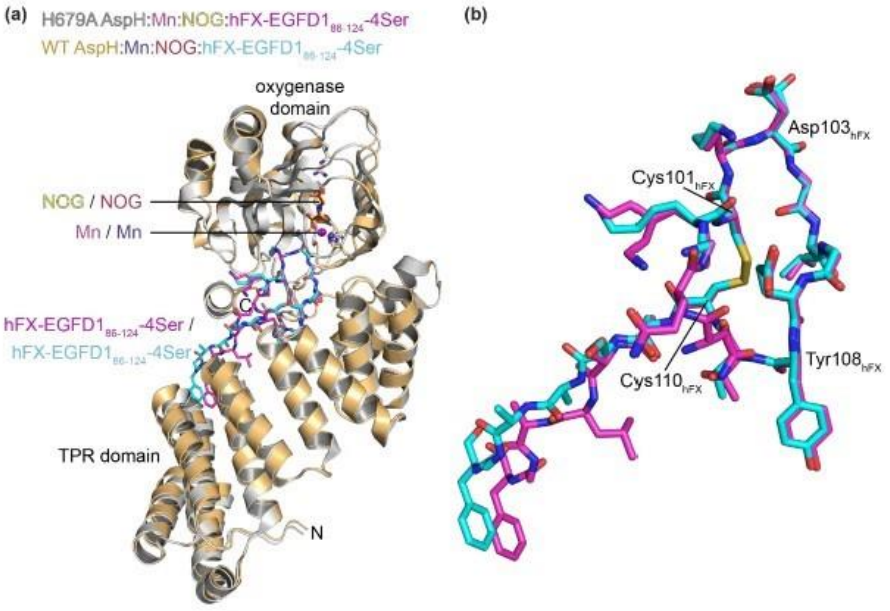
Supporting Figure S19. The substitution of formate in the active site of the H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser structure for NOG and Mn neither substantially alters the overall AspH fold nor the conformation of key active site residues. Colours: grey: H679A His$_6$-AspH$_{315\text{-}758}$; yellow: carbon-backbone of *N*-oxalylglycine (NOG); violet: Mn; magenta: carbon-backbone of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen. w: water.

(**a**) Superimposition of a view from the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure (Supporting Figure S18) with one from the H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser structure (olive: H679A AspH; orange: carbon-backbone of formate; marine blue: carbon-backbone of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide; Supporting Figure S12) reveals similar AspH conformations (Cα RMSD = 0.253 Å); (**b**) superimposition of a view from the active site of the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure with one from the H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser structure (olive: H679A AspH; orange: carbon-backbone of formate; marine blue: carbon-backbone of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide; Supporting Figure S12) reveals that the conformations of those wt AspH residues directly engaged in NOG/2OG binding (*i.e.* S668, R688, H690, and R735)[1,6] do not alter substantially upon substituting formate by NOG and Mn (note that the NOG glycine unit adopts two conformations, one of which superimposes with the formate in the H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser structure). The conformation of H725 does not alter substantially upon coordination of a Mn ion by the side chain imidazole. (**c**) superimposition of the hFX-EGFD1$_{86\text{-}124}$-4Ser substrate peptide from the H679A AspH:Mn:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure with the one from the H679A AspH:hFX-EGFD1$_{86\text{-}124}$-4Ser structure (marine blue: carbon-backbone of the hFX-EGFD1$_{86\text{-}124}$-4Ser peptide) reveals similar peptide conformations (Cα RMSD = 0.17 Å), in particular of the amino acids forming the disulfide bridged (Cys101$_{hFX}$ and Cys110$_{hFX}$) ten-membered non-canonical EGFD macrocycle.

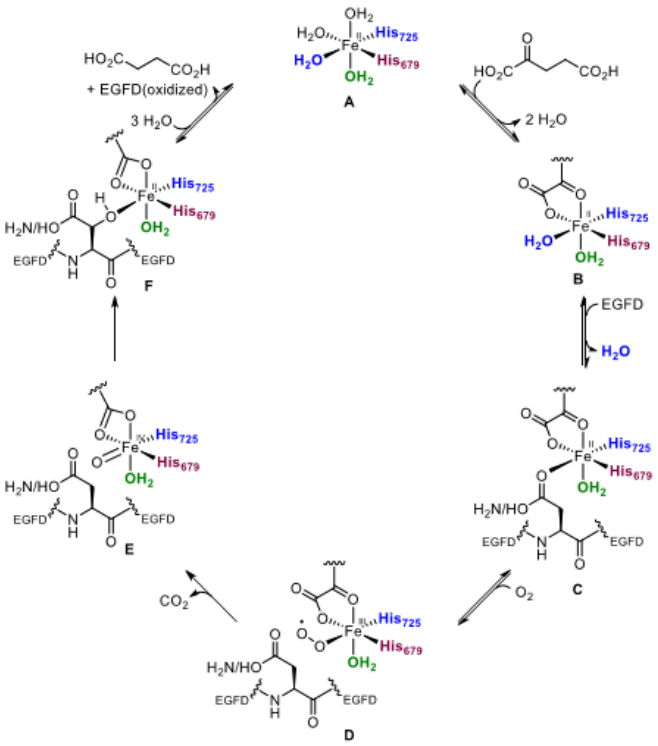


25

Supporting Figure S20. The substitution of H679 by alanine neither substantially alters the overall AspH fold nor the conformation of key active site residues in the presence of a substrate peptide. Colours: grey: H679A His_6-$AspH_{315-758}$; yellow: carbon-backbone of *N*-oxalylglycine (NOG); violet: Mn; magenta: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide (Supporting Figure S2a); red: oxygen; blue: nitrogen. w: water. (**a**) Superimposition of a view from the H679A AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure (Supporting Figure S18) with one from the reported wt AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure (bronze: wt AspH; maroon: carbon-backbone of NOG; cyan: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide; lavender blue: Mn; PDB ID: 5JQY)[1] reveals similar AspH conformations (Cα RMSD = 0.235 Å); (**b**) superimposition of the hFX-$EGFD1_{86-124}$-4Ser substrate peptide from the H679A AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure with the one from the reported wt AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure (cyan: carbon-backbone of the hFX-$EGFD1_{86-124}$-4Ser peptide; PDB ID: 5JQY)[1] reveals similar peptide conformations (Cα RMSD = 0.754 Å), in particular of the amino acids forming the disulfide bridged ($Cys101_{hFX}$ and $Cys110_{hFX}$) ten-membered non-canonical EGFD macrocycle. Note that the $Asp103_{hFX}$ side-chain carboxylate from the reported AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser structure adopts two conformations.[1]

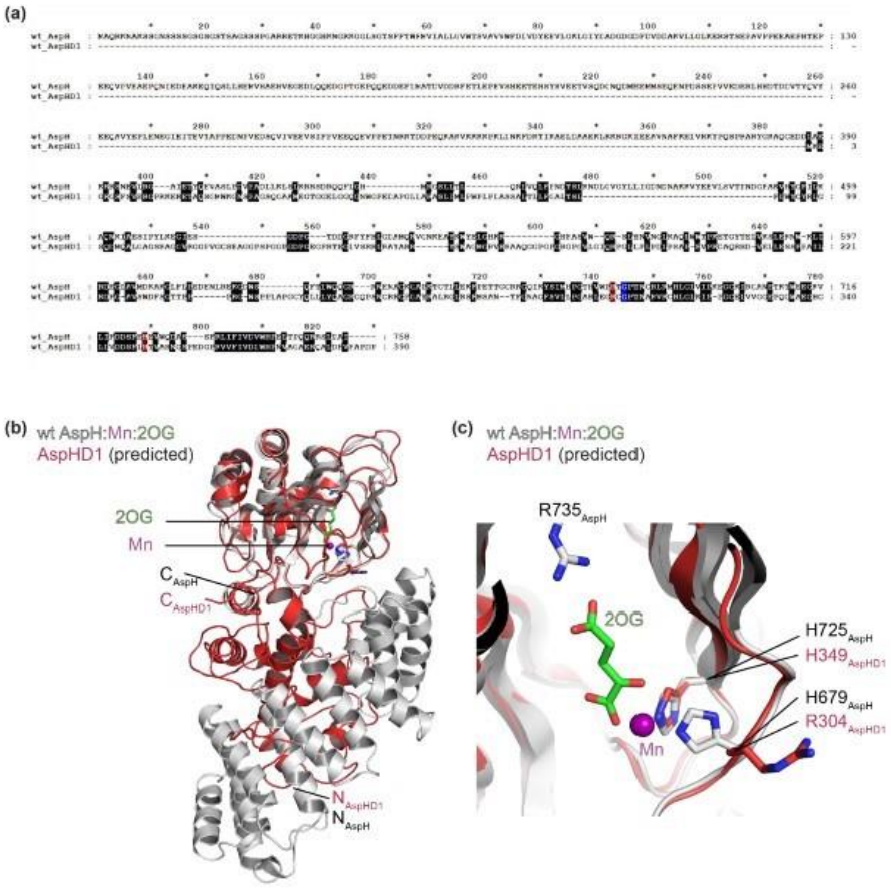


Supporting Figure S21. Proposed outline catalytic cycle for the AspH-catalysed hydroxylation of EGFD Asp- and Asn-residues. Based on reported wt AspH:2OG and wt AspH:2OG:substrate structures[6] and mechanistic studies with other 2OG oxygenases,[7] an outline catalytic cycle for AspH has been proposed.[6] Initially, 2OG coordinates to the AspH Fe(II) cofactor in a bidentate manner through one of its C1-carboxylate oxygen atoms and its C2-ketone oxygen atom (**B**), replacing two ligating water molecules in the Fe(II) complex **A**. The 2OG C1-carboxylate oxygen atom coordinates Fe(II) *trans* to His679 (red).[6] The EGFD substrate then binds AspH with the Asp/Asn-residue undergoing β-hydroxylation displacing the water molecule (blue) binding Fe(II) *trans* to His725 (blue, **C**);[6] however, in the H679A AspH:NOG:hFX-EGFD1$_{86\text{-}124}$-4Ser structure, three water molecules are still coordinated to the metal ion indicating the substrate binding occurs prior to water release. Interactions of the Asp/Asn side chain carboxylate/amide with the active site Fe(II) are likely weak, as several different conformations for the Asp/Asn side chains are observed in the crystalline state.[1,6] Molecular O_2 likely binds to Fe(II) *trans* to the His725 (blue) replacing the substrate side chain carboxylate to give the superoxo-Fe(III) complex **D**. Irreversible oxidative decarboxylation of 2OG occurs, potentially via a peroxyester intermediate,[7b] liberating CO_2 and forming an Fe(IV)-oxo:succinate:substrate complex (*i.e.* the ferryl intermediate, **E**). The *pro-R* hydrogen of the substrate EGFD Asn/Asp-side chain is proximate to the Fe(IV)-oxo moiety of complex **E** and is hydroxylated,[1,6] via direct insertion or a radical mechanism proceeding via an Fe(III) intermediate, to afford the Fe(II):succinate:hydroxylated substrate complex **F**. Finally, three water molecules replace the oxidized EGFD substrate and succinate to reform complex **A**. Note, that the role of the water-occupied coordination site *trans* to the 2OG ketone oxygen is not addressed in this mechanistic proposal; it is likely, this is important as it distinguishes AspH from other 2OG oxygenases;[7] thus, other mechanistic scenarios are feasible.



27

Supporting Figure S22. Human Aspartate β-hydroxylase domain-containing protein 1 is predicted to bear an unusual Fe(II) binding site composed of an Arg- and a His-residue. (**a**) Sequence alignment of human wt AspH (top) with human wt aspartate β-hydroxylase domain-containing protein 1 (AspHD1, bottom) suggests that AspHD1 bears an unusual Fe(II) binding site composed of an Arg- and a His-residue, *i.e.* R304 and H349. Sequences were obtained from the National Center for Biotechnology Information (NCBI) database[8] and processed using GeneDoc 2.7;[9] black shading indicates high sequence conservation; (predicted) metal binding residues are in red (*i.e.* H679 and H725 for wt AspH and R304 and H349 for wt AspHD1); Gly-residues substituting the typical Glu- or Asp-residues present in the Fe(II)-binding facial triad of 2OG hydroxylases are in blue; (**b** and **c**) superimposition of a view from the wt AspH:2OG structure (grey: wt AspH; green: carbon-backbone of 2OG; violet: Mn; PDB ID: 6YYU[6]) with one from the predicted structure of AspHD1 (red: wt AspHD1) which was calculated using the transform-restrained Rosetta software package.[10] Close-up (**c**) of a view from the predicted AspHD1 metal binding site reveals how the side chains of R304 and H349 may interact with the active site metal. Note that in the absence of metal, the side chain of R304 is predicted to orient away from the AspHD1 active site.



28

Supporting Table S1. Crystallization conditions, data collection, and refinement statistics for the H679A and H725A AspH variant complexes.[a)]

	H725A His_6-$AspH_{315-758}$·hFX-$EGFD1_{86-124}$-4Ser (H725A AspH:hFX-$EGFD1_{86-124}$-4Ser)	H679A His_6-$AspH_{315-758}$·Mn^{II}·NOG·hFX-$EGFD1_{86-124}$-4Ser (H679A AspH:Mn:NOG:hFX-$EGFD1_{86-124}$-4Ser)	H679A His_6-$AspH_{315-758}$·hFX-$EGFD1_{86-124}$-4Ser (H679A AspH:hFX-$EGFD1_{86-124}$-4Ser)	H679A His_6-$AspH_{315-758}$ (H679A AspH:acetate)
PDB ID	7E6J	6Q9F	6Q9I	6QA5
Crystallization				
Precipitation conditions	18 mg/mL His_6-H725A-$AspH_{315-758}$ (330 µM), 100 mM PCTP, pH 7.0, 25% w/v PEG1500, 726 µM hFX-$EGFD1_{86-124}$-4Ser	18 mg/mL His_6-H679A-$AspH_{315-758}$ (330 µM), 1 mM $MnCl_2$, 2 mM NOG, 0.1 M Tris-HCl, pH 8.0, 0.2 M sodium chloride, 20% w/v PEG6000, 726 µM hFX-$EGFD1_{86-124}$-4Ser	18 mg/mL His_6-H679A-$AspH_{315-758}$ (330 µM), 0.2 M sodium formate, pH 7.5, 20% w/v PEG3350, 726 µM hFX-$EGFD1_{86-124}$-4Ser	18 mg/mL His_6-H679A-$AspH_{315-758}$ (330 µM), 0.1 M PCTP, pH 9.0, 0.2 M sodium acetate, 25% w/v PEG1500
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Symmetry	orthorhombic	orthorhombic	orthorhombic	orthorhombic
Cell dimensions:				
a, *b*, *c* (Å)	50.24, 86.84, 123.54	49.81, 91.54, 122.90	50.13, 91.36, 123.54	48.78, 70.52, 173.50
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00	90.00, 90.00, 90.00
X-Ray source[b)]	Synchrotron (DLS I03)	Synchrotron (DLS I04)	Synchrotron (DLS I04)	Synchrotron (DLS I04)
Resolution (Å)[c)]	46.54-1.90 (1.93-1.90)	46.16-1.63 (1.69-1.63)	46.45-1.85 (1.92-1.85)	44.72-2.65 (2.74-2.65)
R_{merge}	0.153 (1.792)	0.061 (1.000)	0.077 (1.000)	0.241 (1.319)
I / σI	10.5 (0.8)	35.6 (1.9)	27.1 (1.6)	9.0 (2.0)
CC (1/2)	0.998 (0.699)	0.999 (0.891)	0.998 (0.705)	0.997 (0.608)
Total number unique reflections	43495 (2159)	70950 (3465)	49046 (2451)	18051 (884)
Completeness (%)	100.0 (99.3)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
Multiplicity	13.3 (13.7)	12.9 (13.0)	9.8 (9.5)	12.9 (13.3)
Refinement				
R_{work} / R_{free}	0.170 / 0.213	0.182 / 0.200	0.202 / 0.223	0.197 / 0.224
No. atoms:	3992	4010	3923	3514
B-factors (Å^2):	43.0	46.0	59.0	50.0
R.m.s. deviations:				
Bond lengths (Å)	0.002	0.012	0.007	0.005
Bond angles (°)	0.571	1.199	0.973	0.775

[a)]Experimental details are specified in the Supplementary Information (Section 3); [b)]DLS: Diamond Light Source; [c)]Values in parentheses are for highest-resolution shell.

2. AspH variant inhibition assays

Solutions of the inhibitors (100% DMSO) were dry dispensed across 384-well polypropylene assay plates (Greiner) in an approximately three-fold and 11-point dilution series (100 µM top concentration) using an ECHO 550 acoustic dispenser (Labcyte). DMSO and 2,4-PDCA were used as negative and positive inhibition controls, respectively. The final DMSO concentration was kept constant at 0.5%$_{v/v}$ throughout all experiments (using the DMSO backfill option of the acoustic dispenser). Each reaction was performed in technical duplicates in adjacent wells of the assay plates; additionally, assays were performed in two independent duplicates. Cosubstrate/cofactor stock solutions (L-ascorbic acid, LAA: 50 mM in Milli-Q® Ultrapure water; 2-oxoglutarate, 2OG: 10 mM in Milli-Q® Ultrapure water; ammonium iron(II) sulfate hexahydrate, FAS, $(NH_4)_2Fe(SO_4)_2 \cdot 6H_2O$: 400 mM in 20 mM HCl diluted to 1 mM in Milli-Q® Ultrapure water) were freshly prepared each day from commercially sourced solids (Sigma Aldrich).

The Enzyme Mixture (25 µL/well), containing H679A or H725A His$_6$-AspH$_{315-758}$ variant (0.2 µM) in 25 mM HEPES buffer (pH 7.5), was dispensed across the inhibitor-containing 384-well assay plates with a multidrop dispenser (ThermoFischer Scientific) at 20°C under an ambient atmosphere. The plates were subsequently centrifuged (1000 rpm, 10 s) and incubated for 15 minutes at 20°C. The Substrate Mixture (25 µL/well), containing hFX-EGFD1$_{86-124}$-4Ser peptide (8.0 µM; Supporting Figure 2a), LAA (200 µM), 2OG (for H679A: 220 µM; for H725A: 430 µM), and FAS (for H679A: 8.0 µM; for H725A: 20 µM) in 25 mM HEPES buffer (pH 7.5), was added using the multidrop dispenser. The plates were centrifuged (1000 rpm, 10 s) and after incubating (22 min for H679A; 37 min for H725A), the enzyme reaction was stopped by addition of 10%$_{v/v}$ aqueous formic acid (5 µL/well). The plates were then centrifuged (1000 rpm, 20 s) and analysed by MS.

MS-analyses were performed using a RapidFire RF 365 high-throughput sampling robot (Agilent) attached to an iFunnel Agilent 6550 accurate mass quadrupole time-of-flight (Q-TOF) mass spectrometer operated in the positive ionization mode. Assay samples were aspirated under vacuum for 0.6 s and loaded onto a C4 solid phase extraction (SPE) cartridge. After loading, the C4 SPE cartridge was washed with 0.1%$_{v/v}$ aqueous formic acid to remove non-volatile buffer salts (5.5 s, 1.5 mL/min). The peptide was eluted from the SPE cartridge with 0.1%$_{v/v}$ aqueous formic acid in 80/20$_{v/v}$ acetonitrile/water into the mass spectrometer (5.5 s, 1.5 mL/min) and the SPE cartridge re-equilibrated with 0.1%$_{v/v}$ aqueous formic acid (0.5 s, 1.25 mL/min). The mass spectrometer parameters were: capillary voltage (4000 V), nozzle voltage (1000 V), fragmentor voltage (365 V), gas temperature (280°C), gas flow (13 L/min), sheath gas temperature (350°C), sheath gas flow (12 L/min). For data analysis, the m/z +4 charge states of the hFX-EGFD1$_{86-124}$-4Ser peptide (substrate) and the hydroxylated hFX-EGFD1$_{86-124}$-4Ser peptide (product) were used to extract ion chromatogram data; peak areas were integrated using RapidFire Integrator software (Agilent). Data were exported into Microsoft Excel and used to calculate the % conversion of the hydroxylation reaction using the equation: % conversion = 100 x (integral hydroxylated hFX-EGFD1$_{86-124}$-4Ser peptide) / (integral substrate hFX-EGFD1$_{86-124}$-4Ser peptide + integral hydroxylated hFX-EGFD1$_{86-124}$-4Ser peptide). Normalized dose-response curves (2,4-PDCA and DMSO controls) were obtained from the raw data by non-linear regression (GraphPad Prism 5) and used to determine IC$_{50}$-values. The standard deviation (SD) of two independent IC$_{50}$ determinations (n = 2) was calculated using GraphPad Prism 5. Z'-factors were calculated according to the literature using Microsoft Excel.[5]

3. Crystallography

High-throughput crystallization experiments were performed in 96-well, 3-subwell, low profile Intelliplates (Art Robbins Instruments) using a Phoenix RE liquid dispensing robot (Art Robbins Instruments) and Hampton Research (PEG/Ion, Crystal Screen) or Molecular Dimensions (PACT Premier) crystallization screens. *N*-Terminally His_6-tagged $AspH_{315-758}$ variants (18 mg/mL in 50 mM HEPES buffer, pH 7.5) were mixed with 1 mM $MnCl_2$, 2 mM 2OG or NOG, and, where appropriate, the hFX-EGFD1$_{86-124}$-4Ser peptide as an AspH substrate (note that $MnCl_2$ and 2OG/NOG were present all crystallization experiments, including those for which no metal/ligand-bound structure was obtained). Crystals were grown using the vapour diffusion method at 4°C in 200 nL or 300 nL sitting drops with 2:1, 1:1 or 1:2 sample:well solution ratios; precipitants are listed in the Supporting Table S1. Crystals were cryo-protected using mother liquor supplemented with 25%$_{v/v}$ glycerol before cryo-cooling in liquid N_2. Data were collected at 100 K using synchrotron radiation at Diamond Light Source (DLS) beamlines I03 and I04. For the H679A AspH structures, data were indexed, integrated, and scaled using the HKL2000 program suite;[11] for the H725A AspH:hFX-EGFD1$_{86-124}$-4Ser structure, data were indexed, integrated, and scaled using the Xia2[12] strategy of the beamline auto-processing pipeline (Supporting Table S1). The AspH crystal structures were determined by molecular replacement (MR) using the AutoMR (PHASER[13]) subroutine in PHENIX[14]. The search model used for MR was based on PDB ID 5JTC[15] for the H725A AspH:hFX-EGFD1$_{86-124}$-4Ser structure, PDB ID 5JQY[1] for the H679A AspH:hFX-EGFD1$_{86-124}$-4Ser structure, and PDB ID 6Q9F for all other H679A AspH structures. The structural model was improved by iterative cycles of manual re-building in COOT[16] and crystallographic refinement in phenix.refine[17] (refinement details are summarized in Supporting Table S1).

Crystal structure data for the *N*-terminal His_6-tagged $AspH_{315-758}$ variants H679A and H725A complexed to carboxylic acids or NOG, and, in some cases, substrate peptide (hFX-EGFD1$_{86-124}$-4Ser) and Mn, are deposited in the protein data bank with PDB accession codes: 7E6J (H725A AspH:hFX-EGFD1$_{86-124}$-4Ser), 6Q9F (H679A AspH:Mn:NOG:hFX-EGFD$_{86-124}$-4Ser), 6Q9I (H679A AspH:hFX-EGFD$_{86-124}$-4Ser), and 6QA5 (H679A AspH:acetate). PyMOL[18] was used for the generation of graphical representations; polder omit maps were calculated using Polder Maps[19] in PHENIX[14].

4. References

[1] I. Pfeffer, L. Brewitz, T. Krojer, S. A. Jensen, G. T. Kochan, N. J. Kershaw, K. S. Hewitson, L. A. McNeill, H. Kramer, M. Münzel, R. J. Hopkinson, U. Oppermann, P. A. Handford, M. A. McDonough, C. J. Schofield, *Nat. Commun.* **2019**, *10*, 4910.

[2] a) K. F. Geoghegan, H. B. F. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, A. S. McColl, A. V. Kamath, J. G. Stroh, *Anal. Chem.* **1999**, *267*, 169-184; b) Z. Yan, G. W. Caldwell, P. A. McDonell, *Biochem. Biophys. Res. Commun.* **1999**, *262*, 793-800.

[3] a) P. Fernlund, J. Stenflo, *J. Biol. Chem.* **1983**, *258*, 12509-12512; b) B. A. McMullen, K. Fujikawa, W. Kisiel, T. Sasagawa, W. N. Howald, E. Y. Kwa, B. Weinstein, *Biochemistry* **1983**, *22*, 2875-2884.

[4] L. Brewitz, A. Tumber, C. J. Schofield, *J. Biol. Chem.* **2020**, *295*, 7826-7838.

[5] J.-H. Zhang, T. D. Y. Chung, K. R. Oldenburg, *J. Biomol. Screen.* **1999**, *4*, 67-73.

[6] L. Brewitz, Y. Nakashima, C. J. Schofield, *Chem. Sci.* **2021**, *12*, 1327-1342.

[7] a) H. M. Hanauske-Abel, V. Günzler, *J. Theor. Biol.* **1982**, *94*, 421-455; b) S. Martinez, R. P. Hausinger, *J. Biol. Chem.* **2015**, *290*, 20702-20711.

[8] L. Y. Geer, A. Marchler-Bauer, R. C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, S. H. Bryant, *Nucleic Acids Res.* **2010**, *38*, D492-D496.

[9] K. B. Nicholas, H. B. Nicholas Jr., D. W. Deerfield II., *embnet.news* **1997**, *4*, 1-4.
[10] J. Yang, I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov, D. Baker, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 1496-1503.
[11] Z. Otwinowski, W. Minor, in *Methods Enzymol., Vol. 276*, Academic Press, **1997**, pp. 307-326.
[12] G. Winter, *J. Appl. Cryst.* **2010**, *43*, 186-190.
[13] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, R. J. Read, *J. Appl. Cryst.* **2007**, *40*, 658-674.
[14] P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, P. H. Zwart, *Acta Cryst. D* **2010**, *66*, 213-221.
[15] L. Brewitz, A. Tumber, I. Pfeffer, M. A. McDonough, C. J. Schofield, *Sci. Rep.* **2020**, *10*, 8650.
[16] P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, *Acta Cryst. D* **2010**, *66*, 486-501.
[17] P. V. Afonine, R. W. Grosse-Kunstleve, N. Echols, J. J. Headd, N. W. Moriarty, M. Mustyakimov, T. C. Terwilliger, A. Urzhumtsev, P. H. Zwart, P. D. Adams, *Acta Cryst. D* **2012**, *68*, 352-367.
[18] W. L. DeLano, **2002**, De Lano Scientific, San Carlos.
[19] D. Liebschner, P. V. Afonine, N. W. Moriarty, B. K. Poon, O. V. Sobolev, T. C. Terwilliger, P. D. Adams, *Acta Cryst. D* **2017**, *73*, 148-157.