



Crystallographic Studies on 2-Oxoglutarate Dependent Oxygenases

Wei Shen Aik

Department of Chemistry

Wolfson College

Trinity Term 2014

D.Phil. Thesis

Abstract

The Fe(II) and 2-oxoglutarate dependent oxygenases (2OG oxygenases) catalyse a broad range of oxidative reactions in various organisms. 2OG oxygenases use 2OG and molecular oxygen to catalyse the oxidation of a variety of substrates including small molecules, fatty acids, nucleic acids and proteins. Several human 2OG oxygenases are implicated in diseases. The fat mass and obesity associated protein (FTO) is linked to obesity, whilst collagen prolyl-4-hydroxylases (CPHs), the procollagen lysyl hydroxylases (PLODs), and the hypoxia inducible factor hydroxylases (PHDs) are linked to cancer. Therefore, structure-based inhibition studies on FTO, CPH and related 2OG oxygenases are of significant therapeutic interest.

The obesity-associated FTO is an mRNA *N*⁶-methyladenine (m⁶A) demethylase and a homologue of *E. coli* alkylated DNA repair protein (AlkB), a DNA repair enzyme that demethylates *N*¹-methyladenine (m¹A) and *N*³-methylcytosine (m³C) bases. Human AlkB homologue 5 (ALKBH5) has a similar substrate profile as FTO, i.e. ALKBH5 is another mRNA m⁶A demethylase, making it a target for structural and inhibition studies to improve inhibitor selectivity for FTO. The CPH and PLODs catalyse hydroxylation of collagen prolyl- and lysyl-residues; the resultant 4-hydroxyprolyl- and 4-hydroxylysyl-residues are important for the formation of the collagen triple helix and intermolecular crosslinks, respectively. Another 2OG oxygenase, the 2-oxoglutarate and iron-dependent oxygenase domain-containing protein 2 (OGFOD2), is related by sequence similarity to the PLODs but its biological role is unknown.

This thesis describes crystallographic, biochemical and inhibition studies on five 2OG oxygenases: AlkB, FTO, ALKBH5, OGFOD2 and CPH. Three AlkB-inhibitor complexes were determined; taken together, these structures serve as proof-of-principle that the AlkB subfamily 2OG oxygenases are amenable to structure-based inhibition studies. Several classes of inhibitors of FTO were then identified and their binding modes were investigated through the determination of 7 crystal structures of FTO-inhibitor complexes. Subsequently, a crystal structure of ALKBH5 was determined, which provided insights into its substrate recognition mechanisms. The ALKBH5 structure also serves as a template for inhibitor design. Preliminary structural and biochemical data were also obtained for OGFOD2 and CPH.

Overall, these results contribute to the development of a biophysical understanding of human 2OG oxygenases, and will help to enable the development of selective inhibitors of 2OG oxygenases involved in nucleic acid and collagen modifications.

Acknowledgements

First and foremost, I would like to thank my supervisor, Prof. Christopher Schofield, for giving me the opportunity to conduct this piece of research and for his dedicated supervision. I would also like to express my sincerest gratitude to Mr and Mrs Samuel Yun Ping Lin for financing my D.Phil. studies. Special thanks are also given to my crystallography mentor, Dr Michael McDonough, who has been selflessly imparting to me the necessary skills and knowledge in crystallography; to my cell biology mentor Dr Alexander Wolf, who introduced me to essential cloning and cell biology techniques; and to my biochemistry mentors, Dr Eleanor Bagg and Dr Zhihong Zhang, who mentored me in important protein purification and biochemistry techniques.

I would also like to thank Dr Marina Demetriades for her contributions in the RNA demethylase inhibition projects; Muhammad Khairul Kamilin Hamdan for his contributions in the FTO inhibition assay development; John Scotti for being a great research partner in the CPH project and for his help and support in the ALKBH5 project; Marcus Sims and Dr Luc Henry for their help in organic synthesis; Dr Martin Münzel, Dr Sarah Wilkins and Dr Louise Walport for helpful advice; Dr Kar Kheng Yeoh, Dr Clarisse Lejeune, Jacob Bush, Caitlin Clunie-O'Connor, Cyrille Thinnes and Dr Esther Woon for providing chemical compounds used in this research; Prof. Mukram Mackeen and Dr Lingzhi Gong for their help in mass spectrometry; Dr Hwanho Choi for his contributions in computational modeling; Dr Edward Lowe for collecting SAXS data for CPH; Dr Adam Hardy, Wendy Sobey and Manso Jubier for laboratory and IT support; Dr Ian Clifton and Dr Rasheduzzaman Chowdhury for advice on crystallography; and staff at the beamlines of Diamond Light Source synchrotron for their help in X-ray data collection.

Additionally, I would like to thank Prof. Christopher Schofield, Dr Michael McDonough, Dr Joan Gannon, Dr Sarah Wilkins, Dr Louise Walport, Dr Richard Hopkinson, Dr Ivanhoe Leung, and Sheila Lin for their help in proofreading and their constructive feedback for this thesis. Last but not least, I would like to thank my wife Sheila and my parents for their love and support throughout the process.

Table of Contents

Abstract	i
Acknowledgements	ii
Abbreviations	vii
Chapter 1	1
General Introduction.....	1
1.1 Structural Overview of 2OG Oxygenases	3
1.1.1 The Double-Stranded Beta Helix Fold	3
1.1.2 Structural Characteristics of 2OG Oxygenases	3
1.1.3 Active Site Metal Binding	7
1.1.4 2-Oxoglutarate Binding Site.....	8
1.1.5 Proposed General Catalytic Mechanisms of 2OG Oxygenases	11
1.2 Nucleic Acid Modifying Oxygenases	15
1.2.1 Discovery of the Function of AlkB	18
1.2.2 Human Homologues of AlkB.....	20
1.2.3 TETs	33
1.2.4 TYW5	36
1.3 2OG Oxygenases Involved in Collagen Biosynthesis.....	37
1.3.1 Collagen Prolyl-4-Hydroxylase	38
1.3.2 Procollagen Lysyl Hydroxylases	43
1.3.3 Collagen Prolyl-3-Hydroxylase	45
1.4 Objectives of the Research Described in this Thesis.....	45
1.5 References	47
Chapter 2.....	60
Proof-of-Principle Studies on Structure-based Inhibition of NAOX using AlkB as a Model Enzyme	60
2.1 Introduction	60
2.2 Results and Discussions	62
2.2.1 Inhibition of AlkB by broad-spectrum 2OG oxygenase inhibitors.....	63
2.2.2 Crystal structure of AlkB-IOX1 complex.....	65
2.2.3 Development of AlkB selective inhibitors.....	72
2.2.4 Dynamic Combinatorial Mass Spectrometry	73
2.2.5 Structure-guided inhibitor design.....	79

2.2.6 Crystal structure of the AlkB-18 complex	79
2.2.7 Crystal structure of the AlkB-22 complex	82
2.3 Conclusions	86
2.4 Experimental Procedures	86
2.5 References	89
Chapter 3	92
Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO)	92
3.1 Introduction	92
3.2 Results and Discussions	96
3.2.1 Expression and purification of full-length and truncated ($\Delta 31$) FTO	96
3.2.2 Investigating the enzymatic activities of recombinant hFTO and hFTO $\Delta 31$	100
3.2.3 Optimisation of in vitro demethylation assay for inhibition studies	102
3.2.4 Inhibition studies of FTO	105
3.2.5 Initial Crystallisation Attempts of hFTO $\Delta 31$	109
3.2.6 Recombinant FTO Co-purifies with Nucleic Acids	111
3.2.7 Initial Structural Analysis of FTO	115
3.2.8 Additive and Detergent Screens	117
3.2.9 Co-crystallisation of FTO with 2,4-Pyridine Dicarboxylate	120
3.2.10 Soaking FTO Crystals Grown in NOG with 2,4-PDCA and FG-2216	121
3.2.11 Active site metal substitutes	123
3.2.12 Inhibitor Concentration Optimisation for Crystallographic Studies	123
3.2.13 Crystallographic Data Collection	124
3.2.14 Data Processing and Structure Solution	124
3.2.15 Structure Refinement	127
3.2.16 A Crystal Structure of hFTO $\Delta 31$ in Complex with NOG	129
3.2.17 Crystal Structure of FTO in Complex with 2,4-PDCA	131
3.2.18 Crystal Structure of FTO in Complex with 5-Carboxy-8-hydroxyquinoline	132
3.2.19 Crystal Structure of FTO in Complex with Compound 6	135
3.2.20 Crystal Structure of FTO in Complex with FG-2216	139
3.2.21 Crystal Structure of FTO in Complex with Rhein and Citrate	140
3.2.22 Second Generation FTO Inhibitors	143
3.3 Conclusions	148
3.4 Experimental Procedures	149
3.5 References	157
Chapter 4	162
Crystal structure of human RNA N6-methyladenine demethylase ALKBH5	162

4.1 Introduction	162
4.2 Results and Discussions	164
4.2.1 Protein expression and purification	164
4.2.2 Crystallisation of ALKBH5 ₆₆₋₂₉₂	165
4.2.3 Data Collection and Processing	167
4.2.4 Structure Solution.....	171
4.2.5 Overall ALKBH5 Structure	173
4.2.6 N-Terminal Region of ALKBH5	175
4.2.7 C-Terminal Region of ALKBH5	176
4.2.8 Beta IV-V loop.....	176
4.2.9 Cys200-IOX3 modification	177
4.2.10 Conserved catalytic triad and overall active site architecture	182
4.2.11 2OG binding site	183
4.2.12 Residues involved in substrate binding	186
4.2.13 Substrate Recognition Elements.....	187
4.2.14 Modelling of substrate RNA binding.....	193
4.2.15 In vitro activity of ALKBH5 ₆₆₋₂₉₂	196
4.2.16 End-point inhibition assays	197
4.2.17 Development of ALKBH5 ₆₆₋₂₉₂ inhibition assays.....	198
4.2.18 IOX3 pre-incubation with ALKBH5.....	203
4.3 Conclusions	203
4.4 Experimental Procedures.....	204
4.5 References.....	209
Chapter 5	214
Preliminary Characterisation of OGFOD2	214
5.1 Introduction	214
5.2 Results and Discussions	215
5.2.1 Cellular Localisation Studies.....	216
5.2.2 Attempted substrate identification in cells and in vitro	221
5.2.3 Expression, purification and biophysical studies	222
5.2.4 OGFOD2 catalyses substrate uncoupled 2OG turnover.....	225
5.2.5 Crystallographic studies	226
5.3 Conclusions	230
5.4 Experimental Procedures.....	231
5.5 References.....	243
Chapter 6	246

Structural and biochemical studies on a human collagen prolyl-4-hydroxylase	246
6.1 Introduction	246
6.2 Results and Discussion	247
6.2.1 Expression and Purification	247
6.2.2 In vitro hydroxylation of collagen peptides	251
6.2.3 Dynamic Scanning Fluorometry	256
6.2.4 Crystallographic studies	257
6.2.5 Small Angle X-ray Scattering (SAXS) Studies	263
6.3 Conclusions	266
6.4 Experimental Procedures	267
6.5 References	271
Chapter 7	276
Conclusions	276
7.1 References	281
Appendix	284

Abbreviations

2,4-PDCA	2,4-pyridine dicarboxylic acid
2,5-PDCA	2,5-pyridine dicarboxylic acid
2HG	<i>L</i> -2-hydroxyglutarate
2OG	2-oxoglutarate
3meT	<i>N</i> ³ -methylthymidine deoxyribonucleoside
4C8HQ	4-carboxyl-8-hydroxyquinoline
5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
6meA	<i>N</i> ⁶ -methyladenosine deoxyribonucleoside
ADP	atomic displacement parameters
ALKBH	AlkB homologue
ACCO	1-aminocyclopropane-1-carboxylic acid oxidase
ASCC3	activating signal cointegrator 1 complex subunit 3
AlkB	Alkylated DNA repair protein
AP	apurine/apyrimidine
AspH	aspartyl/asparaginyl beta-hydroxylase
AtsK	alkylsulfatase
BaCPH	<i>Bacillus anthracis</i> collagen prolyl hydroxylase-like
BLAST	basic local alignment search tool
BMI	body mass index
bp	base pairs
BSA	bovine serum albumin
CAS	clavaminic acid synthase
CAT	catalytic domain
CD	circular dichroism
cm ⁵ U	5-carboxymethyluridine
co-IP	co-immunoprecipitation
CPH	collagen prolyl hydroxylase
CrCPH	<i>Chlamydomonas reinhardtii</i> collagen prolyl hydroxylase-like
Δ <i>T</i> _m	melting temperature shift
DAOCS	deacetoxycephalosporin C synthase
DAPI	4',6'-diamidino-2-phenylindole
DCL	dynamic combinatorial library
DCMS	dynamic combinatorial mass spectrometry
DMEM	Dulbecco's Modified Eagle Medium
DMSO	dimethylsulfoxide
DNOC	<i>N</i> -oxalyl- <i>D</i> -cysteine
DSBH	double-stranded beta-helix
dsDNA	double-stranded deoxyribonucleic acid
DSF	dynamic scanning fluorometry
DTT	dithiothreitol
εA	1, <i>N</i> ⁶ -ethenoadenine
εC	1, <i>N</i> ⁴ -ethenocytosine
EA	1, <i>N</i> ⁶ -ethanoadenine
EDTA	ethylenediaminetetraacetic acid

emPAI	exponentially modified Protein Abundance Index
EPR	electron paramagnetic resonance
ER	endoplasmic reticulum
ESI-MS	electrospray ionization-mass spectrometry
f ⁶ A	N ⁶ -formyladenine
FBS	fetal bovine serum
FDH	formaldehyde dehydrogenase
FIH	factor inhibiting HIF/HIF asparaginyl hydroxylase
FITC	fluorescein isothiocyanate
FPLC	fast protein liquid chromatography
FTO	human fat mass and obesity associated protein
<i>FTO</i>	human fat mass and obesity associated gene
Fto	mouse fat mass and obesity associated protein
<i>Fto</i>	mouse fat mass and obesity associated gene
GFP	green fluorescent protein
GWAS	genome-wide association studies
GS	glycine-serine
GGT	glucosyltransferase
GT	galactosyltransferase
HEK293T	human embryonic kidney 293T
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid
HIF	hypoxia inducible factor
hm ⁶ A	N ⁶ -hydroxymethyladenine
HPLC	high-performance liquid chromatography
HppE	hydroxypropylphosphonic acid epoxidase
hTET2	human TET2
I3C	5-amino-2,4,6-triiodosophthalic acid
IC ₅₀	half maximal response inhibitor concentration
IOX1	5-carboxy-8-hydroxyquinoline
IOX3	(1-chloro-4-hydroxyisoquinoline-3-carbonyl)glycine
IPNS	isopenicillin N synthase
IPTG	isopropyl β- <i>D</i> -1-thiogalactopyranoside
IRX	iroquois-class homeodomain protein
JBP	J base (β- <i>D</i> -glucosyl-HOMedU) binding protein
JmjC	jumonji domain-containing
JMJD2A	jumonji domain-containing protein 2A
JMJD3	jumonji domain-containing protein 3
LC	liquid chromatography
LLG	log-likelihood gain
LNOC	N-oxalyl- <i>L</i> -cysteine
m ¹ A	N ¹ -methyladenine
m ¹ G	N ¹ -methylguanine
m ³ C	N ³ -methylcytosine
m ³ T	N ³ -methylthymine
m ³ U	N ³ -methyluracil
m ⁶ A	N ⁶ -methyladenine
MAD	multiwavelength anomalous dispersion
MALDI	matrix-assisted laser desorption ionisation
MCD	magnetic circular dichroism

mchm ⁵ U	5-(methoxycarbonylhydroxymethyl)uracil
mcm ⁵ U	5-methoxycarbonylmethyluridine
MEF	mouse embryonic fibroblast
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MGN3-1	mouse ghrelinoma 3-1
MMS	methyl methane sulfonate
MPD	2-methyl-2,4-pentanediol
MS	mass spectrometry
MS/MS	fragmentation mass spectrometry
MT	methyltransferase
MW	molecular weight
NAD ⁺	oxidised nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NAOX	nucleic acid oxygenases
NgTET1	<i>Naegleria gruberi</i> TET1
NMR	nuclear magnetic resonance
NOG	<i>N</i> -oxalylglycine
NRL	nucleotide recognition lid
OGFOD2	20G and iron-dependent oxygenase domain-containing 2 protein
P3H	proline/prolyl-3-hydroxylase
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PDI	protein disulphide isomerase
PEG	polyethylene glycol
PEI	polyethylenimine
PHD	prolyl-4-hydroxylase domain containing
PHF8	plant homeobox domain finger 8
pI	isoelectric point
PLOD/LH	procollagen lysyl hydroxylase
PPHD	<i>Pseudomonas putida</i> prolyl-4-hydroxylase
PSB	peptide-substrate-binding domain
RMSD	root mean square deviation
RRM	RNA recognition motif
S200	Superdex 200 size exclusion chromatography
S75	Superdex 75 size exclusion chromatography
SAD	single wavelength anomalous dispersion
SART1	U4/U6.U5 tri-snRNP-associated protein 1
SAXS	small-angle X-ray scattering
SC35	serine/arginine-rich splicing factor 2
SDS	sodium dodecyl sulfate
SeMet	selenomethionine
siRNA	small interference ribonucleic acid
SMUG1	uracil-DNA glycosylase
S _N AR	aromatic nucleophilic substitution
SNP	single nucleotide polymorphism
SPA	single point inhibition assay
ssDNA	single-stranded deoxyribonucleic acid
ssRNA	single-stranded ribonucleic acid

SyrB2	syringomycin biosynthesis enzyme 2
TauD	taurine dioxygenase
TCA	tricarboxylic acid
TDG	thymine DNA glycosylase
TET	ten-eleven-translocation
TFE	trifluoroethanol
TFZ	translation function Z score
TLS	translation, libration, screw
ToF	time-of-flight
TPR	tetratricopeptide repeat
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
TYW5	tRNA wybutosine-synthesizing protein 5
UPLC	ultra-performance liquid chromatography
UTX	histone demethylase UTX/lysine-specific demethylase 6B
UV	ultraviolet
VTVH	variable-temperature, variable-field
YTHDF	YTH domain-containing family
yW-72	7-(α -amino- α -carboxypropyl)wyosine

Chapter 1

General Introduction

The Fe(II) and 2-oxoglutarate-dependent oxygenases (2OG oxygenases) are distributed widely in eubacteria and eukaryotes. The 2OG oxygenases catalyse a range of oxidative reactions including desaturation, hydroxylation, epimerisation, and cyclisation (Figure 1 - 1) (1-5). 2OG oxygenases use 2OG and molecular oxygen as cosubstrates and Fe(II) as a cofactor to enable substrate oxidation (1-5). Their substrates range from macromolecules such as nucleic acids and proteins to small molecules including secondary metabolites and antibiotics (1,6).

Structural studies on 2OG oxygenases have revealed insights into their substrate recognition elements, which are often characteristic of different subfamilies. To date, crystal structures of more than 60 different 2OG oxygenases have been determined; these structures have significantly advanced our knowledge of the structure-function relationships and evolutionary paths of 2OG oxygenases. Because many 2OG oxygenases are involved in diseases such as cancer, obesity, and anaemia, some of the structures have been useful for 2OG oxygenase inhibitor development (7).

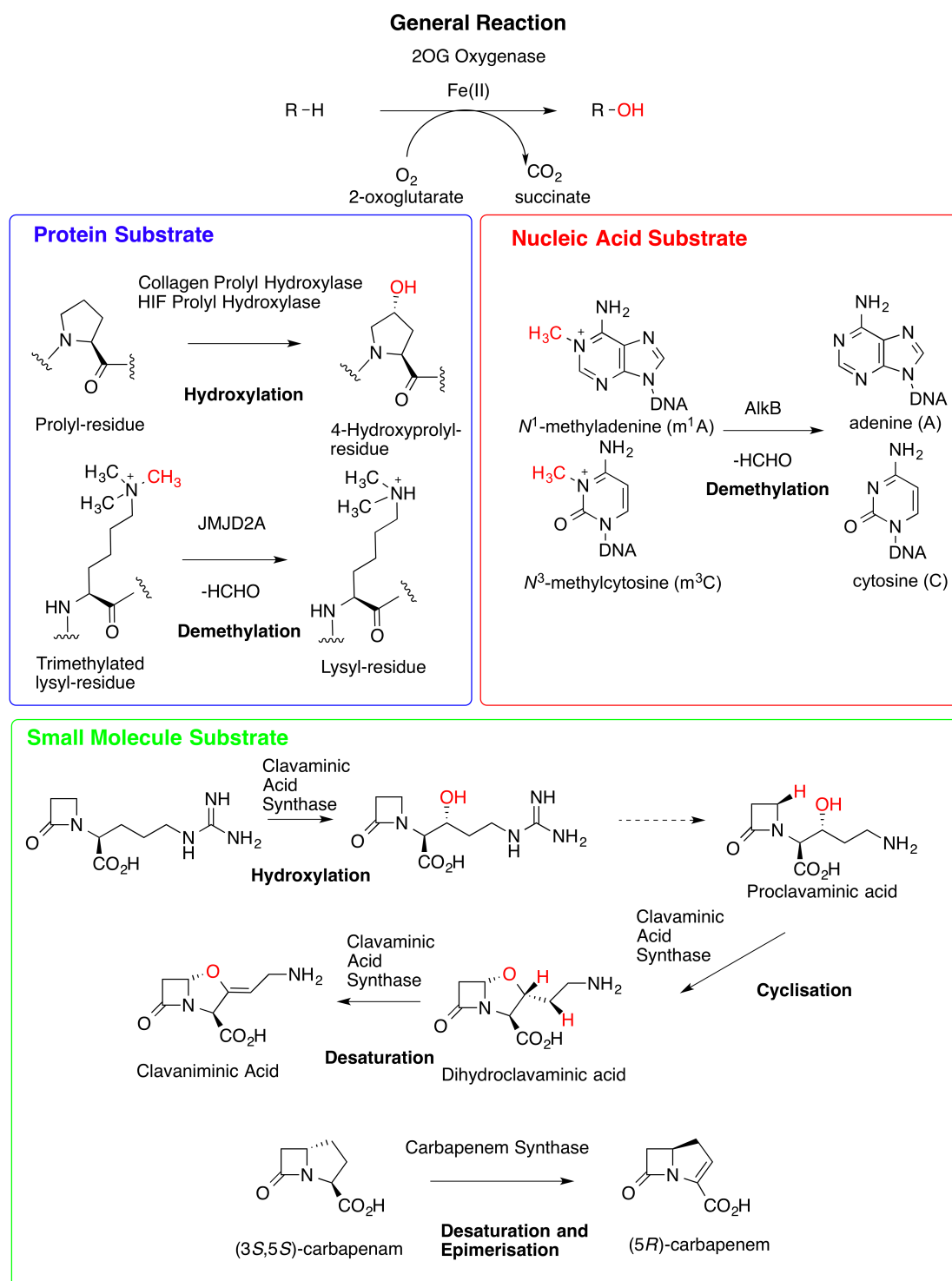


Figure 1 - 1. Examples of reactions catalysed by 2OG oxygenases. Atoms involved in the reactions catalysed by 2OG oxygenases are highlighted in red. Types of catalysed reactions are in bold.

1.1 Structural Overview of 2OG Oxygenases

1.1.1 The Double-Stranded Beta Helix Fold

The ferrous iron and 2-oxoglutarate dependent oxygenase superfamily has a characteristic core fold known as the double-stranded beta-helix (DSBH) fold, which is referred to by some as the 'cupin' or 'jelly roll' fold (5,8). 2OG oxygenases are not the only proteins that adopt the DSBH fold (8). Proteins such as cupins, cNMP-binding domain, copper monooxygenases, thiamin pyrophosphokinases, and viral capsid proteins have also been found to contain a DSBH fold (8). The term 'cupin fold' has sometimes been used to describe the core DSBH folds of 2OG oxygenases (5). The term 'DSBH' is preferable to 'cupin' when describing the fold adopted by 2OG oxygenases to avoid confusion because 2OG oxygenases are not functionally similar to cupins (8).

1.1.2 Structural Characteristics of 2OG Oxygenases

The core DSBH fold of 2OG oxygenases consists of eight β -strands forming two β -sheets: a major β -sheet (strands I, VIII, III, VI) and a minor β -sheet (strands II, VII, IV, V) (5,6). The two sheets form a sandwich structure with the active site opening to the end where the first β -strands (β I and II) are positioned (Figure 1 - 2).

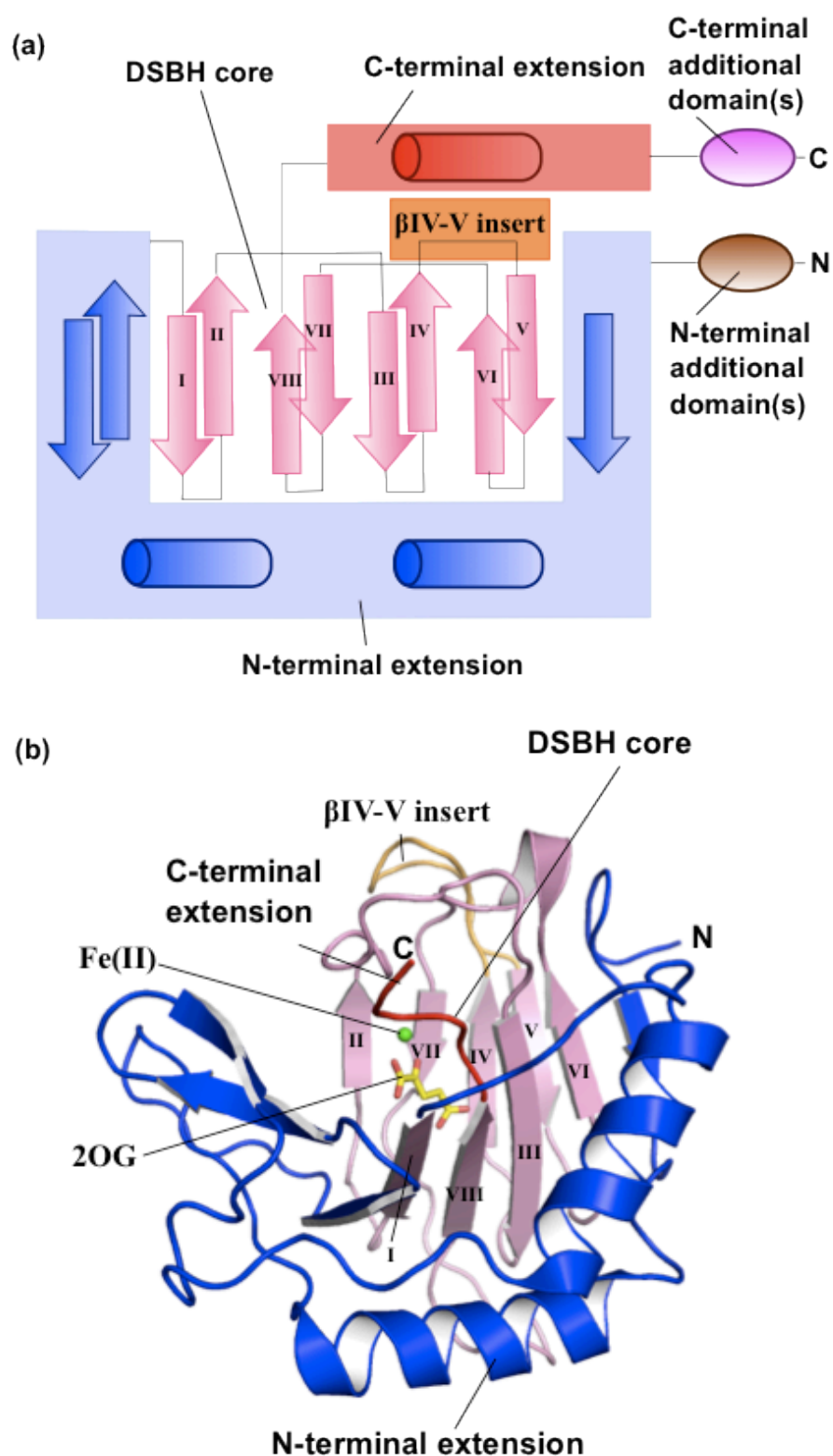


Figure 1 - 2. Structure of 2OG oxygenases. (a) General 2-dimensional topology of a 2OG oxygenase. (b) 3-Dimensional representation of a 2OG oxygenase adapted from the structure of AlkB (PDB ID: 3BIE; resolution: 1.68 Å) (9). Protein structural model was produced using MacPyMOL v1.3 (10).

An N-terminal extension from the DSBH core usually consists of a mixed α/β secondary structure with two or three helices flanking the major β -sheet, and one to three β -strands extending the major β -sheet on both ends (5,6). 2OG oxygenases typically possess β -strands and loops extending from the N-terminus of the DSBH; these elements can help to bind or enclose substrates in the active site (e.g. as occurs in the hypoxia inducible factor (HIF) prolyl hydroxylase 2, PHD2 (Figure 1 - 3a)) (11).

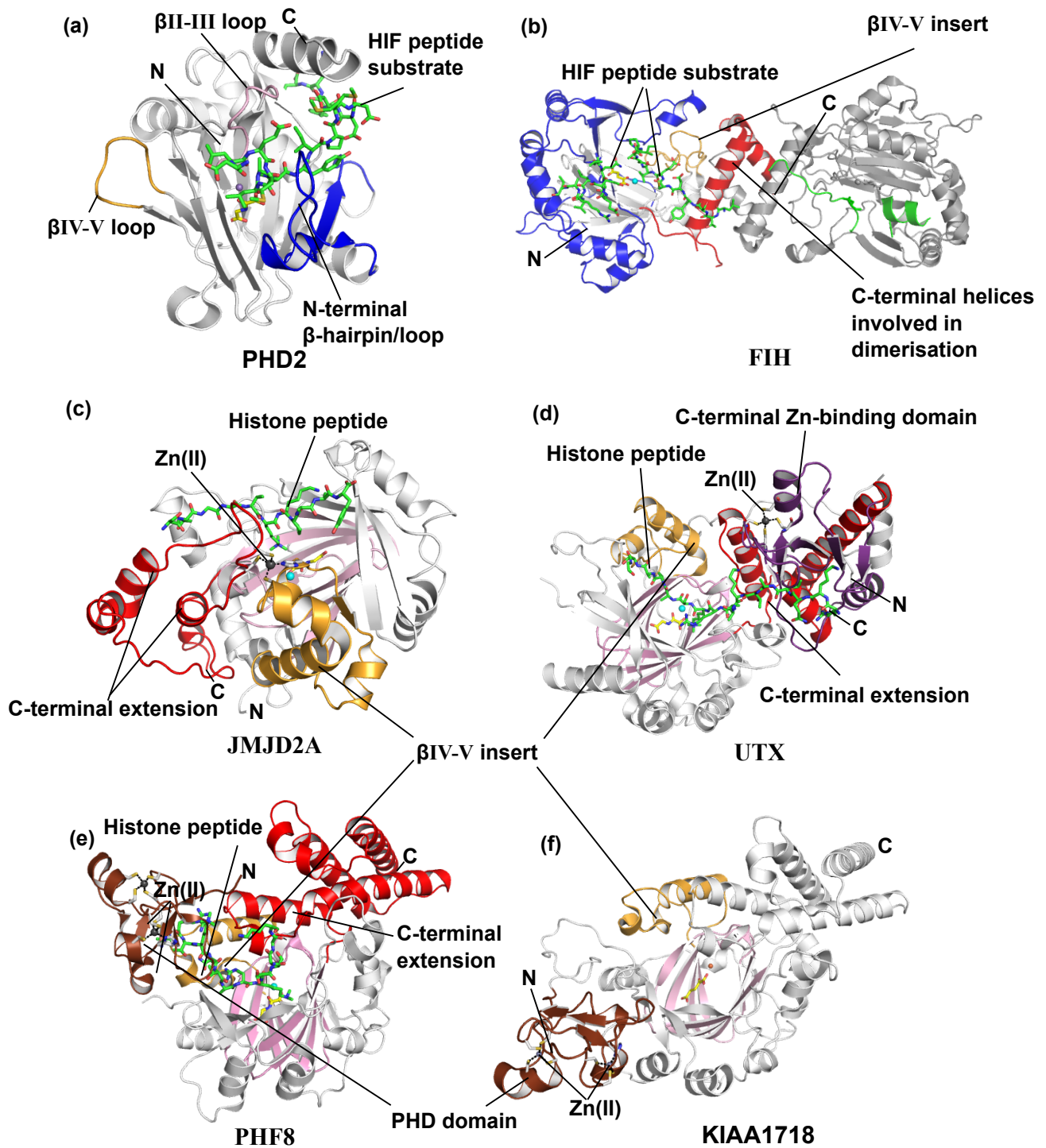
Extensions from the C-terminus of the DSBH core are, in general, less well-conserved; they can vary substantially for different 2OG oxygenases (5,6). Some members of the 2OG oxygenase family have α -helices as part of their C-terminal DSBH extension; these can be important in substrate binding (e.g. as occurs in histone demethylases Jumonji domain-containing protein 2A (JMJD2A) (Figure 1 - 3c) and PHD finger protein 8 (PHF8) (Figure 1 - 3e)) (12,13) or in dimerisation (e.g. as in HIF asparaginyl hydroxylase, FIH (Figure 1 - 3b)) (14).

A region between DSBH β -strands β IV and V (termed the β IV-V insert) sometimes plays an important role in substrate recognition in 2OG oxygenases (5,6). The β IV-V insert ranges from a short turn of a few residues to more than 300 residues long (e.g. as in the human ten-eleven-translocation enzymes, TETs) (15). In several 2OG oxygenases, the β IV-V insert forms additional secondary structures such as helices and strands (e.g. as in clavaminic acid synthase, CAS (Figure 1 - 3g) and taurine dioxygenase, TauD (Figure 1 - 3h)) (16,17).

Some multidomain 2OG oxygenases use additional discrete domains, located N- or C-terminal to the DSBH domain, for substrate recognition, e.g. at either or both termini of the JmjC histone demethylases PHF8, JmjC domain-containing histone demethylation protein 1D (KIAA1718), Jumonji domain-

containing protein 3 (JMJD3), histone demethylase UTX (UTX), and JMJD2A

(Figure 1 - 3b-e) (6,12,13,18-21).



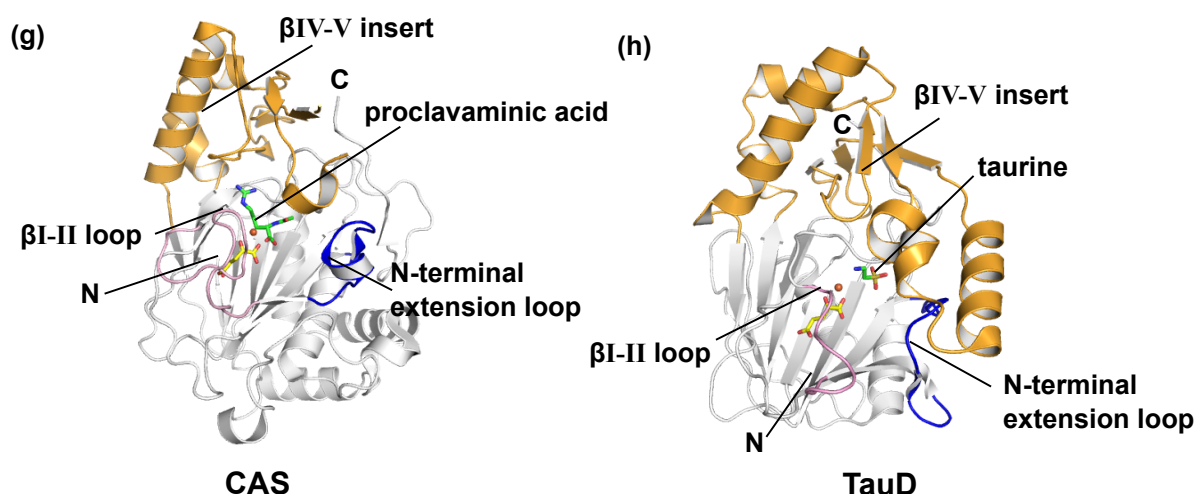


Figure 1 - 3. Representative crystal structures of 2OG oxygenases. View from crystal structures of **(a)** the hypoxia inducible factor (HIF) prolyl hydroxylase 2, PHD2 (PDB ID 3HQR; resolution: 2.00 Å) (11), **(b)** the HIF asparaginyl hydroxylase, FIH (PDB ID 1H2L; resolution: 2.25 Å) (14), **(c)** the histone demethylase JMJD2A-H3K36me3 peptide complex (PDB ID 2OS2; resolution: 2.30 Å) (13), **(d)** the histone demethylase UTX-H3K27me3 peptide complex (PDB ID 3AVR; resolution: 1.80 Å) (19), **(e)** the histone demethylase PHF8-histone H3 peptide containing H3K4me3 and H3K9me2 complex (PDB ID 3KV4; resolution: 2.19 Å) (12), **(f)** the histone demethylase KIAA1718 (PDB ID 3KV6; resolution: 2.89 Å), **(g)** clavaminic acid synthase, CAS (PDB ID 1DRT, resolution: 2.10 Å) (16), and **(h)** taurine dioxygenase, TauD (PDB ID 1OS7; resolution: 2.50 Å) (17). Specific regions are highlighted using the colour scheme as defined in Figure 1 – 2. 2OG/*N*-oxalylglycine (NOG), yellow sticks; substrate, green sticks; Fe(II), orange sphere; Ni(II), cyan sphere; Mn(II), purple sphere. Protein structure models were produced using MacPyMOL v1.3 (10).

1.1.3 Active Site Metal Binding

The DSBH core fold of the 2OG oxygenases provides a scaffold for the Fe(II) cofactor, 2OG cosubstrate, and to some extent, primary substrate binding sites (1,5). At the active site, the single Fe(II) (at least normally) is octahedrally coordinated by conserved residues consisting of two histidines and a carboxylate (aspartate or glutamate). This configuration is referred to as the HXD/E...H motif (1,5). The remaining coordination sites are occupied by water molecules, two of which are displaced on binding of 2OG. The HXD/E...H motif is shared with several other related DSBH-containing oxygenases that are not dependent on 2OG such as isopenicillin N synthase (IPNS), hydroxypropylphosphonic acid epoxidase (HppE), and 1-aminocyclopropane-1-carboxylic acid oxidase (ACCO) (22-24).

In the 2OG oxygenase superfamily, the iron chelating residue side chains of HXD/E originate from the C-terminal end of β -strand II and the loop before β -strand III, while the second histidine of the catalytic triad is located at the N-terminal end of β -strand VII (1,5) (Roman numerals are used to refer to the eight strands of the DSBH). However, there are enzymes in the 2OG oxygenase superfamily that only use two histidine residues to chelate Fe(II) (e.g. syringomycin biosynthesis enzyme 2 (SyrB2) halogenase (Figure 1 - 4b) and aspartyl/asparaginyl beta-hydroxylase (AspH) (unpublished; PDB ID:3RCQ)) (25). Mutagenesis studies on FIH indicated that two histidines were sufficient for hydroxylation (26). The remaining Fe(II) binding sites are coordinated by the oxalate moiety of 2-oxoglutarate (bidentate chelation), and oxygen (1,5).

1.1.4 2-Oxoglutarate Binding Site

The cosubstrate 2OG binding site in 2OG oxygenases is adjacent to the active site Fe(II) and buried within the DSBH core between the major and minor β -sheets, away from the active site opening (5,6). 2OG chelates Fe(II) in a bidentate manner via oxygen atoms from its C2-keto and C1-carboxylate groups (Figure 1 - 4) (1,5).

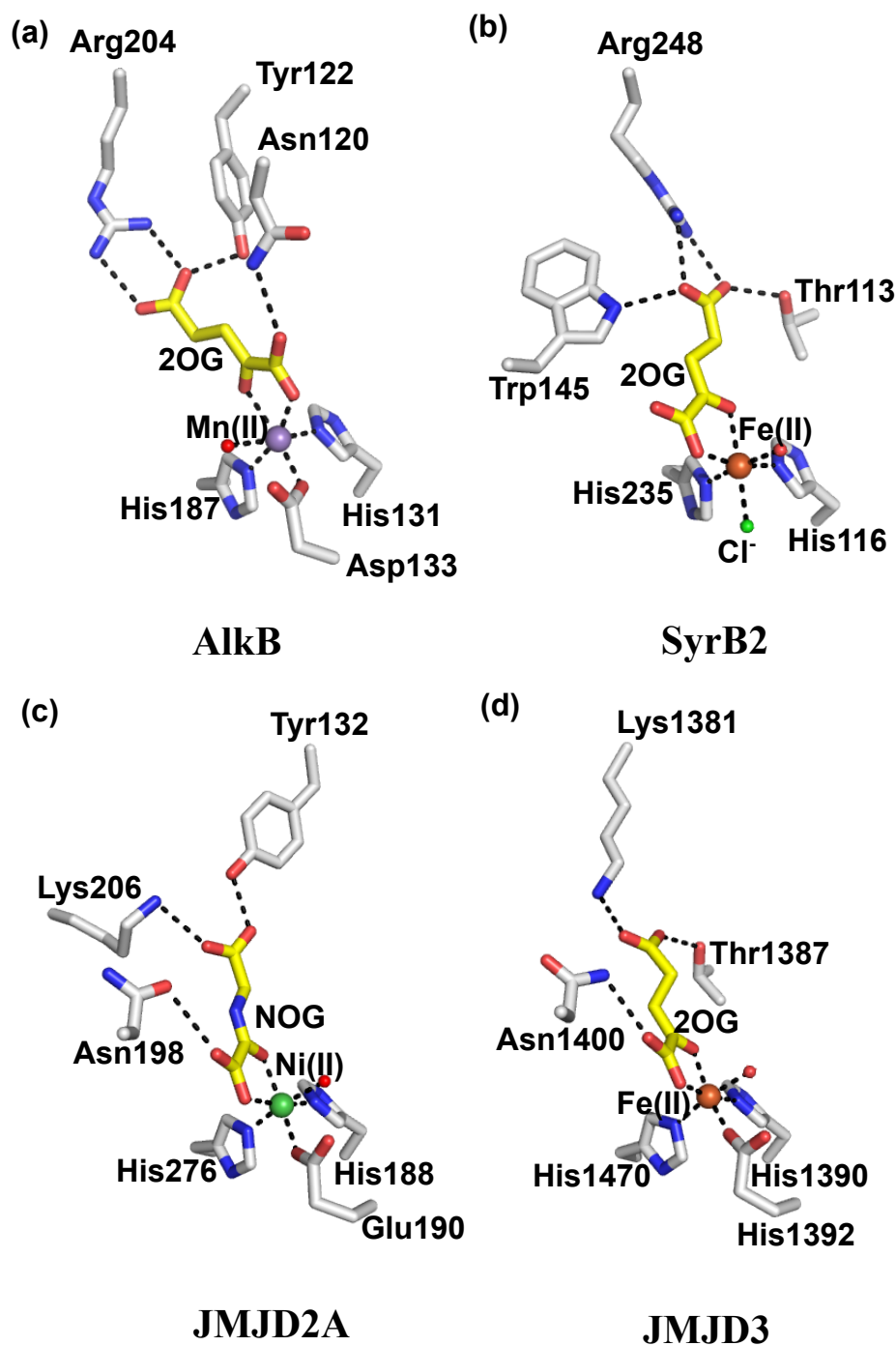


Figure 1 - 4. Views of 2-oxoglutarate co-substrate binding sites. 2OG binding site of (a) AlkB (PDB ID 3B1E; resolution: 1.68 Å) (9), (b) SyrB2 halogenase (PDB ID 2FCT; resolution: 1.60 Å) (25), (c) JmjC histone demethylase JMJD2A (PDB ID 2OQ6; resolution: 2.00 Å) (13), (d) and JmjC histone demethylase JMJD3 (PDB ID 2XUE; resolution: 2.00 Å) (18). Protein structure models were produced using MacPyMOL v1.3 (10).

In all determined structures, the C5-carboxylate group of 2OG interacts electrostatically with the side chain of an arginine- or lysine-residue (1). Binding

by the basic residue is typically supported by an additional interaction including a residue that contains a side chain alcohol or phenol (e.g. tyrosine-, threonine- or serine-residue) (Figure 1 - 4) (1).

Sequence analyses of 2OG oxygenases divide the superfamily into two clades: the classical 2OG oxygenase and the JmjC-domain containing 2OG oxygenase (8). According to Iyer *et al.* 2010, the JmjC oxygenases have several distinct structural features distinguishing them from the classical 2OG oxygenases (8). These include an N-terminal strand-helix-strand motif (previously mistakenly thought to be a separate domain called the JmjN domain (20,27)), an additional strand-helix unit inserted into the extension of the major β -sheet where the active site opening is located, and a helical hairpin with an extended loop between DSBH β IV and V (8).

Another interesting feature of the JmjC oxygenases is that they typically utilise a lysine located on the N-terminal end of β IV to interact with the C5-carboxylate of 2OG (Figure 1 - 4c). Exceptions include the *E. coli* ribosomal protein hydroxylase ycfD, which uses an arginine from β IV (28,29) and from 2OG oxygenases from the UTX/JMJD3 subfamily, the lysine-residue of which is derived from β I (Figure 1 - 4d) (5,6,8,18,19). On the other hand, the classical 2OG oxygenases use an arginine from the N-terminal end of β VIII to interact with the C5-carboxylate of 2OG (Figure 1 - 4a, b) (5,6,8).

Based on the crystal structures of 2OG oxygenases in complex with their 2OG cosubstrate, 2OG is able to bind in two different coordination modes, which may vary between enzymes (1,2,5). The C1-carboxylate of 2OG is observed to coordinate to the metal, either *trans* to the proximal histidine residue (e.g. JMJD2A (13), Figure 1 - 4c) or *trans* to the distal histidine residue (e.g. *E. coli*

alkylated DNA repair protein (AlkB) (9), Figure 1 - 4a) of the facial triad (5). The former coordination mode should enable oxygen to bind *trans* to the distal histidine, and form an oxo-ferryl intermediate projecting towards the C-H bond of the prime substrate for oxidation, hence termed 'in-line' because the oxo-ferryl intermediate is in-line with substrate (2). On the other hand, binding of 2OG with its 1-carboxylate *trans* to the distal histidine would result in an oxo-ferryl intermediate that projects away from the prime substrate, hence termed 'off-line' due to the position of the oxo-ferryl intermediate that is not in-line with substrate (2). The two 2OG binding modes lead to two proposed catalytic mechanisms for 2OG oxygenases: the in-line 'hydroxylation' mechanism and the off-line 'hydroxylation' (2) (see next section).

1.1.5 Proposed General Catalytic Mechanisms of 2OG Oxygenases

A general catalytic mechanism for 2OG oxygenases has been proposed (Figure 1 - 5) through a combination of crystallographic and spectroscopic analyses (2). As a starting point, the Fe(II) in the active site is octahedrally coordinated by side chains of the catalytic triad residues (HXD/E...H), and three water molecules (Figure 1 - 5a, complex A), which can be clearly observed in several crystal structures of 2OG oxygenases such as deacetoxycephalosporin C synthase (DAOCS) and proline-3-hydroxylase (P3H) in the absence of 2OG cosubstrate (30,31). This inactive 2OG oxygenase complex is supported by variable-temperature, variable-field (VTVH) magnetic circular dichroism (MCD) spectroscopic data indicating that a resting clavaminic acid synthase (CAS) (without 2OG binding) has a six-coordinate ferrous site (32).

When 2OG binds in the active site by chelating the Fe(II) in a bidentate manner, two water molecules are displaced (Figure 1 - 5a, complex B) (2). The transition from coordination by water molecules to a bidentate chelation by 2OG in CAS has been observed using UV/visible, circular dichroism (CD), and MCD spectroscopies (32,33). As described above, 2OG is observed crystallographically to bind in two different modes (1-carboxylate of 2OG *trans* to either the proximal or distal histidine of the HXD/E...H motif) (2,5). Figure 1 - 5a depicts the proposed 'in-line' hydroxylation mechanism.

After the binding of 2OG, the primary substrate binds to the active site and promotes the formation of a 5-coordinated Fe(II) through the loss of the third water molecule (Figure 1 - 5a, complex C) (2). The 5-coordination Fe(II) centre is evidenced by crystal structures of FIH, TauD and alkylsulfatase (AtsK) in complex with substrates (14,17,34). However, a crystal structure of CAS in complex with substrate revealed a 6-coordinated Fe(II) but the metal-to-water distance increases from 2.2 to 2.35 Å (16). Several spectroscopic techniques including UV/visible (on CAS, TauD and 2,4-dichlorophenoxyacetate dioxygenase (TfdA)), resonance Raman (on TauD), and MCD (on CAS) (33,35-38) support most of the crystallographic analyses, suggesting the transition of a 6-coordinated state to a 5-coordinated state for the active site metal upon substrate binding (2). However, kinetic studies on collagen prolyl-4-hydroxylase indicated that oxygen may bind prior to substrate binding (39,40), suggesting that there may be some enzyme-dependent variation in the order of oxygen and substrate binding (1). (Note, however, the substrate is a large protein in the case of the collagen prolyl-4-hydroxylases.)

The vacant iron coordination site, due to the departure of the third water molecule, then allows oxygen to bind to the active site Fe(II) (2). However, the subsequent steps leading to substrate oxidation or hydroxylation are less well understood (2). Various spectroscopic experiments (e.g. EPR, Mossbauer and resonance Raman) performed on TauD support the presence of an Fe(IV)-oxo intermediate (41,42), which is proposed to enable hydroxylation (2). To generate the Fe(IV)-oxo species, oxygen is suggested to bind to Fe(II), possibly forming a Fe(III)-superoxo species (Figure 1 - 5a, complex D), followed by a nucleophilic attack on the keto group of 2OG, forming a peroxyhemiketal bicyclic complex (Figure 1 - 5a, complex E) (2). Oxidative decarboxylation then occurs, to generate a succinate, carbon dioxide and an Fe(IV)-oxo intermediate (Figure 1 - 5a, complex F) (4).

If the Fe(IV)-oxo intermediate is located at the appropriate position as proposed for the in-line hydroxylation mechanism, the activated oxygen species can abstract a hydrogen from the substrate (43), forming a substrate radical and an Fe(III)-hydroxide species (Figure 1 - 5a, complex G) (2). Finally, the hydroxyl group and the substrate radical can recombine to give a hydroxylated product (2). The dissociation of product and co-products completes the catalytic cycle (2,6).

In the off-line hydroxylation mechanism, the vacant oxygen binding site is not positioned adjacent to the C-H bond of the substrate (2). Two possibilities were proposed in this situation (1,2). The first possibility is that the 2OG reorients to adopt an "in-line geometry" (1,2), as suggested by studies on the binding of nitric oxide (oxygen analogue) in clavaminic acid synthase (CAS) that showed that 2OG was able to rearrange from *trans* to proximal histidine to *trans*

to the distal histidine (44). The second possibility is that the reaction proceeds to generate an off-line Fe(IV)-oxo species and then lose a CO₂ (Figure 1 - 5b, complex H) (2). An addition of water into the vacant 'CO₂' site produces a dihydroxylated metal (Figure 1 - 5b, complex I), which will then rearrange ('ferryl flip') through the loss of a water molecule, to reposition the Fe(IV)-oxo intermediate close to the substrate for hydrogen abstraction (Figure 1 - 5b, complex J) (1,2,44). Similar catalytic mechanisms involving oxidation of Fe(II) to Fe(IV) are adopted by the heme oxygenases (e.g. cytochrome-P450) (45).

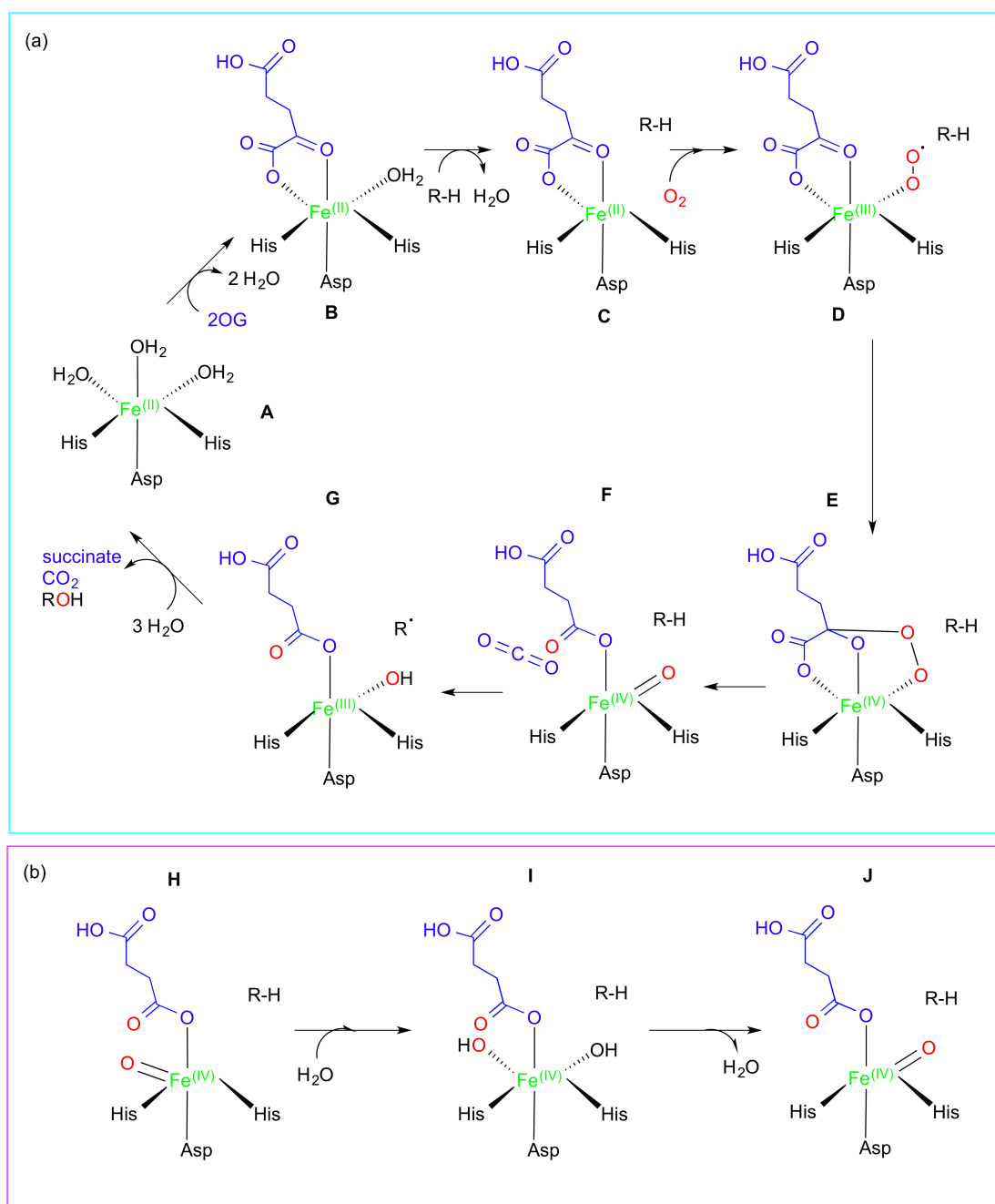


Figure 1 - 5. Proposed catalytic mechanisms of 2OG Oxygenases. (a) Proposed mechanism for 'in-line' hydroxylation. **(b)** Proposed mechanisms of the reorientation of oxo-ferryl intermediate ('ferryl flip') in the 'off-line' hydroxylation mechanism. A water molecule may bind to the Fe(IV)=O species to give a dihydroxylated Fe(IV) species, which isomerises. Atoms from 2OG are highlighted in blue, oxygen atoms are highlighted in red and iron ions are highlighted in green.

1.2 Nucleic Acid Modifying Oxygenases

A subset of Fe(II) and 2OG dependent oxygenases act on nucleic acids. These 2OG oxygenases are termed nucleic acid oxygenases (NAOX). To date, members

from three different subfamilies of 2OG oxygenases have been shown to act on nucleic acids; i.e. the AlkB, TET/JBP and JmjC hydroxylase subfamilies (46). Reactions catalysed by NAOX include DNA demethylation, RNA demethylation, RNA hydroxylation and DNA oxidation (Figure 1 - 6) (46,47). Some of the NAOX are linked to diseases such as cancer (e.g. ALKBH2 and ALKBH8) and obesity (e.g. FTO) (46).

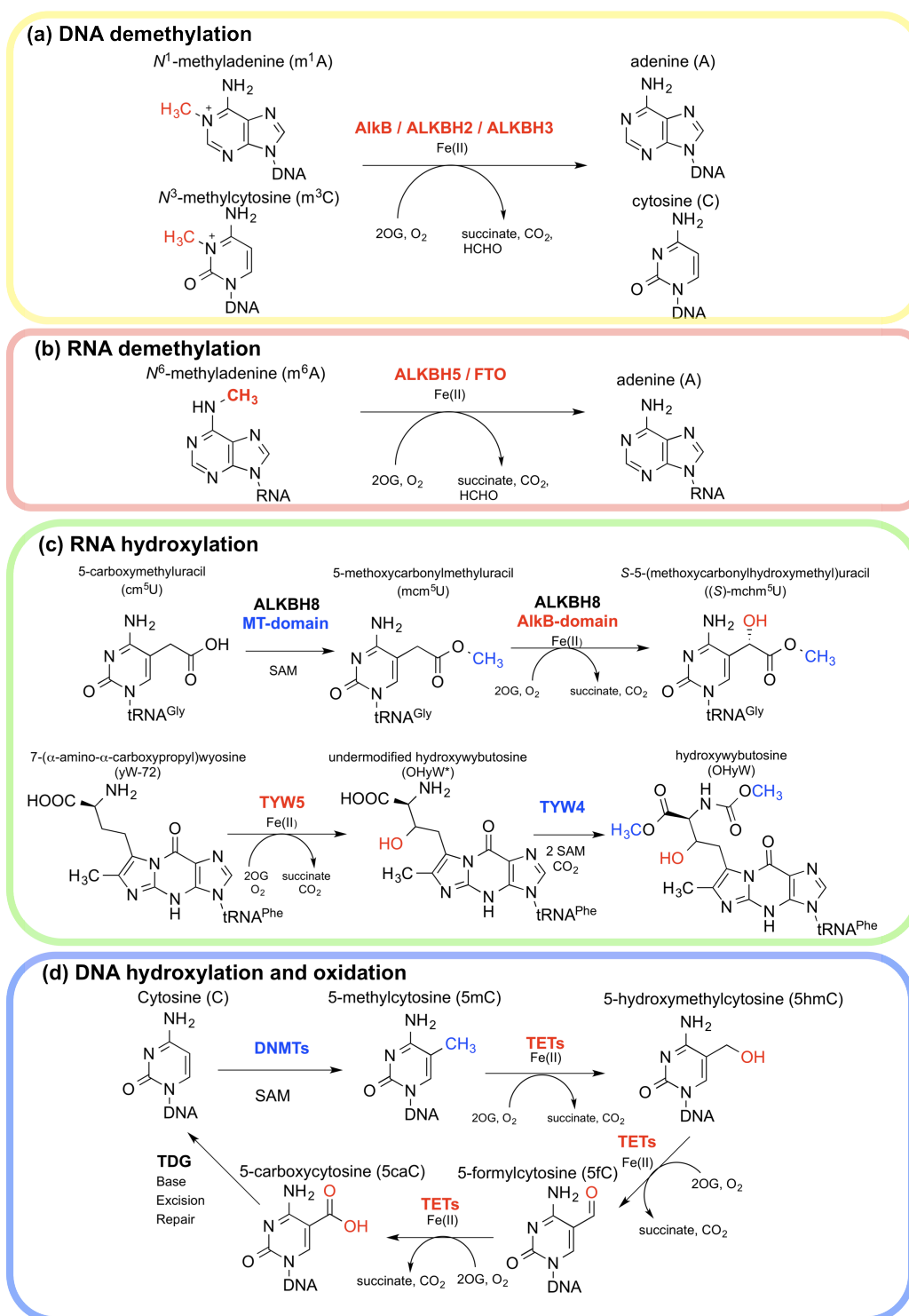


Figure 1 - 6. Summary reaction schemes catalysed by the nucleic acid oxygenases (NAOX). Colour scheme: Atoms and enzymes involved in oxygenase-catalysed reactions (hydroxylation/oxidation/demethylation) are highlighted in red; atoms and enzymes involved in methyl transfer reactions are highlighted in blue.

1.2.1 Discovery of the Function of AlkB

In *Escherichia coli*, one enzyme that is able to remove alkyl groups directly from DNA bases is AlkB (48,49). AlkB was also the first identified member of the AlkB subfamily of 2OG oxygenases. The *alkB* gene was discovered in 1986 when an *alkB* mutant was observed to be more sensitive than was the wild type to the treatment of methyl methanesulfonate (MMS), an alkylating agent (50). Over the next two decades, the mechanisms of action of AlkB in DNA repair remained a mystery. The breakthrough in understanding the role of AlkB came when bioinformatics studies carried out by Aravind and Koonin indicated that AlkB is a member of the Fe(II) and 2-oxoglutarate (2OG) superfamily (51). Subsequently, AlkB was demonstrated to demethylate m¹A and m³C lesions in DNA in the presence of Fe(II) cofactor and 2OG cosubstrate (Figure 1 - 7) (48,49). Further biochemical investigations showed that AlkB prefers single-stranded nucleic acids to double-stranded nucleic acids as substrates (52). Known substrates of AlkB are not limited to just m¹A and m³C, as AlkB is also shown to accept N¹-methylguanine (m¹G), N³-methylthymine (m³T), 1,N⁶-ethenoadenine (εA), 3,N⁴-ethenocytosine (εC), 1,N⁶-ethanoadenine (EA), 3,N⁴-α-hydroxypropanocytosine, and N⁶-methyladenine (m⁶A) as substrates (47).

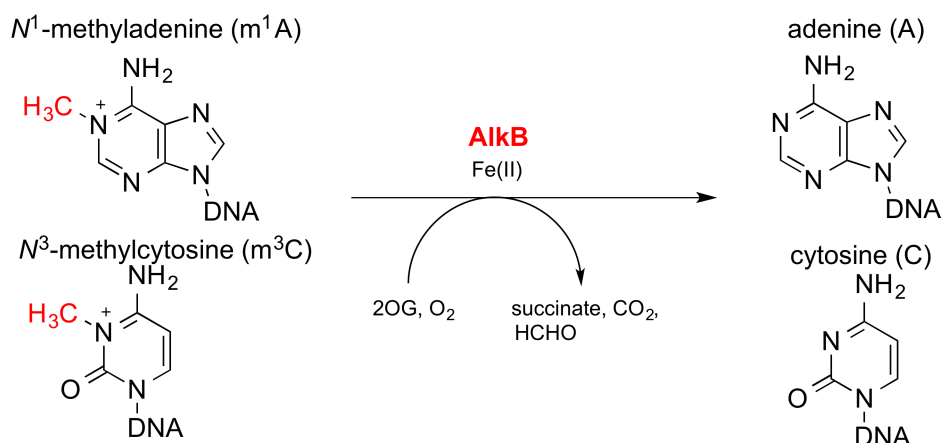


Figure 1 - 7. Reactions catalysed by AlkB.

The structure of AlkB in complex with trinucleotide containing an m^1A base was initially solved to 2.3 Å resolution (53). The AlkB-trinucleotide structure showed that AlkB adopts the conserved DSBH fold and has a N-terminal nucleotide recognition lid (NRL) near the active site opening. Part of this lid contacts the trinucleotide, and its interaction with the substrate induces the flipping of m^1A base into the active site via a direct stacking of adjacent bases. This mechanism was confirmed by structures of AlkB in complex with dsDNA (Figure 1 - 8) and with ssDNA (9,54). High-resolution structures of AlkB-dsDNA complexes were also successfully involved in mechanistic studies through the capture of transient hemiaminal intermediates (55).

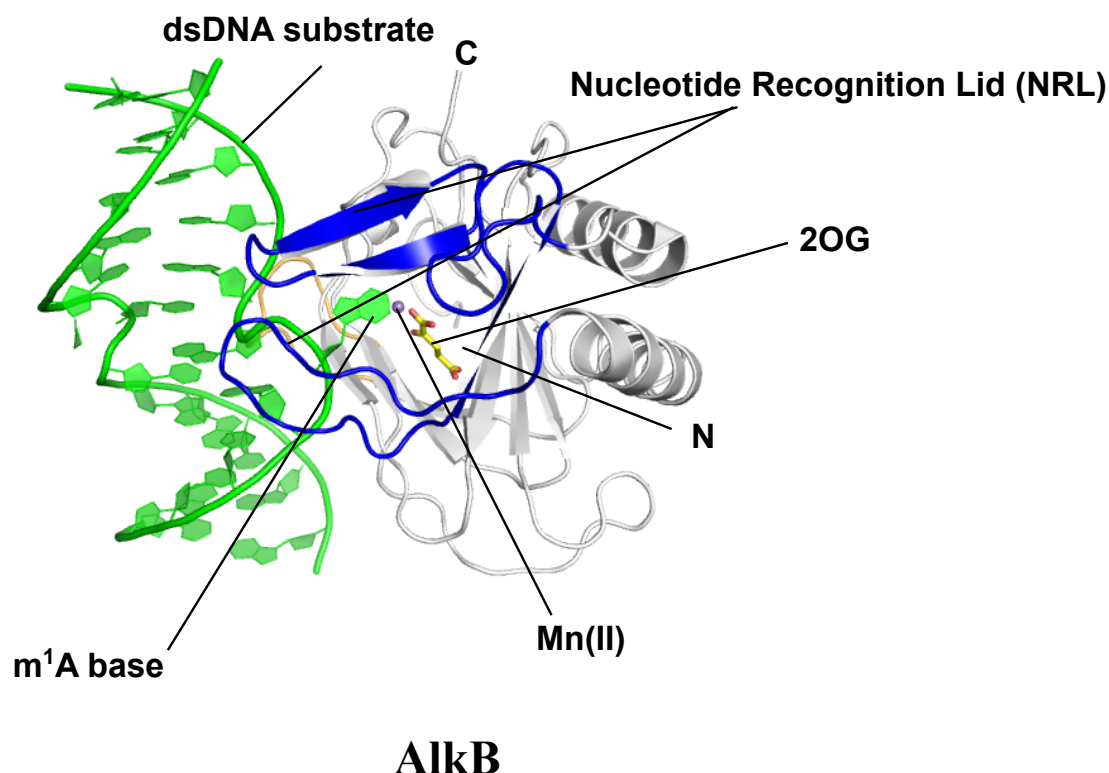


Figure 1 - 8. View from a structure of AlkB-DNA complex. Cartoon representation of the crystal structure of *E. coli* alkylated DNA repair protein (AlkB)-DNA complex (PDB ID 3BIE; resolution: 1.68 Å). N-Terminal nucleotide recognition lid (NRL), blue; β IV-V loop, orange; double-stranded DNA (dsDNA), green; 2-oxoglutarate (2OG), yellow; Mn(II) ion, purple sphere. Protein structure models were produced using MacPyMOL v1.3 (10).

As AlkB is biochemically and structurally well-characterised, preliminary inhibition studies on the NAOX were performed using AlkB as a model target (see Chapter 2) (56).

1.2.2 Human Homologues of AlkB

There are 9 homologues of AlkB (ALKBH1 – 8 and the fat mass and obesity associated protein, FTO). In the following sections, the biochemical functions and structures of the human NAOX are summarised.

ALKBH1

Through sequence comparison using a human cDNA database, ALKBH1 was identified as a human homologue of AlkB, with 34% identity and 59% similarity at the amino acid level (57). RNA blot analyses show that ALKBH1 is ubiquitously expressed in various human adult tissues such as brain, testis, small intestine, prostate and ovary (57). However, in contrast with AlkB, treatment of cells with MMS did not stimulate ALKBH1 expression.

When ALKBH1 was first identified, the biochemical function of AlkB was still unclear. After the successful biochemical characterisation of AlkB as a 2OG oxygenase, ALKBH1 was reported not to have m³C and m¹A demethylase activity (58). In 2008, Westbye *et al.* reported that ALKBH1 is a mitochondrial protein that exhibits m³C demethylase activity *in vitro*, contradicting previous reports (59). Therefore, the status of ALKBH1 as a DNA demethylase, despite being AlkB's most similar human homologue, remains unclear.

The function of ALKBH1 was further confounded by reports claiming that it cleaves DNA at apurine/aprymidine (AP) sites (34,60,61). Recently, ALKBH1 was suggested to be a histone H2A demethylase, as deduced from a comparison of the methylation status of histone H2A between ALKBH1 knockout and wild type MEF cells (62). However, this observation could not be confirmed *in vitro* using methylated histone peptides (62).

Interestingly, ALKBH1 knockout mice display sex-ratio distortion, eye defects, and impairment in placental trophoblast lineage differentiation (63), suggesting that ALKBH1 is important in mammalian development.

ALKBH2 and ALKBH3

Two other human and mouse homologues of AlkB, ALKBH2 and ALKBH3, were later identified by BLAST searches (58,64). ALKBH2 and ALKBH3 were shown to be able to demethylate m³C and m¹A on dsDNA, ssDNA and ssRNA *in vitro* (Figure 1 - 9) (58,64). ALKBH2 prefers dsDNA to ssDNA/ssRNA while ALKBH3 prefers ssDNA to dsDNA (58). Expression of ALKBH2 and ALKBH3 complemented AlkB-mutated-*E. coli* strains by reactivating ssDNA phage in the presence of MMS, which would otherwise suppress ssDNA phage reactivation in the absence of AlkB (58).

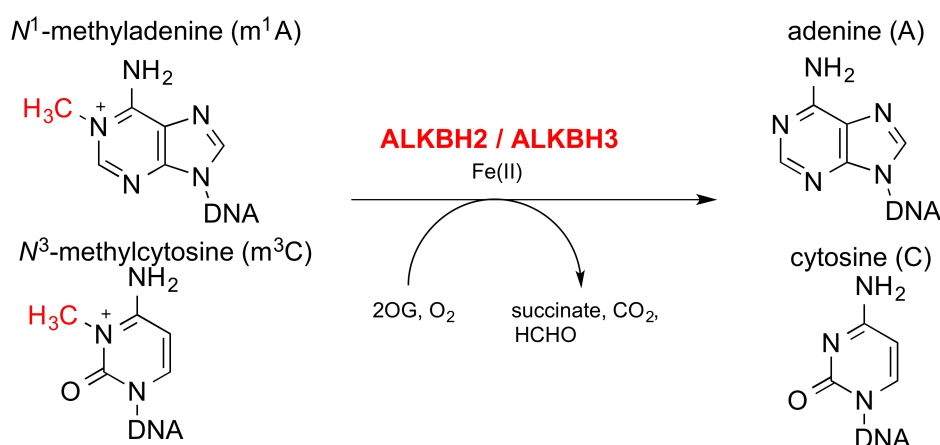


Figure 1 - 9. Reactions catalysed by ALKBH2 and ALKBH3.

ALKBH2 is proposed to be the predominant DNA alkylation repair enzyme in mice because ALKBH2 deficient mice, but not ALKBH3 deficient mice, accumulate m¹A damage in their DNA in the absence of exposure to exogenous methylating agents (65).

Due to its preference for single-stranded nucleic acids, ALKBH3 was previously thought to play a role in RNA repair (52). In 2011, ALKBH3 was reported to associate with the 3' – 5' helicase ASCC3 that unwinds dsDNA to

allow ALKBH3 to repair ssDNA, showing that ALKBH3 plays a role in DNA repair and can act with the help of a helicase (66).

ALKBH2 and ALKBH3 point mutations have been identified in pediatric brain tumour samples, and ALKBH2 is upregulated in some tumour tissues especially in gliomas and medulloblastomas (67). ALKBH3 is found to be overexpressed in prostate cancer (66). Knockdown of ALKBH2 has been shown to aid in cis-platin chemotherapy on the H2199 non-small cell lung cancer cell, suggesting that combination therapies with ALKBH2 inhibitors and cancer drugs may be a useful strategy for cancer treatment (46,68). ALKBH3 and ASCC3 are also required for cell proliferation in a cell-line specific manner and knockdown of either of them decreases cell proliferation (even in the absence of MMS) (66).

Crystal structures of ALKBH2 in complex with dsDNA have been determined (Figure 1 - 10a); they reveal the role of a phenylalanine finger from the NRL in flipping the alkylated base into the active site for repair (9). This contrasts with the base flipping mechanism adopted by AlkB, which is by directly stacking bases adjacent to the methylated base (9,53). A structure of ALKBH3 has also been determined but the structure does not contain a DNA substrate (Figure 1 - 10b) (69). A β -hairpin in the NRL of ALKBH3 was suggested to confer single-stranded nucleic acid selectivity (69).

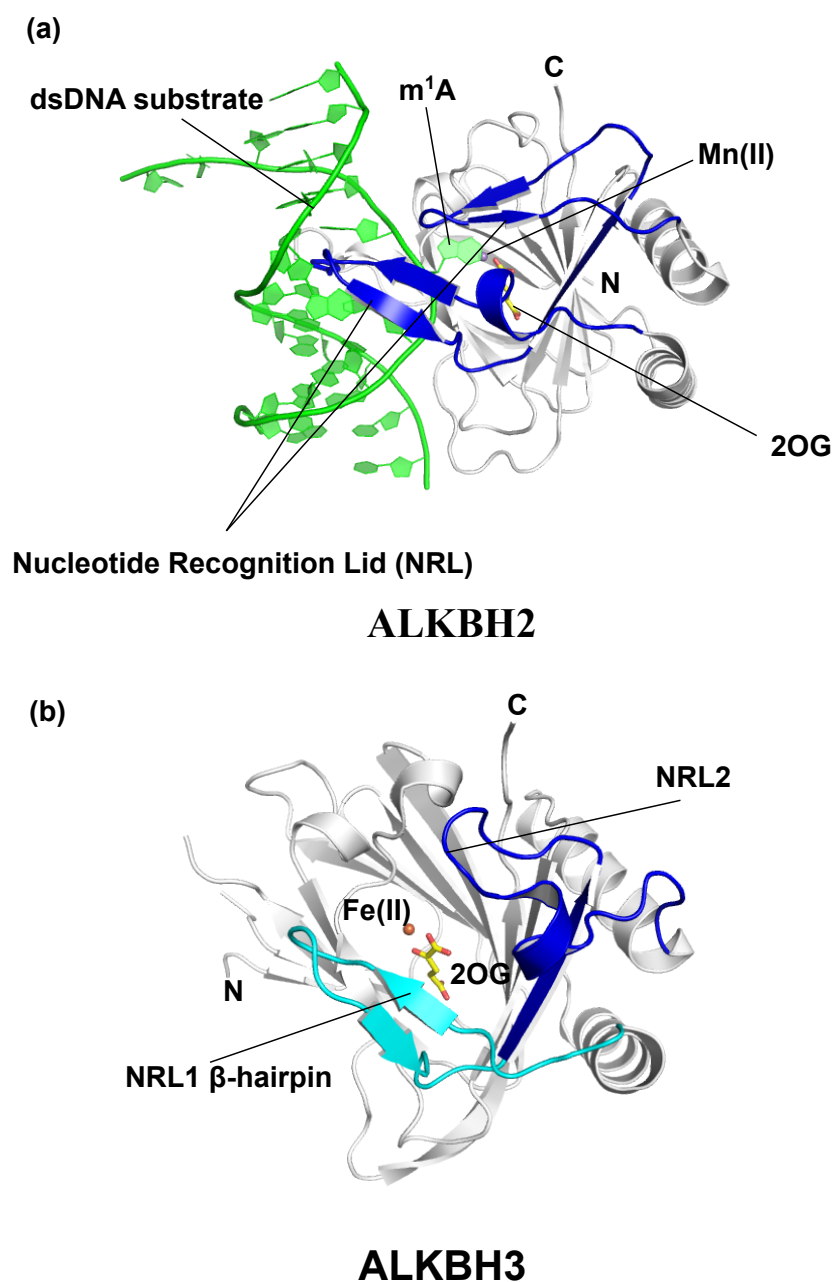


Figure 1 - 10. Views from crystal structures of ALKBH2 and ALKBH3. Cartoon representations of crystal structures of **(a)** the human AlkB homologue 2 (ALKBH2)-DNA complex (PDB ID 3BUC; resolution: 2.59 Å) and **(b)** the human AlkB homologue 3 (ALKBH3) (PDB ID 2IUW; resolution: 1.50 Å). General colour coding: β IV-V loop, orange; double-stranded DNA (dsDNA), green; 2-oxoglutarate (2OG), yellow; Mn(II) ion, purple sphere; Fe(II), orange sphere. In (a), N-terminal nucleotide recognition lid (NRL) is coloured blue. In (b), the β -hairpin in NRL1 (cyan) of ALKBH3 is suggested to confer single-stranded DNA selectivity; NRL2, blue. Protein structure models were produced using MacPyMOL v1.3 (10).

ALKBH4, ALKBH6 and ALKBH7

Five other human AlkB homologues, ALKBH4 – 8, were identified through bioinformatics analyses (70). The functions of ALKBH4, 6 and 7 are still unclear. Specifically, no functional studies or related investigations on ALKBH6 have been reported to date. ALKBH4 was reported to demethylate a monomethylated lysine on actin, potentially regulating actin-myosin interaction (71). Further investigations are needed to confirm this finding. ALKBH7 is localised to the mitochondria (72) and its deletion in mice resulted in an increase in body weight (73). Recently, ALKBH7 was also reported to be required for programmed necrotic cell death upon exposure to DNA-damaging agents (74). However, the substrate and mechanisms of action of ALKBH7 remain elusive.

ALKBH5

ALKBH5 is a nuclear protein and is upregulated via the hypoxia inducible factor (HIF) pathway in hypoxia (75). ALKBH5 is ubiquitously expressed in various tissues; the highest expression is found in the lung and testis (72,76). A truncated construct of ALKBH5 encompassing amino acids 66-292 was shown to have 2OG uncoupled turnover activity *in vitro* (75). However, the truncated construct did not show any demethylase activity against m¹A and m³C in ssDNA (75).

In 2013, Zheng *et al.* discovered that ALKBH5 demethylates N⁶-methyladenine (m⁶A) in mRNA and a deficiency in ALKBH5 caused impaired spermatogenesis in mice (76). Immunofluorescence studies showed that ALKBH5 colocalises with SC35 splice factor in nuclear speckles, suggesting that it may be involved in mRNA processing (76). siRNA knockdown of ALKBH5 in cells

led to an increase in mRNA export to the cytoplasm (76). Additionally, ALKBH5 was shown to prefer ssRNA to dsRNA and stem loop RNA as substrate. ALKBH5 also has a higher activity towards m⁶A in a consensus motif GGm⁶ACU (76). This consensus motif was investigated because m⁶A was predominantly found in Pu[G>A]m⁶AC[A/C/U] through m⁶A-seq (77,78). An independent study on mRNA-bound protein employing UV cross-linking and proteomics identified that ALKBH5 is an mRNA binding protein, supporting findings reported by Zheng *et al.* (76,79).

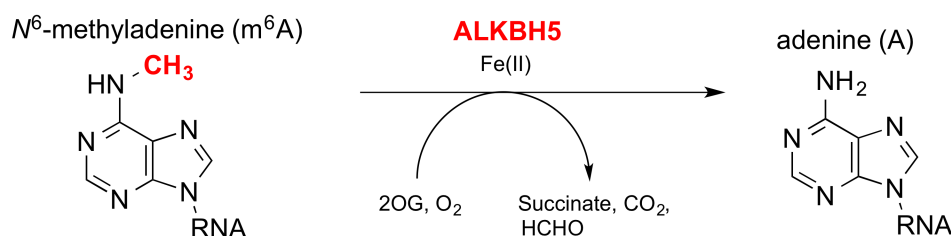


Figure 1 - 11. Reaction catalysed by ALKBH5.

As emerging evidence revealed that ALKBH5 is an RNA demethylase that is involved mRNA processing and fertility in mice, several groups independently carried out crystallographic studies on ALKBH5 to elucidate its mechanisms of substrate recognition and for structure-based inhibitor design (12,80,81). Chapter 4 in this thesis describes structural studies of the first published crystal structure of human ALKBH5 (19).

ALKBH8

ALKBH8 is, to date, the only AlkB subfamily member found to be involved in tRNA wobble uridine hypermodification. It consists of three domains: an N-

terminal RNA recognition motif (RRM), the AlkB-like domain, and a C-terminal methyltransferase (MT) domain.

The MT domain of ALKBH8 modifies 5-carboxymethyluridine (cm^5U) to 5-methoxycarbonylmethyluridine (mcm^5U) (82,83). It was discovered through its homology with *Saccharomyces cerevisiae* Trm9, a methyltransferase that catalyses the same reaction in yeast (84). As shown by ALKBH8 knockout cells, the mcm^5U is important for recoding of the UGA stop codon for the incorporation of selenocysteine (83).

Although the AlkB domain of ALKBH8 was initially postulated to be a demethylase (85), it was later shown to act as a hydroxylase that modifies the product of the MT domain, from mcm^5U to *S*-5-(methoxycarbonylhydroxymethyl)uracil (*S*- mchm^5U) (Figure 1 - 12) (85,86). This discovery was inspired by the identification of 5-carboxyhydroxymethyluridine (mchm^5U) in tRNA of *Bombyx mori* (87) Although both diastereomers *R*- and *S*- mchm^5U occur in mammalian tRNA (86), ALKBH8 was found to only catalyse the formation of the *S* diastereomer (85,86), which led researchers to speculate that the *R* diastereomer could be catalysed by another Fe(II) and 2OG-dependent oxygenase (86).

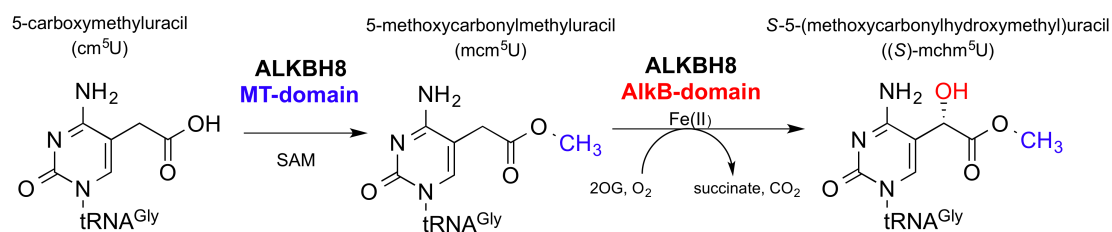


Figure 1 - 12. Reactions catalysed by ALKBH8. Atoms and ALKBH8 domain involved in the methyl transfer reaction are highlighted in blue; atoms and ALKBH8 domain involved in the hydroxylation reaction are highlighted in red.

Consistent with its biochemical function, ALKBH8 localises to the cytoplasm of the cell (72,83). Notably, ALKBH8 overexpression was observed in many cancer cell lines (88). siRNA knockdown of ALKBH8 showed that ALKBH8 is required for bladder cancer progression (88).

A crystal structure of ALKBH8 shows that the RRM and the AlkB domains of ALKBH8 form a cavity, presumably for the binding of tRNA substrate (Figure 1 - 13) (89). ALKBH8 also has a Zn-binding structural motif at the C-terminus of the AlkB domain that has a stabilisation function and this motif is not present in the other ALKBHs (89).

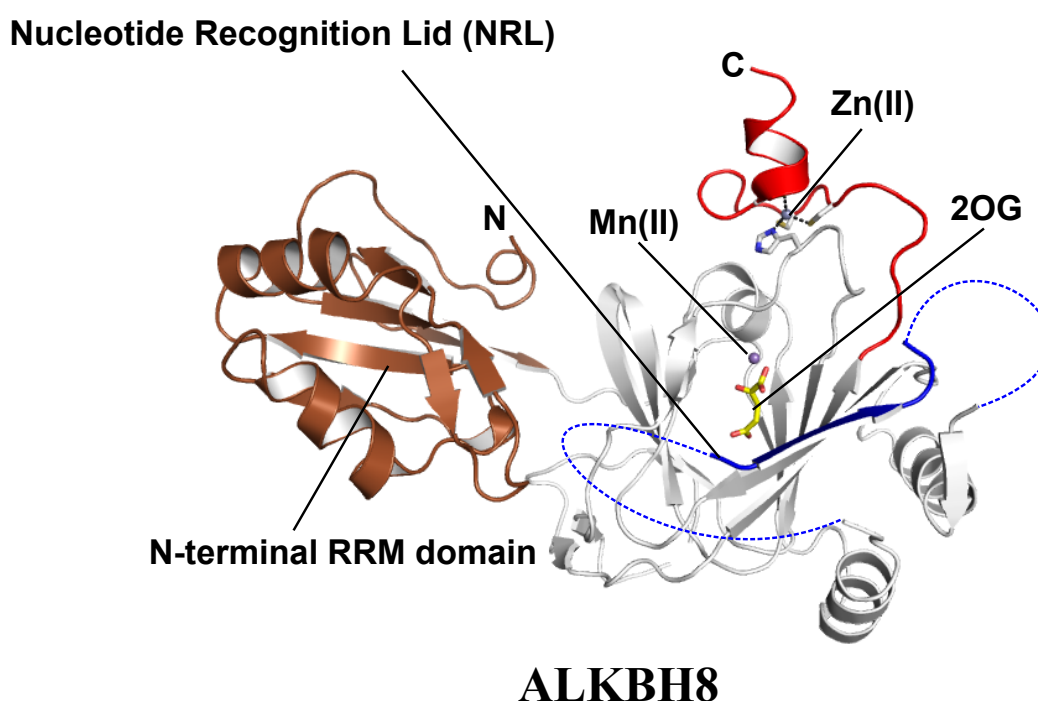


Figure 1 - 13. View from a structure of ALKBH8. Cartoon representation of a crystal structure of human AlkB homologue 8 (ALKBH8) (PDB ID 3THT; resolution 3.01 Å). Color coding: N-terminal nucleotide recognition lid (NRL), blue; N-terminal RNA recognition motif (RRM) domain, brown; C-terminal extension, red; 2-oxoglutarate (2OG), yellow; Mn(II), purple sphere; Zn(II), grey sphere. Note that the majority of the ALKBH8 NRL is disordered (blue dashes). Protein structure models were produced using MacPyMOL v1.3 (10).

FTO

The fat mass and obesity associated protein (FTO) is known for its links to obesity in the human population (90,91). Also known as ALKBH9, FTO was identified as the ninth member of the AlkB subfamily (92,93). The *Fto* gene (originally named Fatso) was first identified as part of a deletion mutant called *Ft* in mice that led to fused toe phenotype (94). Genes affected by the *Ft* mutation include *Fts*, *Ftm*, *Fto*, *Irx3*, *Irx5* and *Irx6* (94). The human *FTO* gene subsequently received considerable attention when several genome-wide association studies (GWAS) concluded that single nucleotide polymorphisms (SNPs) in the first intron of the *FTO* gene is associated with type 2 diabetes, but its association is mediated by body mass index (BMI) (90,91,95). Specifically, homozygous individuals carrying the risk allele A (adenine) for one particular SNP, rs9939609, were found to be on average 3 kg heavier than homozygous individuals carrying the non-risk allele T (thymine) (90).

FTO is most highly expressed in the brain especially in the hypothalamus, which is important in the control of energy homeostasis (92). *Fto* knockout mice exhibit higher frequency of postnatal lethality and leanness (96). The lean phenotype is attributed to increased energy expenditure of the mice, demonstrating that alteration of *Fto* levels have a direct effect on energy homeostasis (96). Conversely, overexpression of *Fto* resulted in an increase in overall body weight of mice, especially in fat mass, as a result of increased food intake (97). In humans, homozygotes of the AA risk allele of rs9939609 SNP in the *FTO* gene showed increased food intake and preference for high fat food (98). This phenotype is potentially mediated by ghrelin, a hunger stimulating peptide, the levels of which are regulated by FTO (98). Recently, the intron of *FTO* was

reported to be a long-range enhancer of IRX3 (99). IRX3 expression was then shown to affect body mass and composition, suggesting that the link between *FTO* and body mass may be more complex than expected (99). Besides obesity, *FTO* has been also reported to be associated with melanoma (100).

The *FTO* gene encodes for an Fe(II) and 2OG dependent oxygenase that was initially shown to demethylate *N*³-methylthymine (m³T) and *N*³-methyluracil (m³U) in ssDNA and ssRNA (Figure 1 - 14) (92,101). Subsequently, *N*⁶-methyladenine (m⁶A) in RNA (Figure 1 - 14) was discovered to be a potential physiological substrate of FTO (102). Immunofluorescence studies revealed that FTO localises in nuclear speckles, and partially colocalises with splicing factor SC35, suggesting that FTO might play a role in the regulation of mRNA processing (102).

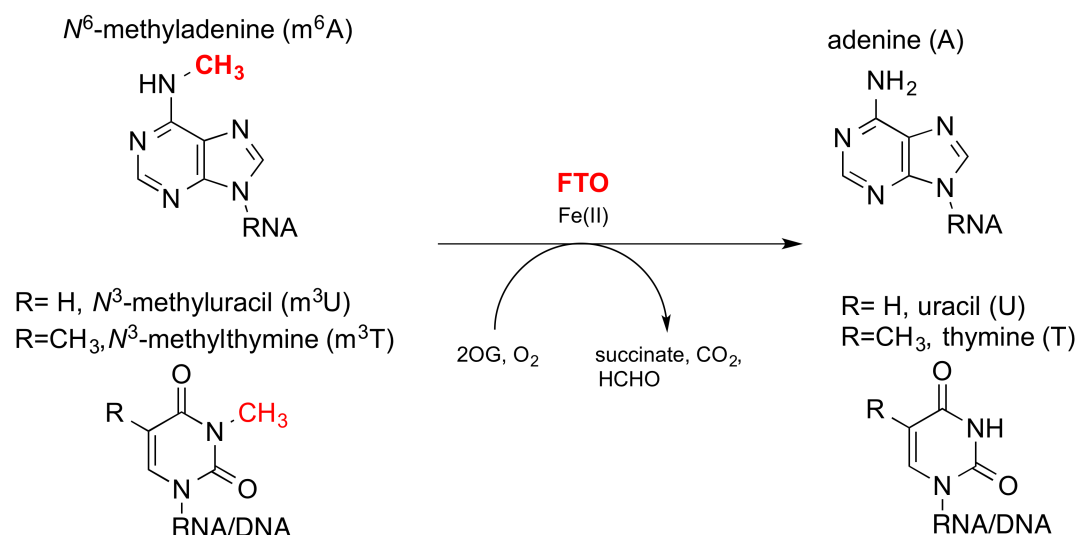


Figure 1 - 14. Reactions catalysed by FTO.

m⁶A levels were later shown to positively regulate mRNA degradation through an m⁶A reader, YTHDF2 (103). Interestingly, Karra *et al.* reported that

overexpression of FTO in MGN3-1 and HEK293T cells led to a reduction in m⁶A levels on ghrelin mRNA and a concomitant increase in ghrelin mRNA, supporting the finding by Wang *et al.* that a decrease in m⁶A levels increases the stability of mRNA (98,103). Taken together, these findings indicate that FTO could be regulating food intake and energy metabolism via alteration of target mRNA levels through the demethylation of m⁶A.

Similar to the ten-eleven-translocation enzymes (TETs) (see below), FTO can also catalyse the formation of N⁶-hydroxymethyladenine (hm⁶A) and sequential oxidation of hm⁶A to N⁶-formyladenine (f⁶A) (Figure 1 - 15) (104). Both hm⁶A and f⁶A decompose in water to form adenine. Whether these intermediates are physiologically relevant remains to be further investigated.

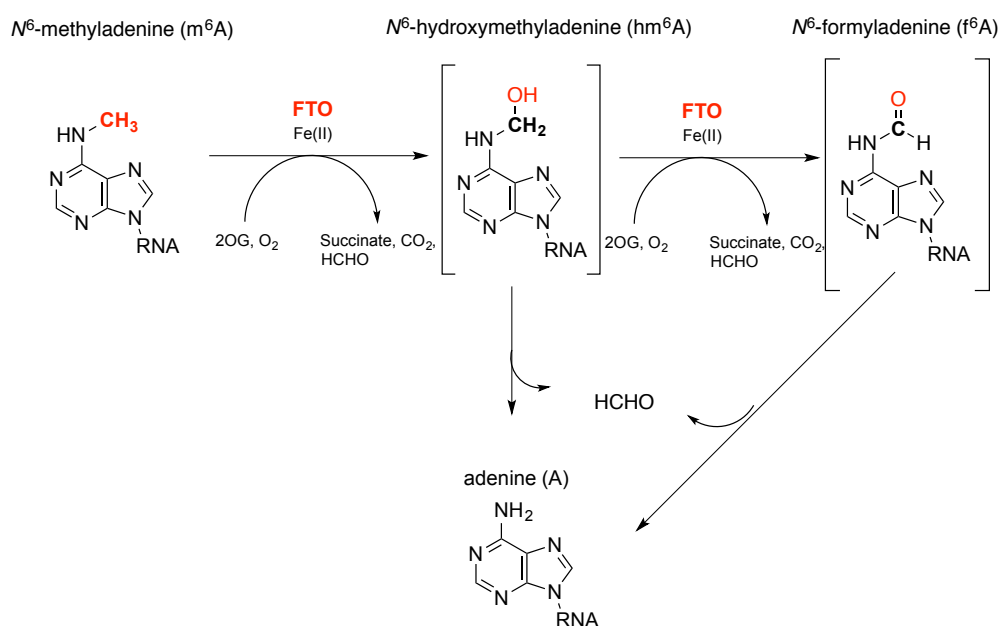


Figure 1 - 15. FTO catalyses the formation of N⁶-hydroxymethyladenine (hm⁶A) and N⁶-formyladenine (f⁶A). hm⁶A and f⁶A decompose in aqueous solution to form adenine and formaldehyde.

A crystal structure of FTO was determined in 2010 and showed that FTO consists of an N-terminal AlkB-like domain that contains the DSBH fold, and a C-

terminal five helical domain of unknown function (Figure 1 - 16) (105). While the actual function of the FTO C-terminal domain is not known, its deletion was shown to impair catalytic activity (106). Like the other AlkB subfamily proteins, FTO possesses an N-terminal NRL for nucleic acid recognition. A loop between DSBH β I and II, termed the L1 loop, is predicted to confer single-stranded nucleic acid selectivity (105).

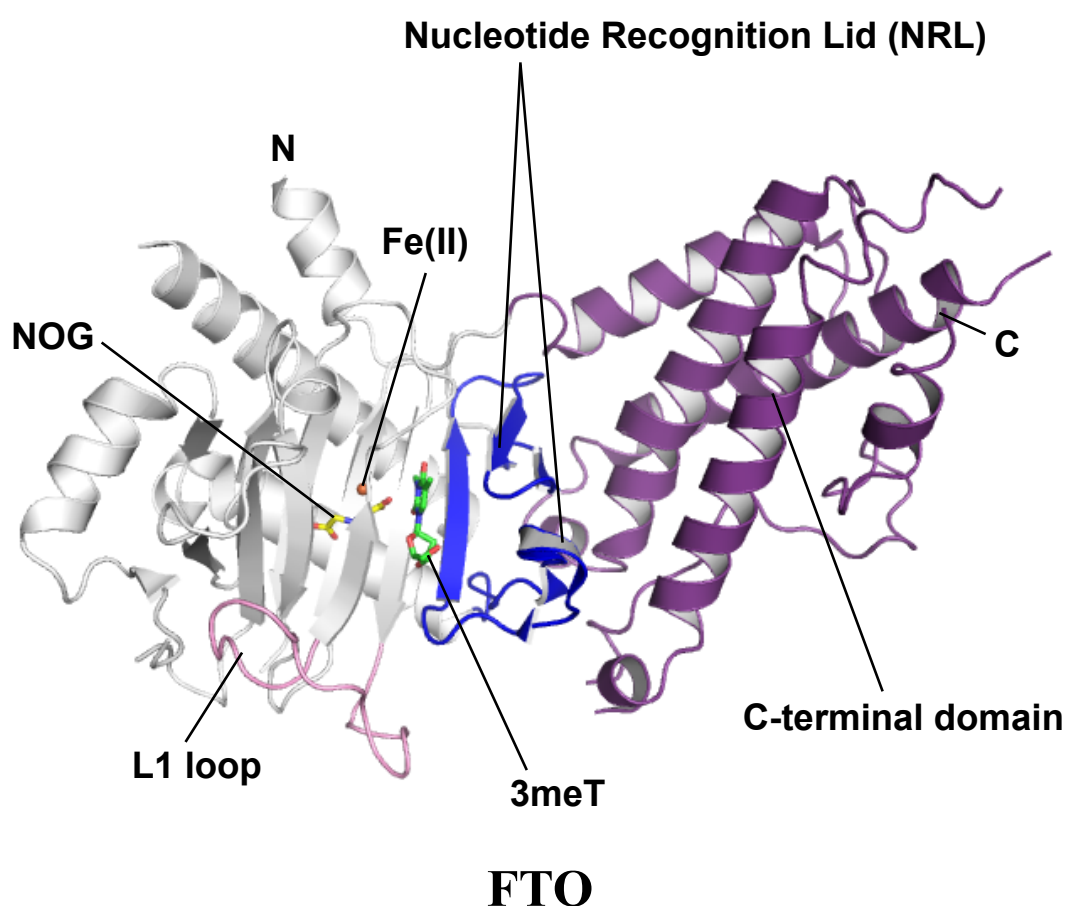


Figure 1 - 16. Structure of FTO. Cartoon representations of the crystal structures of the human fat mass and obesity associated protein (FTO) (PDB ID 3LFM; resolution: 2.50 Å). Colour coding: N-terminal nucleotide recognition lid (NRL), blue; β IV-V loop, orange; L1 loop (β I-II loop) of FTO suggested to confer single-stranded nucleic acid selectivity, pink; C-terminal domain of FTO, purple; Fe(II) ion, orange sphere; *N*-oxalylglycine (NOG), yellow sticks; *N*³-methylthymidine (3meT) substrate, green sticks. Protein structure models were produced using MacPyMOL v1.3 (10).

FTO is viewed as a potential anti-obesity target and it calls for medicinal chemistry efforts for its inhibition (46). Chapter 3 describes structure-based inhibition of FTO to identify classes of inhibitors that can potentially be further developed into functional probes and anti-obesity therapy.

1.2.3 TETs

The human ten-level-translocation enzymes (TET1–3) were first identified as DNA 5-methylcytosine (5mC) hydroxylases that catalyse hydroxylation of 5mC on CpG islands to 5-hydroxymethylcytosine (5hmC) (107). The TETs were shown to also catalyse subsequent oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) (Figure 1 - 17) (108,109). 5caC can then be repaired by thymine DNA glycosylase (TDG) by base excision, thus reversing methylation on cytosine (108). Another proposed route for reversal of 5hmC to cytosine is via deamination of 5hmC to form 5-hydroxyuracil, which can then be removed by DNA glycosylases such as TDG and single-stranded selective monofunctional uracil-DNA glycosylase (SMUG1); cytosine is then reinstalled through the base excision repair pathway (110). TETs are implicated in stem cell differentiation and cancer through their roles in altering the epigenetic markers consisting of various cytosine derivatives (46,111).

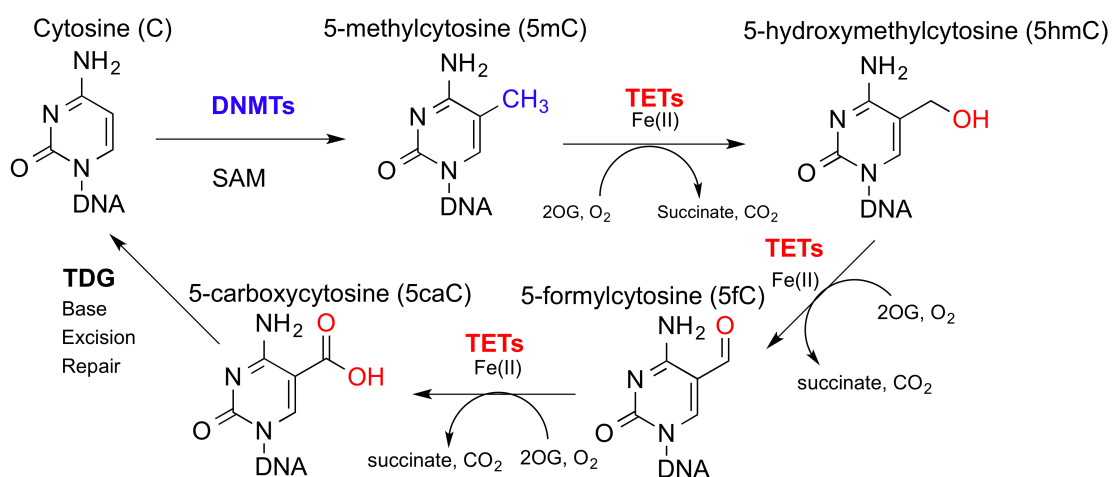


Figure 1 - 17. Reactions catalysed by TETs and reversal of methylcytosine (5mC) to cytosine (C). Atoms and enzymes involved in oxygenase-catalysed reactions (hydroxylation/oxidation) are highlighted in red; atoms and enzymes involved in methyl transfer reaction are highlighted in blue.

Recently, crystal structures of a human TET2 (hTET2)-dsDNA and a *Naegleria gruberi* TET-like enzyme (NgTET1)-dsDNA complexes were determined (Figure 1 - 18) (112,113). The N-terminal extension β -strand(s) of hTET2 and NgTET1 at the active site opening, together with their neighbouring loops, form the key substrate recognition element that contacts the DNA minor groove to flip the 5mC base into the active site.

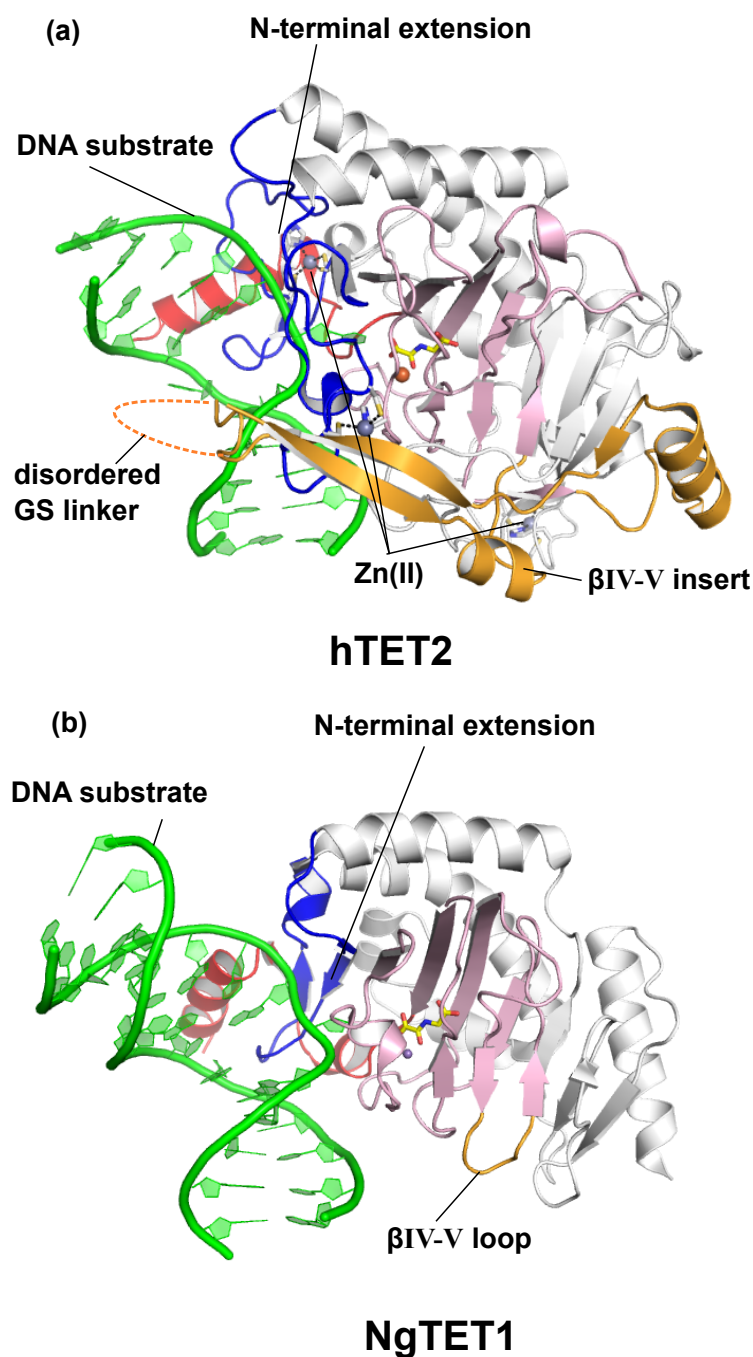


Figure 1 - 18. Views from crystal structures of human TET2 and *Naegleria gruberi* TET1. Cartoon representations of crystal structures of (a) the human TET2 (hTET2)-DNA complex (PDB ID 4NM6; resolution: 2.03 Å) and (b) the *Naegleria gruberi* TET1 (NgTET1)-DNA complex (PDB ID 4LT5; resolution: 2.89 Å). Colour coding: N-terminal extension loop, blue; DSBH core, pink; βIV-V insert, orange; C-terminal extension, red; Fe(II), orange sphere; Mn(II), purple sphere; Zn(II), grey sphere; 2-oxoglutarate (2OG), yellow sticks; DNA substrate, green. Disordered GS-linker of hTET2 is indicated by dashes. Protein structure models were produced using MacPyMOL v1.3 (10).

One other striking feature of the structure of the hTET2 (113), as previously predicted from its sequence (15), is the extensive β IV-V insert that is partially disordered. The ordered region consists of mixed β -strands and α -helices that flank an anti-parallel β -hairpin projecting towards the DNA substrate. An artificial glycine-serine (GS) linker replacing the actual sequence forms the disordered region, which is situated at the apex of the β -hairpin. However, the β IV-V insert of NgTET1 only consists of a relatively short loop of four amino acids.

The human TET2 structure also has three Zn-binding sites at the N-terminal extension: two of them flank the active site opening, apparently stabilising the N-terminal β -strands and loops that are important for substrate interaction (113). However, the NgTET1 structure does not contain any Zn-binding sites (112).

1.2.4 TYW5

The human tRNA wybutosine-synthesizing protein 5 (TYW5) is the only member of the JmjC oxygenase that is known to act on nucleic acid (8,114). TYW5 hydroxylates 7-(α -amino- α -carboxypropyl)wyosine (yW-72), a wybutosine precursor (Figure 1 - 19), in the biosynthesis process of hydroxywybutosine, which is found near the wobble uridine base of tRNA^{Phe} (114).

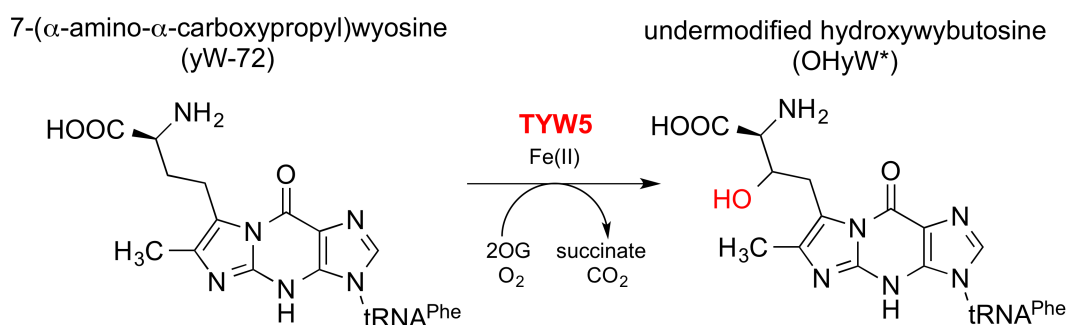


Figure 1 - 19. The reaction catalysed by TYW5.

A crystal structure of TYW5 (Figure 1 - 20) showed that it dimerises via its C-terminal helices and has a JmjC oxygenase architecture including the use of a lysine (Lys175) from DSBH β IV for 2OG binding (115).

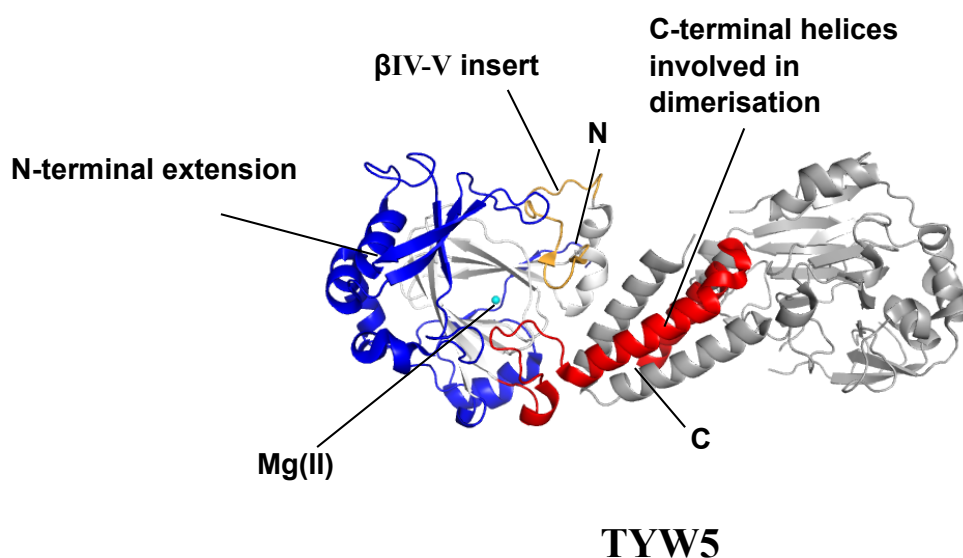


Figure 1 - 20. View from a crystal structure of TYW5. Cartoon representations of a crystal structure of human tRNA wybutosine-synthesizing protein 5 (TYW5) (PDB ID 3AL5; resolution: 2.50 Å). Colour coding: N-terminal extension, blue; β IV-V insert, orange; C-terminal extension dimerisation helices, red; neighbouring monomer molecule that is part of a dimer, grey; Mg(II) ion, cyan sphere.

1.3 2OG Oxygenases Involved in Collagen Biosynthesis

2OG oxygenases also catalyse post-translational modifications of collagen (3,116), a superfamily of proteins that plays a mainly structural role in various

mammalian tissues (117). Collagen undergoes multiple post-translational modifications such as prolyl-4-hydroxylation, prolyl-3-hydroxylation, lysyl hydroxylation and glycosylation during the collagen biosynthesis process (117). These processes occur in the lumen of the endoplasmic reticulum (ER) of the cell. Three subfamilies of 2OG oxygenases are involved in modifying collagens: the collagen prolyl-4-hydroxylases (CPHs), the collagen prolyl-3-hydroxylases (lepreicans/P3Hs), and the procollagen lysyl hydroxylases (PLODs/LHs) (117). 4-hydroxyprolyl residues are important for the stabilisation of the triple helical structure of collagen; and hydroxylysines serve as glycosylation sites and are also important for intermolecular cross-links of collagen (116). The function of 3-hydroxyprolyl residues on collagen is still unclear (118), but it has been demonstrated to destabilise collagen triple helix (119,120).

1.3.1 Collagen Prolyl-4-Hydroxylase

Collagen prolyl-4-hydroxylases (CPHs) catalyse the hydroxylation of proline residues in the position Y of the Gly-X-Y repeated motif in collagen, forming 4-hydroxyprolines (Figure 1 - 21) (121). The X position of the Gly-X-Y motif is usually a proline residue. There are at least 3 isoforms of CPH (CPH1, CPH2, and CPH3) in humans and mice (121). CPH (α subunit) and protein disulphide isomerase (β subunit) form an $\alpha_2\beta_2$ heterotetramer, which has a molecular weight of ~240 kDa (122,123).

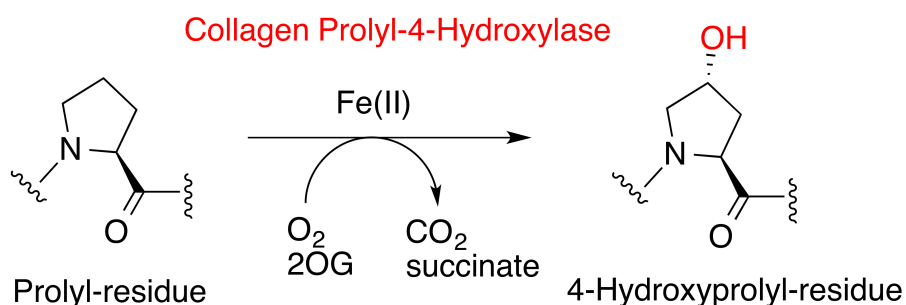


Figure 1 - 21. Reaction catalysed by collagen prolyl-4-hydroxylase.

The activity of CPH was first reported in 1965, in which ascorbate and Fe(II) were found to be requirements for CPH activity in chick embryo microsomal and soluble fractions (124,125). However, ascorbate and Fe(II) alone were not sufficient for *in vitro* activity using partially purified CPH from chick embryo. Hutton *et al.* subsequently demonstrated that 2OG is the other component required for the *in vitro* activity of CPH (126). This important discovery demonstrated that CPH is a 2OG-dependent enzyme and thus marked the first identification of a Fe(II) and 2OG dependent oxygenase.

Most reported CPH activity assays rely on 2OG decarboxylation activity (i.e. detecting the release of $^{14}\text{CO}_2$) of partially purified cell lysates containing CPH for activity quantification (116). Kinetic studies have been carried out using the method stated above and have demonstrated that the K_m of collagen peptides is dependent on peptide length, with a lower K_m for longer peptide chain length (116,121-123,127). For example, for human CPH1, the K_m values for peptides with the sequence PPG, (PPG)₅ and (PPG)₂₀ are 20 mM, 2 mM and 50 μM respectively (121-123). Other types of assays used to quantify the hydroxylation activity of CPH include the detection of hydroxyl[^{14}C]proline, $^3\text{H}_2\text{O}$ release, and O_2 uptake (116). Expression of active CPH was also reported to be achievable by *E. coli* expression using the Origami strain that lacks thioredoxin, resulting in an

oxidising cellular environment to enhance disulphide bond formation (128). Furthermore, gene optimisation is shown to be able to boost the yield of CPH in *E. coli* expression (129).

Recently, CPH was reported to promote breast cancer tumour metastasis (130). CPH expression was induced through the HIF pathway in breast cancer cells, resulting in deposition of collagen fibers that are essential for invasion and metastasis (130). This result was further supported by the reduction in tumour fibrosis and metastasis when mice were treated with a prolyl hydroxylase inhibitor, 3,4-dihydroxybenzoate (130). Therefore, inhibition and structural investigations of CPH are potentially of therapeutic significance.

A structure of the $\alpha_2\beta_2$ tetramer complex of CPH has yet to be determined. The α subunit (517 residues) of CPH is divided into the N-terminal double domain (residues 1 – 244) and the C-terminal catalytic domain (CAT) (residues 245 – 517), which contains the DSBH fold (131-133). The N-terminal double domain can be further divided into the N-terminal dimerisation domain (residues 1 – 143) and the peptide-substrate-binding domain (PSB) (residues 144 – 244) (131,132,134). Structures of the N-terminal double domain have been determined recently, showing a dimer formed via the two α -helices of the N-terminal dimerisation domain (Figure 1 - 22a) (134). The PSB domains, which adopt a tetratricopeptide repeat (TPR) fold (135), are positioned at the two ends of the dimer. Overall, the N-terminal double domain is shaped like the letter 'Z' (134). However, the structure of the CAT domain of human CPH remains unsolved. Nevertheless, the structures of homologous CPH CAT domain from *Chlamydomonas reinhardtii* (CrCPH) and *Bacillus anthracis* (BaCPH) have been reported to contain the conserved DSBH fold of 2OG oxygenases (Figure 1 - 22c

and d) (136-138). Additionally, the structure of CrCPH in complex with a proline-serine peptide substrate fragment revealed that the β -hairpin from the N-terminal extension and the DSBH β II-III loop enclose the substrate (136), reminiscent of the hypoxia inducible factor prolyl-4-hydroxylase (PHD2) (Figure 1 - 22e) (11). Thus it is probable that the human CPH CAT domain adopts similar substrate binding elements.

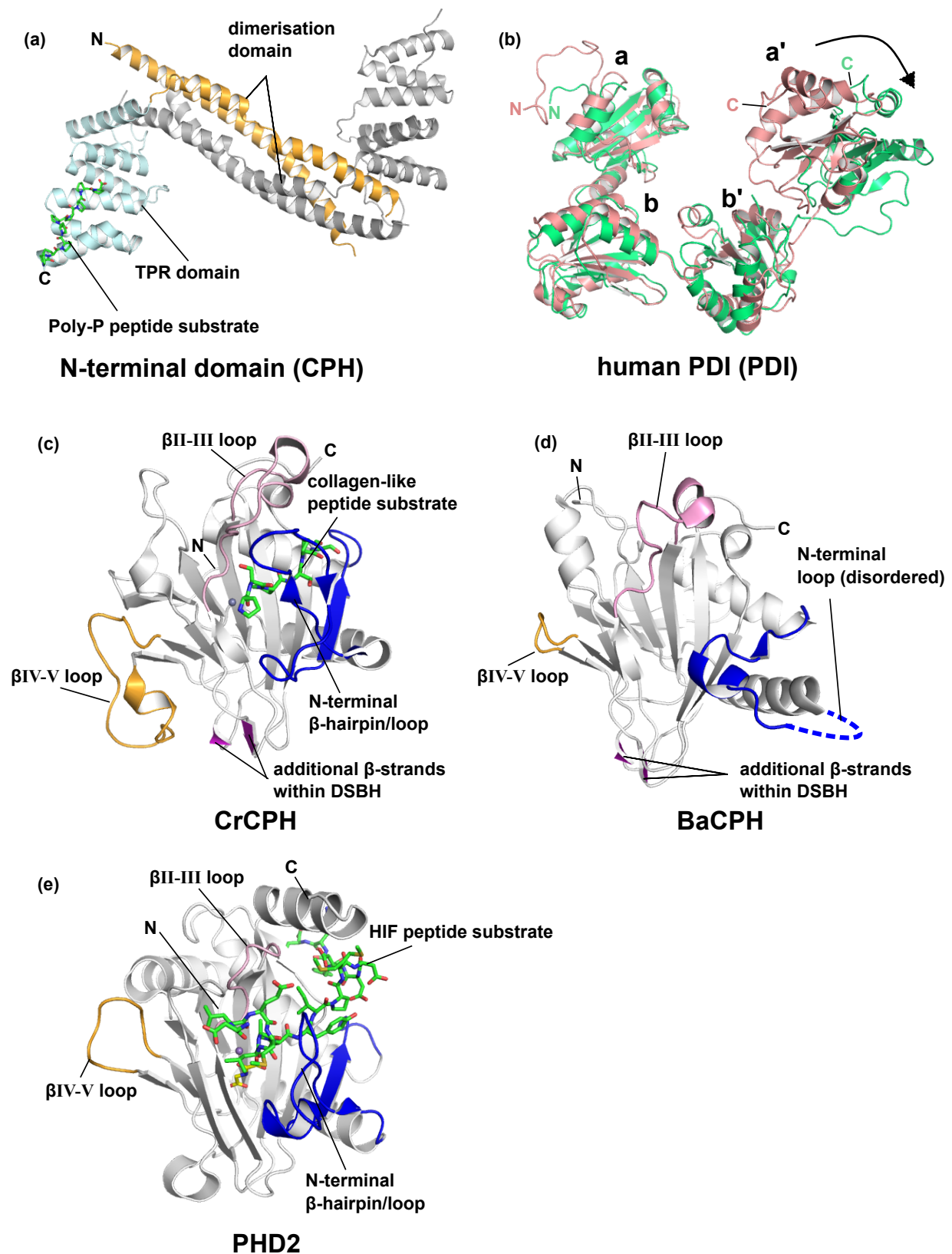


Figure 1 - 22. Views from crystal structures of prolyl-4-hydroxylases and protein disulphide isomerase. (a) Crystal structure of the dimer of the N-terminal double domain of human collagen prolyl-4-hydroxylase (PDB ID 4BTB; resolution: 1.90 Å). Colour coding: N-terminal dimerisation domain, orange; peptide substrate binding domain (TPR), light cyan; collagen-peptide substrate fragment, green sticks; neighbouring monomer molecule, grey. (b) Superimposition of views from crystal structures of the reduced

(salmon) (PDB ID 4EKZ; resolution: 2.51 Å) and oxidised (lime green) (PDB ID 4EL1) β -subunit (human protein disulphide isomerase, PDI) of human collagen prolyl hydroxylase tetramer. Note that domain a' of the oxidised PDI undergoes a conformational change relative to the reduced form (indicated by a black arrow). Cartoon representations of crystal structures of **(c)** the *Chlamydomonas reinhardtii* prolyl-4-hydroxylase (CrCPH)-peptide substrate complex (PDB ID 3GZE; resolution: 1.98 Å), **(d)** the *Bacillus anthracis* prolyl-4-hydroxylase (PDB ID 3ITQ; resolution: 1.40 Å) and **(e)** the human HIF prolyl hydroxylase 2 (PHD2)-HIF peptide substrate complex (PDB ID 3HQR; resolution 2.00 Å). Colour coding: N-terminal extension β -hairpin/loop, blue; β II-III loop, pink; β IV-V loop, orange; additional β -strands within the DSBH core, magenta; Mn(II) ion, purple sphere; 2-oxoglutarate (2OG), yellow sticks; peptide substrate, green sticks. Disordered regions are indicated by dashes. Protein structure models were produced using MacPyMOL v1.3 (10).

A structure of PDI (β -subunit of the CPH tetramer) showed that it has four thioredoxin fold homologous domains, named **a**, **b**, **b'** and **a'** (from N to C-termini) (Figure 1 - 22b) (139-141). Domains **a** and **a'** contain the active site residue motif CGHC but domains **b** and **b'** lack the active site motif (142). Crystal structures of both the reduced and oxidised human PDI revealed a conformational change involving a 45° rotation of domain a' around the **x** linker, which is a linker region between domains **b'** and **a'** (143).

1.3.2 Procollagen Lysyl Hydroxylases

The procollagen lysyl hydroxylases (PLODs) hydroxylate lysine residues at the Y position of the collagen Gly-X-Y motif. There are three identified isoforms of procollagen lysyl hydroxylases (PLOD1, PLOD2 and PLOD3) (117). PLOD2 can be alternatively spliced in two forms, PLOD2a and PLOD2b. PLOD1 acts on lysines in the α -helical region of collagen while PLOD2 is a collagen telopeptide lysyl hydroxylase (144). However, the specificity of PLOD3 is unknown. PLOD1 is localised in the lumen of ER, while PLOD3 is observed to localise in the ER as a membrane-associated form as well as in the extracellular space (145-148).

In addition to their role in intermolecular crosslinks of collagen, hydroxylysines on collagen also serve as carbohydrate attachment sites. PLOD3

was later found to also have collagen glucosyltransferase (GGT) and galactosyltransferase (GT) activities (149-151), thus implying that PLOD3 is a multifunctional enzyme capable of catalysing the formation of hydroxylysine, then galactosylhydroxylysine, and finally glucosylgalactosylhydroxylysine (Figure 1 - 23).

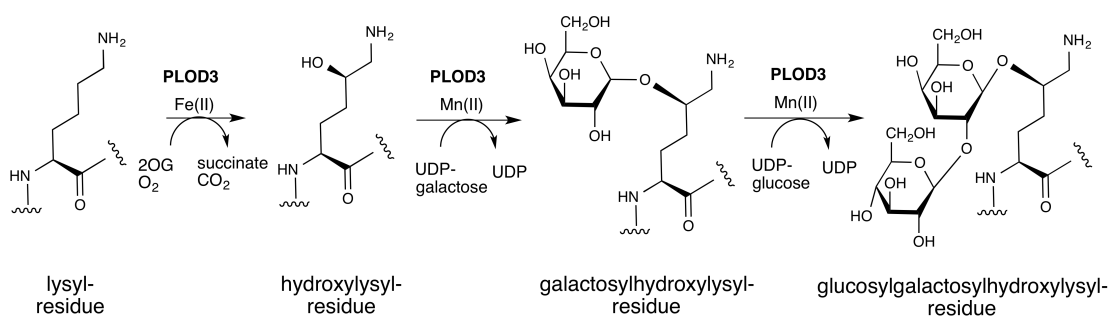


Figure 1 - 23. Reactions catalysed by PLOD3. PLOD3 has lysyl hydroxylase, galactosyltransferase, and glucosyltransferase activities.

The PLODs are implicated in several diseases. Mutation in the *PLOD1* gene results in the Ehlers-Danlos syndrome type VI, a condition caused by an absence of hydroxylysine in collagen triple helix (152). Misregulation of PLOD2 is suggested to cause the Bruck Syndrome, a result of an over-hydroxylation of collagen telopeptide lysine residues i.e. fibrosis (144). Like the CPHs, PLOD1 and PLOD2 are upregulated in human breast cancer cells via the HIF pathway and are essential for metastasis (153).

The closest relative of the human PLODs is the 2-oxoglutarate and iron-dependent oxygenase domain-containing 2 protein (OGFOD2), which shows 25-30% sequence identity with the PLODs (51,154). Whether or not the function of OGFOD2 is related to post-translational modification of collagen or collagen-related protein remains to be investigated. Crystal structures of members of the

PLOD subfamily have yet to be reported. Chapter 5 of this thesis describes preliminary functional and structural characterisation of OGFOD2.

1.3.3 Collagen Prolyl-3-Hydroxylase

The third subfamily of 2OG oxygenase involved in collagen biosynthesis is the collagen prolyl-3-hydroxylases (leprecans/P3Hs) (155). Leprecans specifically catalyse 3-hydroxylation of a specific few prolines (Figure 1 - 24) on the X position of the Gly-X-Y motif (155), in contrast with CPHs and the PLODs that hydroxylate a proline or lysine residue on the Y position. There are also three isoforms for the leprecans (P3H1, P3H2, and P3H3)(155). Mutations on *P3H1* cause osteogenesis imperfecta in both human and mice (156,157).

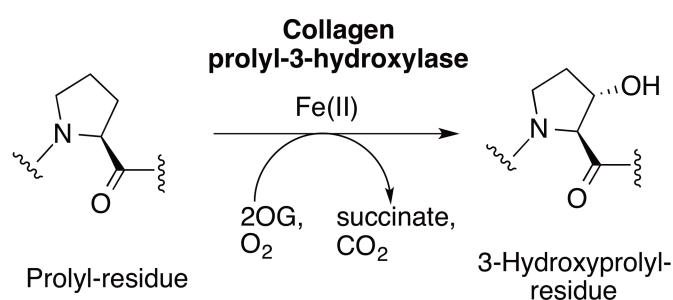


Figure 1 - 24. Reaction catalysed by collagen prolyl-3-hydroxylases (leprecans).

1.4 Objectives of the Research Described in this Thesis

As described in the previous sections (Sections 1.2 and 1.3), several NAOX and collagen modifying oxygenases (e.g. ALKBH2, ALKBH3, ALKBH8, FTO, CPH and PLODs) are upregulated in diseases such as cancer and obesity (46,130,153). Thus inhibition of these 2OG oxygenases may present potential therapeutic opportunities. In view of the need for inhibition of 2OG oxygenases, structural

studies of 2OG oxygenases are important because crystal structures of 2OG oxygenase-ligand complexes will aid in inhibitor design.

This thesis describes structural and inhibition studies concerning 5 different 2OG oxygenases – AlkB, FTO, ALKBH5, OGFOD2 and CPH, with a focus on determining their structures in complex with inhibitors using X-ray crystallography. The first part of the thesis (Chapters 2 – 4) focuses on structure-based inhibition studies of AlkB, FTO, and ALKBH5, with the overall aim to develop inhibitors for FTO, which is associated with obesity in the human population (90,91). The objectives of the work detailed in the chapters of Part 1 are as follows:

- (i) Chapter 2 – to determine structures of the AlkB-inhibitor complexes as proof-of-principle that the AlkB subfamily proteins can be subjected to structure-based inhibitor design.
- (ii) Chapter 3 – to apply structure-based inhibition studies on the human fat mass and obesity associated protein FTO by developing an inhibition assay, and to determine crystal structures of FTO in complex with inhibitors.
- (iii) Chapter 4 – to determine a structure of human ALKBH5 to aid in selective inhibitor development against FTO.

In light of emerging evidence that 2OG oxygenases involved in collagen biosynthesis promote cancer metastasis (130,153), structural and biochemical studies on collagen-modifying 2OG oxygenases and related enzymes were also carried out. The second part of this thesis (Chapters 5 and 6) describes preliminary structural and functional studies on OGFOD2 and CPH, both related

to collagen biosynthesis (OGFOD2 is related to the procollagen lysyl hydroxylases, PLODs). The objectives of the work detailed in Part 2 are:

- (iv) Chapter 5 – to investigate the structure and function of OGFOD2 to explore the potential roles of a novel 2OG oxygenase that is related to the PLODs.
- (v) Chapter 6 – to work towards obtaining a crystal structure of the $\alpha_2\beta_2$ tetramer complex of CPH and PDI, and to develop an *in vitro* biochemical assay to quantify the enzymatic activity of CPH. The preliminary structural and biochemical work is aimed to pave way for inhibition studies of CPH in the future.

1.5 References

1. Clifton, I.J., McDonough, M.A., Ehrismann, D., Kershaw, N.J., Granatino, N. and Schofield, C.J. (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem*, **100**, 644-669.
2. Hausinger, R.P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol*, **39**, 21-68.
3. Loenarz, C. and Schofield, C.J. (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci*, **36**, 7-18.
4. Schofield, C.J. and Zhang, Z. (1999) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol*, **9**, 722-731.
5. McDonough, M.A., Loenarz, C., Chowdhury, R., Clifton, I.J. and Schofield, C.J. (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struct Biol*, **20**, 659-672.
6. Aik, W., McDonough, M.A., Thalhammer, A., Chowdhury, R. and Schofield, C.J. (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struct Biol*, **22**, 691-700.
7. Rose, N.R., McDonough, M.A., King, O.N.F., Kawamura, A. and Schofield, C.J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev*, **40**, 4364-4397.

8. Iyer, L.M., Abhiman, S., de Souza, R.F. and Aravind, L. (2010) Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. *Nucleic Acids Res*, **38**, 5261-5279.
9. Yang, C.G., Yi, C., Duguid, E.M., Sullivan, C.T., Jian, X., Rice, P.A. and He, C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature*, **452**, 961-965.
10. Schrodinger, LLC. (2010) The PyMOL Molecular Graphics System, Version 1.3r1.
11. Chowdhury, R., McDonough, M.A., Mecinovic, J., Loenarz, C., Flashman, E., Hewitson, K.S., Domene, C. and Schofield, C.J. (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure*, **17**, 981-989.
12. Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y. and Cheng, X. (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol*, **17**, 38-43.
13. Ng, S.S., Kavanagh, K.L., McDonough, M.A., Butler, D., Pilka, E.S., Lienard, B.M., Bray, J.E., Savitsky, P., Gileadi, O., von Delft, F. *et al.* (2007) Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature*, **448**, 87-91.
14. Elkins, J.M., Hewitson, K.S., McNeill, L.A., Seibel, J.F., Schlemminger, I., Pugh, C.W., Ratcliffe, P.J. and Schofield, C.J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem*, **278**, 1802-1806.
15. Iyer, L.M., Tahiliani, M., Rao, A. and Aravind, L. (2009) Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle*, **8**, 1698-1710.
16. Zhang, Z., Ren, J., Stammers, D.K., Baldwin, J.E., Harlos, K. and Schofield, C.J. (2000) Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat Struct Biol*, **7**, 127-133.
17. O'Brien, J.R., Schuller, D.J., Yang, V.S., Dillard, B.D. and Lanzilotta, W.N. (2003) Substrate-induced conformational changes in Escherichia coli taurine/alpha-ketoglutarate dioxygenase and insight into the oligomeric structure. *Biochemistry*, **42**, 5547-5554.
18. Kruidenier, L., Chung, C.W., Cheng, Z., Liddle, J., Che, K., Joberty, G., Bantscheff, M., Bountra, C., Bridges, A., Diallo, H. *et al.* (2012) A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature*, **488**, 404-408.
19. Sengoku, T. and Yokoyama, S. (2011) Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev*, **25**, 2266-2277.
20. Chen, Z., Zang, J., Whetstone, J., Hong, X., Davrazou, F., Kutateladze, T.G., Simpson, M., Mao, Q., Pan, C.H., Dai, S. *et al.* (2006) Structural insights into histone demethylation by JMJD2 family members. *Cell*, **125**, 691-702.
21. Chen, Z., Zang, J., Kappler, J., Hong, X., Crawford, F., Wang, Q., Lan, F., Jiang, C., Whetstone, J., Dai, S. *et al.* (2007) Structural basis of the recognition of a methylated histone tail by JMJD2A. *Proc Natl Acad Sci U S A*, **104**, 10818-10823.

22. Zhang, Z., Ren, J.S., Clifton, I.J. and Schofield, C.J. (2004) Crystal structure and mechanistic implications of 1-aminocyclopropane-1-carboxylic acid oxidase--the ethylene-forming enzyme. *Chem Biol*, **11**, 1383-1394.
23. Higgins, L.J., Yan, F., Liu, P.H., Liu, H.W. and Drennan, C.L. (2005) Structural insight into antibiotic fosfomycin biosynthesis by a mononuclear iron enzyme. *Nature*, **437**, 838-844.
24. Roach, P.L., Clifton, I.J., Fulop, V., Harlos, K., Barton, G.J., Hajdu, J., Andersson, I., Schofield, C.J. and Baldwin, J.E. (1995) Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature*, **375**, 700-704.
25. Blasiak, L.C., Vaillancourt, F.H., Walsh, C.T. and Drennan, C.L. (2006) Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature*, **440**, 368-371.
26. Hewitson, K.S., Holmes, S.L., Ehrismann, D., Hardy, A.P., Chowdhury, R., Schofield, C.J. and McDonough, M.A. (2008) 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.
27. Balciunas, D. and Ronne, H. (2000) Evidence of domain swapping within the jumonji family of transcription factors. *Trends Biochem Sci*, **25**, 274-276.
28. van Staalduinen, L.M., Novakowski, S.K. and Jia, Z. (2014) Structure and Functional Analysis of YcfD, a Novel 2-Oxoglutarate/Fe-Dependent Oxygenase Involved in Translational Regulation in Escherichiacoli. *J Mol Biol*.
29. Chowdhury, R., Sekirnik, R., Brissett, N.C., Krojer, T., Ho, C.H., Ng, S.S., Clifton, I.J., Ge, W., Kershaw, N.J., Fox, G.C. *et al.* (2014) Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature*.
30. Clifton, I.J., Hsueh, L.C., Baldwin, J.E., Harlos, K. and Schofield, C.J. (2001) Structure of proline 3-hydroxylase. Evolution of the family of 2-oxoglutarate dependent oxygenases. *Eur J Biochem*, **268**, 6625-6636.
31. Valegard, K., van Scheltinga, A.C., Lloyd, M.D., Hara, T., Ramaswamy, S., Perrakis, A., Thompson, A., Lee, H.J., Baldwin, J.E., Schofield, C.J. *et al.* (1998) Structure of a cephalosporin synthase. *Nature*, **394**, 805-809.
32. Pavel, E.G., Zhou, J., Busby, R.W., Gunsior, M., Townsend, C.A. and Solomon, E.I. (1998) Circular dichroism and magnetic circular dichroism spectroscopic studies of the non-heme ferrous active site in clavamate synthase and its interaction with alpha-ketoglutarate cosubstrate. *Journal of the American Chemical Society*, **120**, 743-753.
33. Zhou, J., Kelly, W.L., Bachmann, B.O., Gunsior, M., Townsend, C.A. and Solomon, E.I. (2001) Spectroscopic studies of substrate interactions with clavamate synthase 2, a multifunctional alpha-KG-dependent non-heme iron enzyme: Correlation with mechanisms and reactivities. *Journal of the American Chemical Society*, **123**, 7388-7398.
34. Muller, I., Kahnert, A., Pape, T., Sheldrick, G.M., Meyer-Klaucke, W., Dierks, T., Kertesz, M. and Uson, I. (2004) Crystal structure of the alkylsulfatase AtsK: insights into the catalytic

- mechanism of the Fe(II) alpha-ketoglutarate-dependent dioxygenase superfamily. *Biochemistry*, **43**, 3075-3088.
35. Ryle, M.J., Padmakumar, R. and Hausinger, R.P. (1999) Stopped-flow kinetic analysis of *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase: Interactions with alpha-ketoglutarate, taurine, and oxygen. *Biochemistry*, **38**, 15278-15286.
 36. Hegg, E.L., Whiting, A.K., Saari, R.E., McCracken, J., Hausinger, R.P. and Que, L. (1999) Herbicide-degrading alpha-keto acid-dependent enzyme TfdA: Metal coordination environment and mechanistic insights. *Biochemistry*, **38**, 16714-16726.
 37. Ho, R.Y.N., Mehn, M.P., Hegg, E.L., Liu, A., Ryle, M.J., Hausinger, R.P. and Que, L. (2001) Resonance Raman studies of the Iron(II)-alpha-keto acid chromophore in model and enzyme complexes. *Journal of the American Chemical Society*, **123**, 5022-5029.
 38. Zhou, J., Gunsior, M., Bachmann, B.O., Townsend, C.A. and Solomon, E.I. (1998) Substrate binding to the alpha-ketoglutarate-dependent non-heme iron enzyme clavamate synthase 2: Coupling mechanism of oxidative decarboxylation and hydroxylation. *Journal of the American Chemical Society*, **120**, 13539-13540.
 39. Myllyla, R., Tuderman, L. and Kivirikko, K.I. (1977) Mechanism of the prolyl hydroxylase reaction. 2. Kinetic analysis of the reaction sequence. *Eur J Biochem*, **80**, 349-357.
 40. Tuderman, L., Myllyla, R. and Kivirikko, K.I. (1977) Mechanism of the prolyl hydroxylase reaction. 1. Role of co-substrates. *Eur J Biochem*, **80**, 341-348.
 41. Price, J.C., Barr, E.W., Tirupati, B., Bollinger, J.M., Jr. and Krebs, C. (2003) The first direct characterization of a high-valent iron intermediate in the reaction of an alpha-ketoglutarate-dependent dioxygenase: a high-spin FeIV complex in taurine/alpha-ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry*, **42**, 7497-7508.
 42. Proshlyakov, D.A., Henshaw, T.F., Monterosso, G.R., Ryle, M.J. and Hausinger, R.P. (2004) Direct detection of oxygen intermediates in the non-heme Fe enzyme taurine/alpha-ketoglutarate dioxygenase. *Journal of the American Chemical Society*, **126**, 1022-1023.
 43. Price, J.C., Barr, E.W., Glass, T.E., Krebs, C. and Bollinger, J.M., Jr. (2003) Evidence for hydrogen abstraction from C1 of taurine by the high-spin Fe(IV) intermediate detected during oxygen activation by taurine:alpha-ketoglutarate dioxygenase (TauD). *Journal of the American Chemical Society*, **125**, 13008-13009.
 44. Zhang, Z., Ren, J., Harlos, K., McKinnon, C.H., Clifton, I.J. and Schofield, C.J. (2002) Crystal structure of a clavamate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: evidence for metal centered rearrangements. *FEBS Lett*, **517**, 7-12.
 45. Montellano, P.R. (2000) The mechanism of heme oxygenase. *Curr Opin Chem Biol*, **4**, 221-227.
 46. Thalhammer, A., Aik, W., Bagg, E.A.L. and Schofield, C.J. (2012) The potential of 2-oxoglutarate oxygenases acting on nucleic acids as therapeutic targets. *Drug Discovery Today: Therapeutic Strategies*, **9**, e91-e100.
 47. Zheng, G., Fu, Y. and He, C. (2014) Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics. *Chem Rev*.

48. Falnes, P.O., Johansen, R.F. and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature*, **419**, 178-182.
49. Treweek, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature*, **419**, 174-178.
50. Kataoka, H., Yamamoto, Y. and Sekiguchi, M. (1983) A new gene (alkB) of *Escherichia coli* that controls sensitivity to methyl methane sulfonate. *J Bacteriol*, **153**, 1301-1307.
51. Aravind, L. and Koonin, E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol*, **2**, RESEARCH0007.
52. Falnes, P.O., Bjoras, M., Aas, P.A., Sundheim, O. and Seeberg, E. (2004) Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res*, **32**, 3456-3461.
53. Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R. and Hunt, J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879-884.
54. Holland, P.J. and Hollis, T. (2010) Structural and mutational analysis of *Escherichia coli* AlkB provides insight into substrate specificity and DNA damage searching. *PLoS One*, **5**, e8680.
55. Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C.G., Cui, Q. and He, C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*, **468**, 330-333.
56. Woon, E.C., Demetriades, M., Bagg, E.A., Aik, W., Krylova, S.M., Ma, J.H., Chan, M., Walport, L.J., Wegman, D.W., Dack, K.N. *et al.* (2012) Dynamic combinatorial mass spectrometry leads to inhibitors of a 2-oxoglutarate-dependent nucleic acid demethylase. *J Med Chem*, **55**, 2173-2184.
57. Wei, Y.F., Carter, K.C., Wang, R.P. and Shell, B.K. (1996) Molecular cloning and functional analysis of a human cDNA encoding an *Escherichia coli* AlkB homolog, a protein involved in DNA alkylation damage repair. *Nucleic Acids Res*, **24**, 931-937.
58. Aas, P.A., Otterlei, M., Falnes, P.O., Vagbo, C.B., Skorpen, F., Akbari, M., Sundheim, O., Bjoras, M., Slupphaug, G., Seeberg, E. *et al.* (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature*, **421**, 859-863.
59. Westbye, M.P., Feyzi, E., Aas, P.A., Vagbo, C.B., Talstad, V.A., Kavli, B., Hagen, L., Sundheim, O., Akbari, M., Liabakk, N.B. *et al.* (2008) Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J Biol Chem*, **283**, 25046-25056.
60. Muller, T.A., Yu, K., Hausinger, R.P. and Meek, K. (2013) ALKBH1 is dispensable for abasic site cleavage during base excision repair and class switch recombination. *PLoS One*, **8**, e67403.

61. Muller, T.A., Andrzejak, M.M. and Hausinger, R.P. (2013) A covalent protein-DNA 5'-product adduct is generated following AP lyase activity of human ALKBH1 (AlkB homologue 1). *Biochem J*, **452**, 509-518.
62. Ougland, R., Lando, D., Jonson, I., Dahl, J.A., Moen, M.N., Nordstrand, L.M., Rognes, T., Lee, J.T., Klungland, A., Kouzarides, T. *et al.* (2012) ALKBH1 is a histone H2A dioxygenase involved in neural differentiation. *Stem Cells*, **30**, 2672-2682.
63. Nordstrand, L.M., Svard, J., Larsen, E., Nilsen, A., Ougland, R., Furu, K., Lien, G.F., Rognes, T., Namekawa, S.H., Lee, J.T. *et al.* (2010) Mice lacking Alkbh1 display sex-ratio distortion and unilateral eye defects. *PLoS One*, **5**, e13827.
64. Duncan, T., Trewick, S.C., Koivisto, P., Bates, P.A., Lindahl, T. and Sedgwick, B. (2002) Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A*, **99**, 16660-16665.
65. Ringvoll, J., Nordstrand, L.M., Vagbo, C.B., Talstad, V., Reite, K., Aas, P.A., Lauritzen, K.H., Liabakk, N.B., Bjork, A., Doughty, R.W. *et al.* (2006) Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *Embo J*, **25**, 2189-2198.
66. Dango, S., Mosammaparast, N., Sowa, M.E., Xiong, L.J., Wu, F., Park, K., Rubin, M., Gygi, S., Harper, J.W. and Shi, Y. (2011) DNA unwinding by ASCC3 helicase is coupled to ALKBH3-dependent DNA alkylation repair and cancer cell proliferation. *Mol Cell*, **44**, 373-384.
67. Cetica, V., Genitori, L., Giunti, L., Sanzo, M., Bernini, G., Massimino, M. and Sardi, I. (2009) Pediatric brain tumors: mutations of two dioxygenases (hABH2 and hABH3) that directly repair alkylation damage. *J Neurooncol*, **94**, 195-201.
68. Lee, S.Y., Luk, S.K., Chuang, C.P., Yip, S.P., To, S.S. and Yung, Y.M. (2010) TP53 regulates human AlkB homologue 2 expression in glioma resistance to Photofrin-mediated photodynamic therapy. *Br J Cancer*, **103**, 362-369.
69. Sundheim, O., Vagbo, C.B., Bjoras, M., Sousa, M.M., Talstad, V., Aas, P.A., Drablos, F., Krokan, H.E., Tainer, J.A. and Slupphaug, G. (2006) Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *Embo J*, **25**, 3389-3397.
70. Kurowski, M.A., Bhagwat, A.S., Papaj, G. and Bujnicki, J.M. (2003) Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics*, **4**, 48.
71. Li, M.M., Nilsen, A., Shi, Y., Fusser, M., Ding, Y.H., Fu, Y., Liu, B., Niu, Y., Wu, Y.S., Huang, C.M. *et al.* (2013) ALKBH4-dependent demethylation of actin regulates actomyosin dynamics. *Nat Commun*, **4**, 1832.
72. Tsujikawa, K., Koike, K., Kitae, K., Shinkawa, A., Arima, H., Suzuki, T., Tsuchiya, M., Makino, Y., Furukawa, T., Konishi, N. *et al.* (2007) Expression and sub-cellular localization of human ABH family molecules. *J Cell Mol Med*, **11**, 1105-1116.

73. Solberg, A., Robertson, A.B., Aronsen, J.M., Rognmo, O., Sjaastad, I., Wisloff, U. and Klungland, A. (2013) Deletion of mouse Alkbh7 leads to obesity. *J Mol Cell Biol*, **5**, 194-203.
74. Fu, D., Jordan, J.J. and Samson, L.D. (2013) Human ALKBH7 is required for alkylation and oxidation-induced programmed necrosis. *Genes Dev*, **27**, 1089-1100.
75. Thalhammer, A., Bencokova, Z., Poole, R., Loenarz, C., Adam, J., O'Flaherty, L., Schodel, J., Mole, D., Giaslakitotis, K., Schofield, C.J. *et al.* (2011) Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1alpha (HIF-1alpha). *PLoS One*, **6**, e16210.
76. Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vagbo, C.B., Shi, Y., Wang, W.L., Song, S.H. *et al.* (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell*, **49**, 18-29.
77. Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201-206.
78. Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E. and Jaffrey, S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, **149**, 1635-1646.
79. Baltz, A.G., Munschauer, M., Schwanhaussner, B., Vasile, A., Murakawa, Y., Schueler, M., Youngs, N., Penfold-Brown, D., Drew, K., Milek, M. *et al.* (2012) The mRNA-bound proteome and its global occupancy profile on protein-coding transcripts. *Mol Cell*, **46**, 674-690.
80. Feng, C., Liu, Y., Wang, G., Deng, Z., Zhang, Q., Wu, W., Tong, Y., Cheng, C. and Chen, Z. (2014) Crystal structures of the human RNA demethylase alkbh5 reveal basis for substrate recognition. *J Biol Chem*, **289**, 11571-11583.
81. Chen, W., Zhang, L., Zheng, G., Fu, Y., Ji, Q., Liu, F., Chen, H. and He, C. (2014) Crystal structure of the RNA demethylase ALKBH5 from zebrafish. *FEBS Lett*, **588**, 892-898.
82. Songe-Moller, L., van den Born, E., Leihne, V., Vagbo, C.B., Kristoffersen, T., Krokan, H.E., Kirpekar, F., Falnes, P.O. and Klungland, A. (2010) Mammalian ALKBH8 possesses tRNA methyltransferase activity required for the biogenesis of multiple wobble uridine modifications implicated in translational decoding. *Mol Cell Biol*, **30**, 1814-1827.
83. Fu, D., Brophy, J.A., Chan, C.T., Atmore, K.A., Begley, U., Paules, R.S., Dedon, P.C., Begley, T.J. and Samson, L.D. (2010) Human AlkB homolog ABH8 Is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. *Mol Cell Biol*, **30**, 2449-2459.
84. Kalhor, H.R. and Clarke, S. (2003) Novel methyltransferase for modified uridine residues at the wobble position of tRNA. *Mol Cell Biol*, **23**, 9283-9292.

85. Fu, Y., Dai, Q., Zhang, W., Ren, J., Pan, T. and He, C.A. (2010) The AlkB Domain of Mammalian ABH8 Catalyzes Hydroxylation of 5-Methoxycarbonylmethyluridine at the Wobble Position of tRNA. *Angew Chem Int Edit*, **49**, 8885-8888.
86. van den Born, E., Vagbo, C.B., Songe-Moller, L., Leihne, V., Lien, G.F., Leszczynska, G., Malkiewicz, A., Krokan, H.E., Kirpekar, F., Klungland, A. *et al.* (2011) ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nat Commun*, **2**, 172.
87. Kawakami, M., Takemura, S., Kondo, T., Fukami, T. and Goto, T. (1988) Chemical structure of a new modified nucleoside located in the anticodon of Bombyx mori glycine tRNA². *J Biochem*, **104**, 108-111.
88. Shimada, K., Nakamura, M., Anai, S., De Velasco, M., Tanaka, M., Tsujikawa, K., Oujii, Y. and Konishi, N. (2009) A novel human AlkB homologue, ALKBH8, contributes to human bladder cancer progression. *Cancer Res*, **69**, 3157-3164.
89. Pastore, C., Topalidou, I., Forouhar, F., Yan, A.C., Levy, M. and Hunt, J.F. (2012) Crystal structure and RNA binding properties of the RNA recognition motif (RRM) and AlkB domains in human AlkB homolog 8 (ABH8), an enzyme catalyzing tRNA hypermodification. *J Biol Chem*, **287**, 2130-2143.
90. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R.B., Elliott, K.S., Lango, H., Rayner, N.W. *et al.* (2007) A Common Variant in the FTO Gene Is Associated with Body Mass Index and Predisposes to Childhood and Adult Obesity. *Science*, **316**, 889-894.
91. Dina, C., Meyre, D., Gallina, S., Durand, E., Korner, A., Jacobson, P., Carlsson, L.M., Kiess, W., Vatin, V., Lecoecur, C. *et al.* (2007) Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet*, **39**, 724-726.
92. Gerken, T., Girard, C.A., Tung, Y.C., Webby, C.J., Saudek, V., Hewitson, K.S., Yeo, G.S., McDonough, M.A., Cunliffe, S., McNeill, L.A. *et al.* (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*, **318**, 1469-1472.
93. Sanchez-Pulido, L. and Andrade-Navarro, M.A. (2007) The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. *BMC Biochem*, **8**, 23.
94. Peters, T., Ausmeier, K., Dildrop, R. and Ruther, U. (2002) The mouse Fused toes (Ft) mutation is the result of a 1.6-Mb deletion including the entire Iroquois B gene cluster. *Mamm Genome*, **13**, 186-188.
95. Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, **316**, 1341-1345.
96. Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Bruning, J.C. and Ruther, U. (2009) Inactivation of the Fto gene protects from obesity. *Nature*, **458**, 894-898.

97. Church, C., Moir, L., McMurray, F., Girard, C., Banks, G.T., Teboul, L., Wells, S., Bruning, J.C., Nolan, P.M., Ashcroft, F.M. *et al.* (2010) Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet*, **42**, 1086-1092.
98. Karra, E., O'Daly, O.G., Choudhury, A.I., Yousseif, A., Millership, S., Neary, M.T., Scott, W.R., Chandarana, K., Manning, S., Hess, M.E. *et al.* (2013) A link between FTO, ghrelin, and impaired brain food-cue responsivity. *J Clin Invest*, **123**, 3539-3551.
99. Smemo, S., Tena, J.J., Kim, K.H., Gamazon, E.R., Sakabe, N.J., Gomez-Marin, C., Aneas, I., Credidio, F.L., Sobreira, D.R., Wasserman, N.F. *et al.* (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, **507**, 371-375.
100. Iles, M.M., Law, M.H., Stacey, S.N., Han, J., Fang, S., Pfeiffer, R., Harland, M., Macgregor, S., Taylor, J.C., Aben, K.K. *et al.* (2013) A variant in FTO shows association with melanoma risk not due to BMI. *Nat Genet*, **45**, 428-432, 432e421.
101. Jia, G., Yang, C.G., Yang, S., Jian, X., Yi, C., Zhou, Z. and He, C. (2008) Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett*, **582**, 3313-3319.
102. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G. *et al.* (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol*, **7**, 885-887.
103. Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G. *et al.* (2014) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, **505**, 117-120.
104. Fu, Y., Jia, G., Pang, X., Wang, R.N., Wang, X., Li, C.J., Smemo, S., Dai, Q., Bailey, K.A., Nobrega, M.A. *et al.* (2013) FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat Commun*, **4**, 1798.
105. Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y. and Chai, J. (2010) Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature*, **464**, 1205-1209.
106. Church, C., Lee, S., Bagg, E.A., McTaggart, J.S., Deacon, R., Gerken, T., Lee, A., Moir, L., Mecinovic, J., Quwailid, M.M. *et al.* (2009) A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet*, **5**, e1000599.
107. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930-935.
108. He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L. *et al.* (2011) Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, **333**, 1303-1307.
109. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300-1303.

110. Guo, J.U., Su, Y., Zhong, C., Ming, G.L. and Song, H. (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell*, **145**, 423-434.
111. Bacher, U., Haferlach, C., Schnittger, S., Kohlmann, A., Kern, W. and Haferlach, T. (2010) Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies. *Ann Hematol*, **89**, 643-652.
112. Hashimoto, H., Pais, J.E., Zhang, X., Saleh, L., Fu, Z.Q., Dai, N., Correa, I.R., Jr., Zheng, Y. and Cheng, X. (2014) Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature*, **506**, 391-395.
113. Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P. and Xu, Y. (2013) Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell*.
114. Noma, A., Ishitani, R., Kato, M., Nagao, A., Nureki, O. and Suzuki, T. (2010) Expanding role of the jumonji C domain as an RNA hydroxylase. *J Biol Chem*, **285**, 34503-34507.
115. Kato, M., Arais, Y., Noma, A., Nagao, A., Suzuki, T., Ishitani, R. and Nureki, O. (2011) Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res*, **39**, 1576-1585.
116. Kivirikko, K.I. and Myllyla, R. (1982) Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol*, **82 Pt A**, 245-304.
117. Myllyharju, J. and Kivirikko, K.I. (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet*, **20**, 33-43.
118. Tiainen, P., Pasanen, A., Sormunen, R. and Myllyharju, J. (2008) Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV. *J Biol Chem*, **283**, 19432-19439.
119. Mizuno, K., Hayashi, T., Peyton, D.H. and Bachinger, H.P. (2004) The peptides acetyl-(Gly-3(S)Hyp-4(R)Hyp)10-NH₂ and acetyl-(Gly-Pro-3(S)Hyp)10-NH₂ do not form a collagen triple helix. *J Biol Chem*, **279**, 282-287.
120. Jenkins, C.L., Bretscher, L.E., Guzei, I.A. and Raines, R.T. (2003) Effect of 3-hydroxyproline residues on collagen stability. *Journal of the American Chemical Society*, **125**, 6422-6427.
121. Myllyharju, J. (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol*, **22**, 15-24.
122. Kivirikko, K.I. and Myllyharju, J. (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biology*, **16**, 357-368.
123. Kivirikko, K.I. and Pihlajaniemi, T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv Enzymol Relat Areas Mol Biol*, **72**, 325-398.
124. Peterkofsky, B. and Udenfriend, S. (1965) Enzymatic Hydroxylation of Proline in Microsomal Polypeptide Leading to Formation of Collagen. *Proc Natl Acad Sci U S A*, **53**, 335-342.

125. Prockop, D.J. and Juva, K. (1965) Synthesis of Hydroxyproline in Vitro by the Hydroxylation of Proline in a Precursor of Collagen. *Proc Natl Acad Sci U S A*, **53**, 661-668.
126. Hutton, J.J., Jr., Trappel, A.L. and Udenfriend, S. (1966) Requirements for alpha-ketoglutarate, ferrous ion and ascorbate by collagen proline hydroxylase. *Biochem Biophys Res Commun*, **24**, 179-184.
127. Kukkola, L., Koivunen, P., Pakkanen, O., Page, A.P. and Myllyharju, J. (2004) Collagen prolyl 4-hydroxylase tetramers and dimers show identical decreases in Km values for peptide substrates with increasing chain length: mutation of one of the two catalytic sites in the tetramer inactivates the enzyme by more than half. *J Biol Chem*, **279**, 18656-18661.
128. Neubauer, A., Neubauer, P. and Myllyharju, J. (2005) High-level production of human collagen prolyl 4-hydroxylase in Escherichia coli. *Matrix Biol*, **24**, 59-68.
129. Neubauer, A., Soini, J., Bollok, M., Zenker, M., Sandqvist, J., Myllyharju, J. and Neubauer, P. (2007) Fermentation process for tetrameric human collagen prolyl 4-hydroxylase in Escherichia coli: improvement by gene optimisation of the PDI/beta subunit and repeated addition of the inducer anhydrotetracycline. *J Biotechnol*, **128**, 308-321.
130. Gilkes, D.M., Chaturvedi, P., Bajpai, S., Wong, C.C., Wei, H., Pitcairn, S., Hubbi, M.E., Wirtz, D. and Semenza, G.L. (2013) Collagen prolyl hydroxylases are essential for breast cancer metastasis. *Cancer Res*, **73**, 3285-3296.
131. Hieta, R., Kukkola, L., Permi, P., Pirila, P., Kivirikko, K.I., Kilpelainen, I. and Myllyharju, J. (2003) The peptide-substrate-binding domain of human collagen prolyl 4-hydroxylases. Backbone assignments, secondary structure, and binding of proline-rich peptides. *J Biol Chem*, **278**, 34966-34974.
132. Myllyharju, J. and Kivirikko, K.I. (1999) Identification of a novel proline-rich peptide-binding domain in prolyl 4-hydroxylase. *Embo J*, **18**, 306-312.
133. Myllyharju, J. and Kivirikko, K.I. (1997) Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *Embo J*, **16**, 1173-1180.
134. Anantharajan, J., Koski, M.K., Kursula, P., Hieta, R., Bergmann, U., Myllyharju, J. and Wierenga, R.K. (2013) The structural motifs for substrate binding and dimerization of the alpha subunit of collagen prolyl 4-hydroxylase. *Structure*, **21**, 2107-2118.
135. Pekkala, M., Hieta, R., Bergmann, U., Kivirikko, K.I., Wierenga, R.K. and Myllyharju, J. (2004) The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. *Journal of Biological Chemistry*, **279**, 52255-52261.
136. Koski, M.K., Hieta, R., Hirsila, M., Ronka, A., Myllyharju, J. and Wierenga, R.K. (2009) The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem*, **284**, 25290-25301.

137. Koski, M.K., Hieta, R., Bollner, C., Kivirikko, K.I., Myllyharju, J. and Wierenga, R.K. (2007) The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. *J Biol Chem*, **282**, 37112-37123.
138. Culpepper, M.A., Scott, E.E. and Limburg, J. (2010) Crystal structure of prolyl 4-hydroxylase from *Bacillus anthracis*. *Biochemistry*, **49**, 124-133.
139. Alanen, H.I., Salo, K.E., Pekkala, M., Siekkinen, H.M., Pirneskoski, A. and Ruddock, L.W. (2003) Defining the domain boundaries of the human protein disulfide isomerases. *Antioxid Redox Signal*, **5**, 367-374.
140. Freedman, R.B., Gane, P.J., Hawkins, H.C., Hlodan, R., McLaughlin, S.H. and Parry, J.W. (1998) Experimental and theoretical analyses of the domain architecture of mammalian protein disulphide-isomerase. *Biol Chem*, **379**, 321-328.
141. Pirneskoski, A., Klappa, P., Lobell, M., Williamson, R.A., Byrne, L., Alanen, H.I., Salo, K.E., Kivirikko, K.I., Freedman, R.B. and Ruddock, L.W. (2004) Molecular characterization of the principal substrate binding site of the ubiquitous folding catalyst protein disulfide isomerase. *J Biol Chem*, **279**, 10374-10381.
142. Darby, N.J. and Creighton, T.E. (1995) Functional properties of the individual thioredoxin-like domains of protein disulfide isomerase. *Biochemistry*, **34**, 11725-11735.
143. Wang, C., Li, W., Ren, J., Fang, J., Ke, H., Gong, W., Feng, W. and Wang, C.C. (2013) Structural insights into the redox-regulated dynamic conformations of human protein disulfide isomerase. *Antioxid Redox Signal*, **19**, 36-45.
144. van der Slot, A.J., Zuurmond, A.M., Bardoel, A.F., Wijmenga, C., Pruijs, H.E., Sillence, D.O., Brinckmann, J., Abraham, D.J., Black, C.M., Verzijl, N. *et al.* (2003) Identification of PLOD2 as telopeptide lysyl hydroxylase, an important enzyme in fibrosis. *J Biol Chem*, **278**, 40967-40972.
145. Salo, A.M., Wang, C., Sipila, L., Sormunen, R., Vapola, M., Kervinen, P., Ruotsalainen, H., Heikkinen, J. and Myllyla, R. (2006) Lysyl hydroxylase 3 (LH3) modifies proteins in the extracellular space, a novel mechanism for matrix remodeling. *J Cell Physiol*, **207**, 644-653.
146. Myllyla, R., Wang, C., Heikkinen, J., Juffer, A., Lampela, O., Risteli, M., Ruotsalainen, H., Salo, A. and Sipila, L. (2007) Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). *J Cell Physiol*, **212**, 323-329.
147. Kellokumpu, S., Sormunen, R., Heikkinen, J. and Myllyla, R. (1994) Lysyl hydroxylase, a collagen processing enzyme, exemplifies a novel class of lumenally-oriented peripheral membrane proteins in the endoplasmic reticulum. *J Biol Chem*, **269**, 30524-30529.
148. Suokas, M., Lampela, O., Juffer, A.H., Myllyla, R. and Kellokumpu, S. (2003) Retrieval-independent localization of lysyl hydroxylase in the endoplasmic reticulum via a peptide fold in its iron-binding domain. *Biochem J*, **370**, 913-920.
149. Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M. and Myllyla, R. (2000) Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. *J Biol Chem*, **275**, 36158-36163.

150. Wang, C., Luosujarvi, H., Heikkinen, J., Risteli, M., Uitto, L. and Myllyla, R. (2002) The third activity for lysyl hydroxylase 3: galactosylation of hydroxyllysyl residues in collagens in vitro. *Matrix Biol*, **21**, 559-566.
151. Wang, C., Risteli, M., Heikkinen, J., Hussa, A.K., Uitto, L. and Myllyla, R. (2002) Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). *J Biol Chem*, **277**, 18568-18573.
152. Steinmann, B., Eyre, D.R. and Shao, P. (1995) Urinary pyridinoline cross-links in Ehlers-Danlos syndrome type VI. *Am J Hum Genet*, **57**, 1505-1508.
153. Gilkes, D.M., Bajpai, S., Wong, C.C., Chaturvedi, P., Hubbi, M.E., Wirtz, D. and Semenza, G.L. (2013) Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Mol Cancer Res*, **11**, 456-466.
154. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol*, **215**, 403-410.
155. Vranka, J.A., Sakai, L.Y. and Bachinger, H.P. (2004) Prolyl 3-hydroxylase 1, enzyme characterization and identification of a novel family of enzymes. *J Biol Chem*, **279**, 23615-23621.
156. Marini, J.C., Cabral, W.A. and Barnes, A.M. (2010) Null mutations in LEPRE1 and CRTAP cause severe recessive osteogenesis imperfecta. *Cell Tissue Res*, **339**, 59-70.
157. Vranka, J.A., Pokidysheva, E., Hayashi, L., Zientek, K., Mizuno, K., Ishikawa, Y., Maddox, K., Tufa, S., Keene, D.R., Klein, R. *et al.* (2010) Prolyl 3-hydroxylase 1 null mice display abnormalities in fibrillar collagen-rich tissues such as tendons, skin, and bones. *J Biol Chem*, **285**, 17253-17262.

Chapter 2

Proof-of-Principle Studies on Structure-based Inhibition of NAOX using AlkB as a Model Enzyme

2.1 Introduction

As described in Chapter 1, AlkB was the first member of the nucleic acid oxygenase subfamily to be identified (1-4). *Escherichia coli* AlkB acts as a DNA repair enzyme that preferentially demethylates *N*¹-methyladenine (m¹A) and *N*³-methylcytosine (m³C) on single-stranded DNA (Figure 2 - 1a) (1,2). AlkB has also been shown to be promiscuous as it accepts other base substrates including 1,*N*⁶-ethenoadenine (εA), 3,*N*⁴-ethenoecytosine (εC), and *N*⁶-methyladenine (m⁶A) (Figure 2 - 1b) (5-7). Crystal structures of AlkB in complex with trinucleotide DNA and double-stranded DNA (dsDNA) have revealed a base flipping mechanism via direct stacking of bases adjacent to the 'repaired' base (8,9).

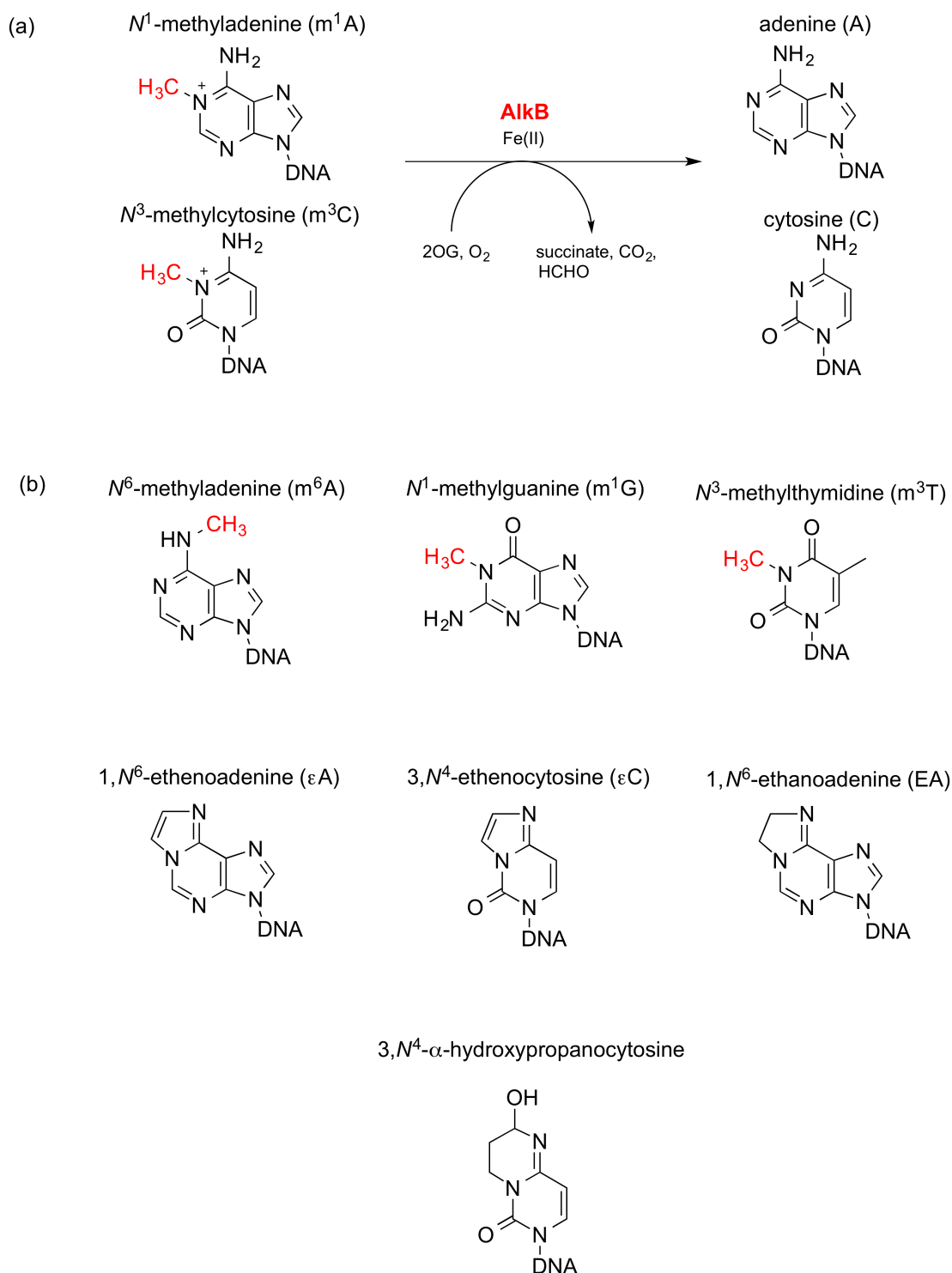


Figure 2 - 1. Nucleobase substrates reported to be accepted by AlkB. (a) Reaction catalysed by AlkB showing demethylation of *N*¹-methyladenine (m¹A) and *N*³-methylcytosine (m³C) (2,10). (b) Structures of substrates accepted by AlkB in addition to m¹A and m³C (11).

Nine AlkB homologues have been identified in humans (ALKBH1-8 and FTO) (10,12-16), some of which are associated with disease (17). For instance,

FTO is linked to obesity while ALKBH8 is overexpressed in bladder cancer cells (18,19). Prior to this study, structure-guided inhibition studies had not been carried out on the nucleic acid oxygenases, and there had yet been any reported crystal structures of AlkB in complex with small-molecule inhibitors.

As a proof of principle that the nucleic acid oxygenases can be subjected to structure-based inhibition studies to yield selective inhibitors, structural analyses of AlkB in complex with inhibitors were carried out alongside *in vitro* inhibition investigations. This chapter describes work on the structure-based inhibition of AlkB.

2.2 Results and Discussions

Some of the results described in Sections 2.2.1 and 2.2.2 and associated experimental procedures described in this chapter have been reported in Hopkinson *et al.* (2013) '5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation.' *Chem. Sci.*, **4**, 3110 - reproduced by permission of The Royal Society of Chemistry.

Some of the results described in Sections 2.2.3 - 2.2.8 and associated experimental procedures described in this chapter are reprinted (adapted) with permission from Woon *et al.* (2012) 'Dynamic Combinatorial Mass Spectrometry Leads to Inhibitors of a 2-Oxoglutarate-Dependent Nucleic Acid Demethylase. *J. Med. Chem.*, **55**, 2173-2184.' Copyright (2012) American Chemical Society.

2.2.1 Inhibition of AlkB by broad-spectrum 2OG oxygenase inhibitors

As the most well-characterised nucleic acid demethylase subfamily, AlkB was considered to be a suitable model for inhibition studies with respect to the human homologues (e.g. the fat mass and obesity associated protein FTO and human RNA *N*⁶-methyladenine demethylase ALKBH5; see Chapters 3 and 4). Inhibition of AlkB was initially investigated using broad spectrum inhibitors such as *N*-oxalylglycine (NOG), 2,4-pyridine dicarboxylic acid (2,4-PDCA), and 5-carboxy-8-hydroxyquinoline (IOX1), which have been shown to inhibit various 2OG oxygenases from different subfamilies (20). A formaldehyde dehydrogenase (FDH) assay was used to obtain the half maximal response inhibitor concentration (IC₅₀ values) of these broad spectrum 2OG oxygenase inhibitors against AlkB.

AlkB catalyses the formation of a hemiaminal by hydroxylating a methylated base on DNA, and the hemiaminal subsequently fragments to give an unmethylated base and formaldehyde (11,21). The latter can be oxidised, as catalysed by formaldehyde dehydrogenase (FDH) using NAD⁺, generating NADH (Figure 2 - 2) (22). NADH production can be monitored by fluorescence with an excitation wavelength of 355 nm and an emission wavelength of 460 nm (23). Therefore, the rate of increase in fluorescence at wavelength 460 nm, which corresponds to the increase of NADH, can be used to indirectly quantify the demethylation reaction of AlkB (24). Further, inhibition of AlkB can be quantified by measuring the initial rates of reaction in the presence of an inhibitor.

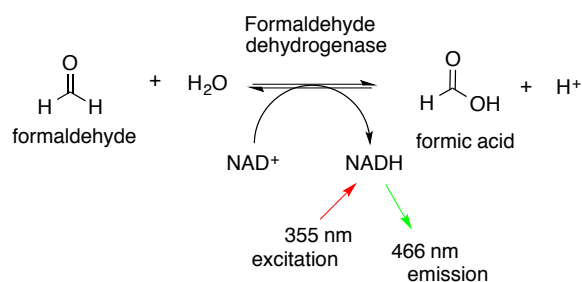


Figure 2 - 2. Formaldehyde dehydrogenase reaction scheme. NADH is generated concomitant with oxidation of formaldehyde to give formic acid.

Using an FDH-based assay for AlkB developed by Dr Eleanor Bagg, the initial rates of reaction of AlkB were investigated with different inhibitor concentrations to obtain IC₅₀ values of three broad spectrum 2OG oxygenase inhibitors, all 2OG competitors (25). The IC₅₀ values of NOG, 2,4-PDCA and IOX1 were 2.7 μ M, 1.4 μ M and 15 μ M respectively (chemical structures and IC₅₀ curves shown in Figure 2 - 3a-c), indicating that 2,4-PDCA is the most potent of the three inhibitors against AlkB. Rather surprisingly, IOX1 was found to be the least potent of these inhibitors. This result is different from results with several other 2OG oxygenases (e.g. JMJD2A, JMJD3 and PHF8), for which IOX1 is often more potent than NOG and 2,4-PDCA (20). Structural studies of IOX1 in complex with various JmjC enzymes have revealed an interesting binding mode of IOX1 in the active site, where IOX1 causes translocation of the active site metal (i.e. metal is moved out of its usual position in the active site) (20,26). This observation has also led to speculation that metal translocation may contribute to, or correlate with, the inhibitory potency of the compound (20). To investigate the binding mode of IOX1 in AlkB as a representative of nucleic acid demethylases, crystallographic studies of the AlkB-IOX1 complex were then carried out.

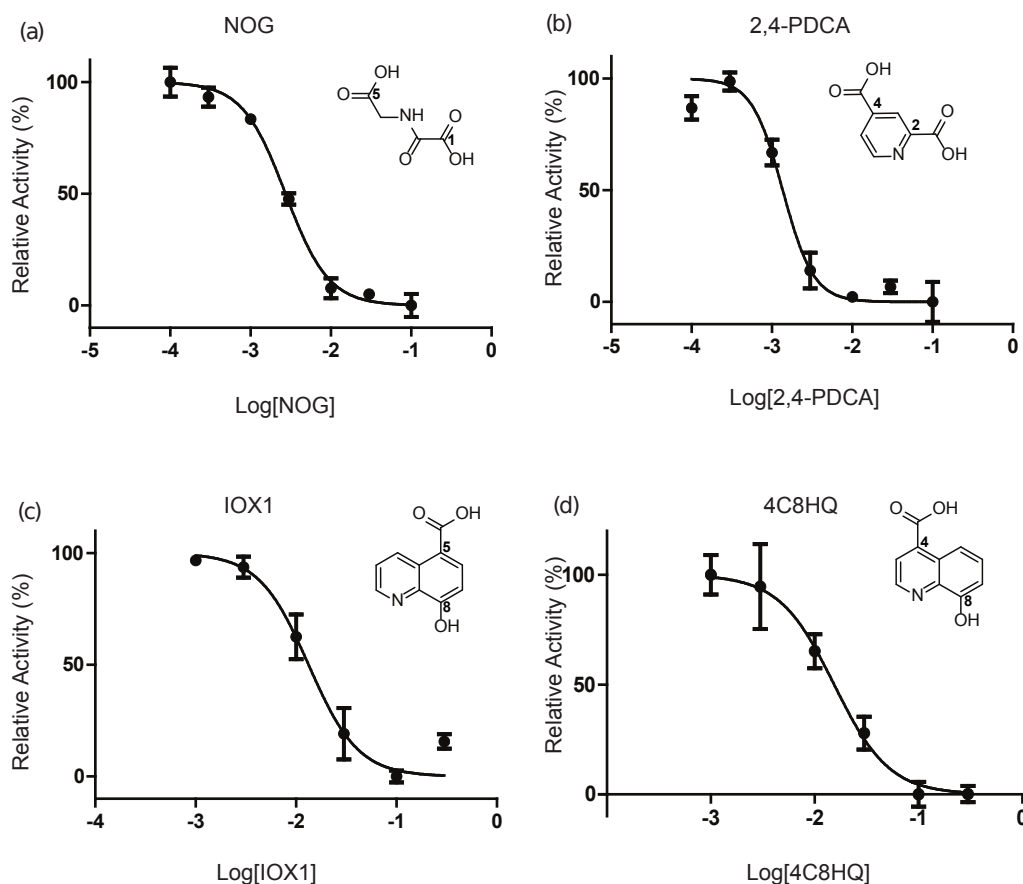


Figure 2 - 3. IC₅₀ curves of four broad spectrum inhibitors when tested with AlkB. Dose-response curves of AlkB relative activity measured using the FDH-based assay in the presence of (a) *N*-oxalylglycine (NOG) (IC₅₀ = 2.7 μM), (b) 2,4-pyridine dicarboxylic acid (2,4-PDCA) (IC₅₀ = 1.4 μM), (c) 5-carboxy-8-hydroxyquinoline (IOX1) (IC₅₀ = 15 μM), and (d) 4-carboxy-8-hydroxyquinoline (4C8HQ) (IC₅₀ = 14 μM). The chemical structures of the inhibitors used are shown.

2.2.2 Crystal structure of AlkB-IOX1 complex

A crystallisable construct of AlkB with an N-terminal 11 residue truncation was used in crystallographic studies. The protein was purified by Dr Eleanor Bagg. Crystallisation trials were set up in the presence of manganese chloride and IOX1 using an Index screen (Hampton Research), a sparse matrix crystallisation screen, and an optimisation screen derived from conditions that gave AlkB-15 crystals (0.2 M NaCl, 0.1 M HEPES pH 7.5, and 25% PEG3350) (see Section 2.2.4 and Section 2.2.6). AlkB was found to crystallise in the presence of IOX1 in the

optimisation screen (see details on crystallisation in Section 2.4.2, Table 2 - 2).

AlkB crystals typically had a cuboid morphology (Figure 2 - 4).

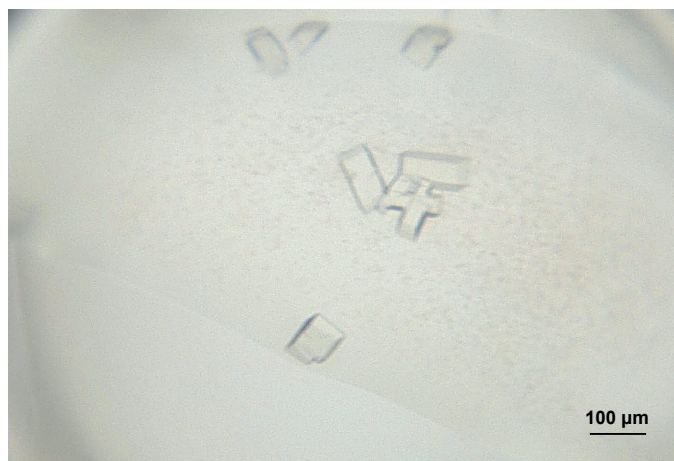


Figure 2 - 4. A representative image of AlkB crystals showing their cuboid morphology.

Crystals of AlkB grown in the presence of IOX1 were harvested and one was found to diffract to 1.2 Å at the Diamond Light Source synchrotron Beamline I04. The space group was found to be *P*1 with unit cell constants $a = 36.84$ Å, $b = 38.59$ Å, $c = 40.45$ Å, $\alpha = 77.62^\circ$, $\beta = 75.01^\circ$, $\gamma = 66.10^\circ$. Data were collected for 360° at 100 K (see Table 2 – 1, Section 2.2.7, and Experimental Procedure Section 2.4.2). The structure of AlkB-IOX1 complex was solved by molecular replacement using AlkB-**18** (PDB ID 3T4V, see Section 2.2.6), with compound **18** removed, as a model.

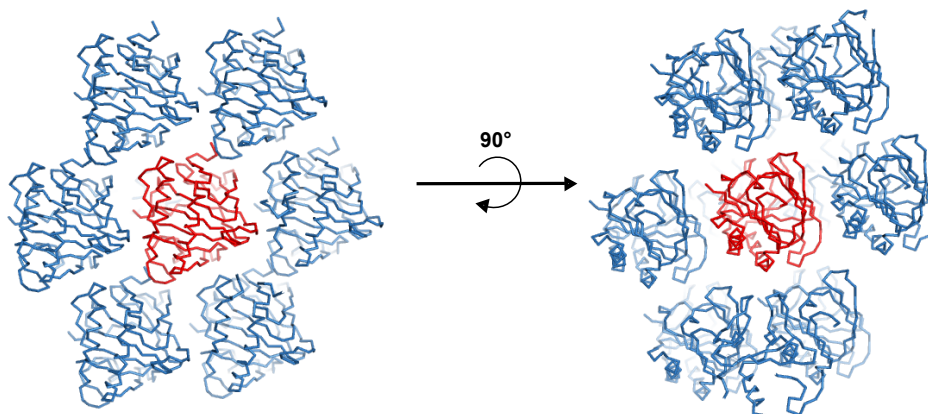
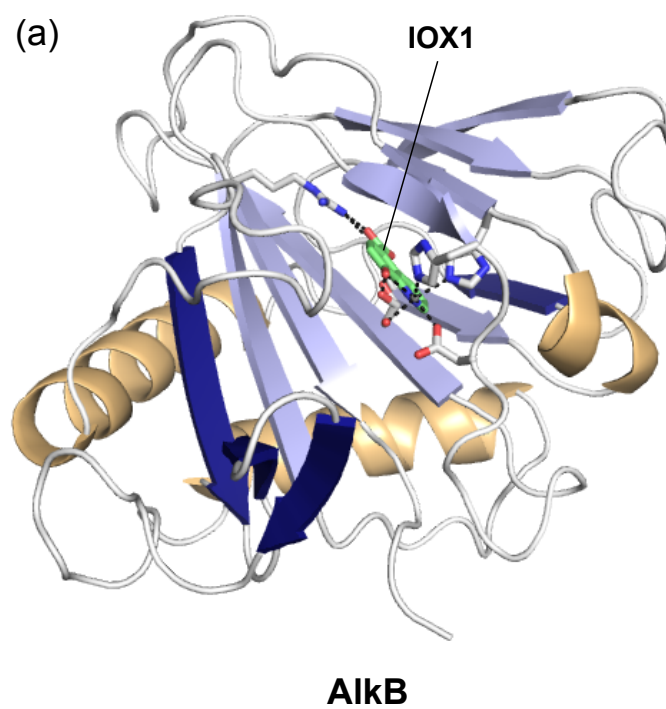


Figure 2 - 5. Crystal packing of AlkB protein molecules in two orientations rotated at 90° relative to each other. There is one molecule of AlkB per asymmetric unit in a *P1* space group. Protein structure models were produced using MacPyMOL v1.3 (27).

The AlkB-IOX1 structure reveals that IOX1 occupies the 2OG binding pocket and chelates the active site metal in a bidentate manner, using the oxygen from the 8-hydroxyl group and the nitrogen from the quinoline ring (Figure 2 - 6a, b). The 5-carboxy group of IOX1 is positioned to interact with the 2OG binding Arg204 and Tyr122, which interact with the co-substrate 2OG C5-carboxylate during catalysis. However, the position of the IOX1 5-carboxylate group does not overlap perfectly with the 2OG C5-carboxylate group (Figure 2 - 6g). In the case of IOX1, the carboxylate is shifted to form only a single hydrogen bond with Arg204 of AlkB with one of the oxygens from its carboxylate group; the other 5-carboxylate oxygen is positioned to hydrogen bond with the hydroxyl group of the side chain of Ser145. During the assignment of IOX1, extra electron density was observed in the 2OG binding pocket near Arg204, part of which resembled that of a carboxylate group. This extra density was later assigned to an acetate molecule due to its shape and because acetate had been

used in the IOX1 synthesis process (Jacob Bush, personal communication). Both IOX1 and the acetate molecules were both refined to 0.5 occupancy levels.



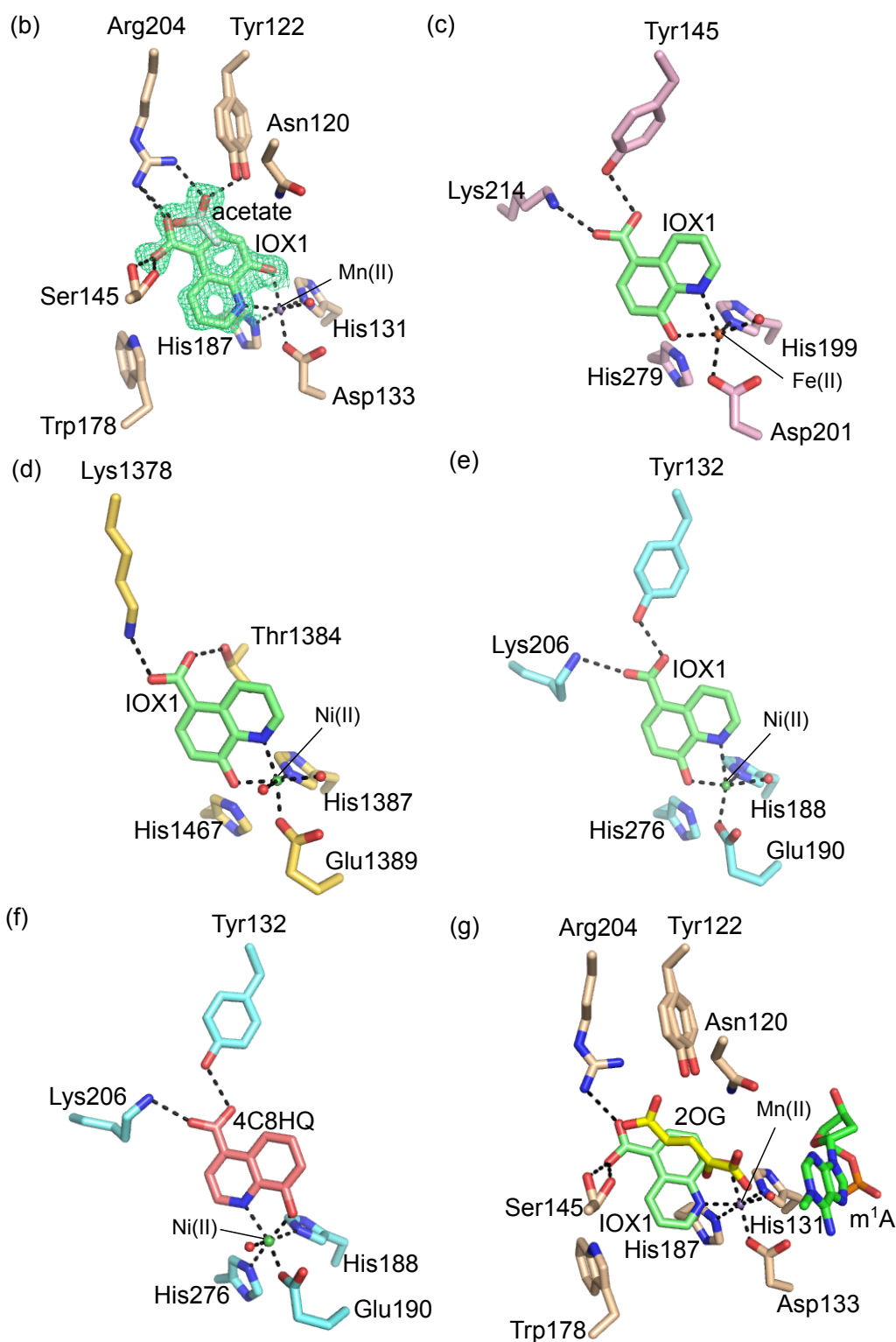


Figure 2 - 6. Views from crystal structures of related 2OG oxygenases in complex with IOX1 and 4C8HQ. (a) Overall view from a crystal structure of the AlkB-IOX1 complex (PDB ID 4JHT; resolution: 1.18 Å). DSBH β -strands (light blue), additional N-terminal extension β -strands (dark blue), helices (light orange), loops (white), catalytic triad HXD...H and 2OG-interacting Arg204 and Ser145 (white sticks), and IOX1 (green sticks) are indicated. (b) View of the active site of the AlkB-IOX1 structure (resolution 1.18 Å). Electron density for IOX1 and acetate (3.0σ mF_o-DF_c OMIT, green mesh) is shown. Both IOX1 and acetate were refined to 0.5 occupancies. Views from the active sites of crystal structures of (c) the FIH-IOX1 complex (PDB ID 4BIO; resolution: 2.45 Å) (20), (d) the JMJD3-IOX1 complex (PDB ID 2XXZ; resolution: 1.80 Å) (20), (e) the JMJD2A-IOX1 complex (PDB ID 3NJY; resolution: 2.60 Å) (26), and (f) the JMJD2A-4C8HQ

complexes (PDB ID 4BIS; resolution: 2.49 Å) (20) are shown. **(g)** Superimposition of 2OG and m¹A nucleotide substrate from a structure of the AlkB-dsDNA complex (PDB ID 3BIE; resolution: 1.68 Å) (9) onto a structure of the AlkB-IOX1 complex (PDB ID 4JHT; resolution: 1.18 Å). Oxygen (red), nitrogen (blue), Mn(II) (grey sphere), Ni(II) (green sphere), and water (red sphere) are indicated. Protein structure models were produced using MacPyMOL v1.3 (27).

Interestingly, IOX1 adopts an alternative binding mode when compared with reported crystal structures of IOX1 in complex with other 2OG oxygenases including the hypoxia inducible factor asparaginyl hydroxylase (FIH) (Figure 2 - 6c), and the histone demethylases JMJD3 (Figure 2 - 6d) and JMJD2A (Figure 2 - 6e) (20,26). The 8-hydroxyl group of IOX1 is situated *trans* to Asp133 of the HXD...H motif in AlkB, whilst the isoquinoline nitrogen is positioned *trans* to the proximal histidine, His131, of the conserved HXD/E...H motif (Figure 2 - 6b). This arrangement differs from that observed in structures of FIH, JMJD2A and JMJD3 in complex with IOX1; where the 8-hydroxyl group is positioned *trans* to the proximal histidine of the conserved HXD/E...H motif while the isoquinoline nitrogen is positioned *trans* to the aspartate or glutamate of the HXD/E...H motif (Figure 2 - 6c-e) (20,26). Another striking difference is that the active site metal in AlkB remains in almost the same position as that observed when AlkB is in complex with 2OG (8). This contrasts with the translocated metal position observed in the structures of JMJD2A, JMJD3 and FIH in complex with IOX1. The translocated metal (~1.5 Å from its usual position) may in some cases, such as in JMJD2A and JMJD3, partially contribute to the high inhibitory potency of IOX1 (IC₅₀ 0.2 μM and 0.14 μM respectively) (20). The binding mode of IOX1 in AlkB, that does not cause active site metal translocation, might explain why IOX1 has a lower inhibitory potency in AlkB than in JMJD2A and JMJD3.

An analogue of IOX1, 4-carboxy-8-hydroxyquinoline (4C8HQ), was predicted to bind to the active site of JmjC oxygenases without causing active site metal translocation by docking studies (performed by Dr Richard Hopkinson), leading to speculation that 4C8HQ is less potent against JmjC oxygenases compared to IOX1. Indeed, 4C8HQ (synthesised by Cyrille Thinnes) was later found to be less potent than IOX1 in most histone demethylases such as JMJD2A and JMJD3 (IC_{50} 1.4 μ M and 9 μ M respectively) (20). Additionally, a crystal structure of JMJD2A in complex with 4-carboxy-8-hydroxyquinoline determined by Dr Rasheduzzaman Chowdhury (PDB ID: 4BIS; resolution: 2.49 Å) (20) (Figure 2 - 6f) revealed that metal translocation was not observed crystallographically, suggesting that metal translocation might contribute to the potency of IOX1 in certain oxygenases. When tested against AlkB, 4C8HQ has a similar potency (IC_{50} = 14 μ M) as IOX1 (IC_{50} = 15 μ M) (Figure 2 - 3c, d).

Table 2 - 1. IC_{50} values (μ M) of 5-carboxy-8-hydroxyquinoline (IOX1), 4-carboxy-8-hydroxyquinoline (4C8HQ), N-oxalylglycine (NOG), and 2,4-pyridine dicarboxylate (2,4-PDCA) against 2OG oxygenases AlkB, JMJD2A, JMJD3, PHF8 and FIH (20).

	IOX1	4C8HQ	NOG	2,4-PDCA
AlkB	15	14	2.7	1.4
JMJD2A	0.20	1.4	N/A	2.0
JMJD3	0.14	9.0	0.30	33
PHF8	13	17	>300	150
FIH	7.6	7.5	N/A	N/A

2.2.3 Development of AlkB selective inhibitors

(Dynamic combinatorial mass spectrometry (DCMS) and organic synthesis were carried out by Dr Esther Woon and Dr Marina Demetriades, dynamic scanning fluorometry (DSF) analyses were carried out by Dr Eleanor Bagg. *In vitro* assays and capillary electrophoresis were carried out by Dr Svetlana Krylova and David Wegman from Professor Sergey Krylov's laboratory in York University, Toronto, Canada.)

A combination of DCMS (Figure 2 - 7) and structure-guided design approaches were then adopted to identify selective inhibitors against AlkB.

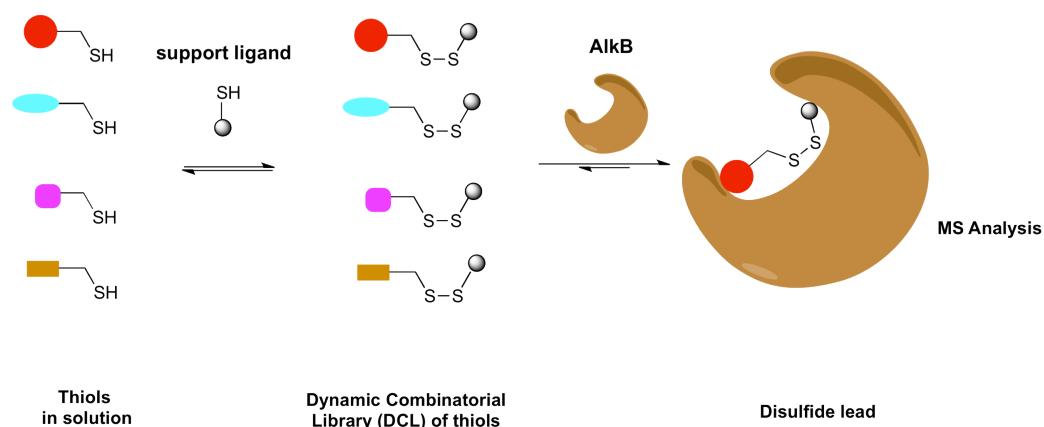


Figure 2 - 7. Schematic representation of the dynamic combinatorial mass spectrometry technique, which was used to identify potential binders for AlkB. A thiol-containing support ligand that could bind into the active site of AlkB was incubated with a library of thiols, resulting in reversible disulphide formation between the support ligand and the thiols. Disulphides that form stable interactions with AlkB protein can be detected using non-denaturing electrospray ionisation-mass spectrometry (ESI-MS). (Note it is possible that a support ligand already bound to AlkB can bind reversibly with the thiols.)

2.2.4 Dynamic Combinatorial Mass Spectrometry

Initially, DCMS was carried out to identify potential compounds that would bind to the active site of AlkB. *N*-oxalyl-*L*-cysteine (LNOC) (Figure 2 - 8a) was used as the support ligand to form dynamic disulphides with a library of 37 thiols in solution (Figure 2 - 8b). LNOC was chosen as a support ligand on the basis of structural analysis and the fact NOG is a broad spectrum 2OG oxygenase inhibitor (20,25).

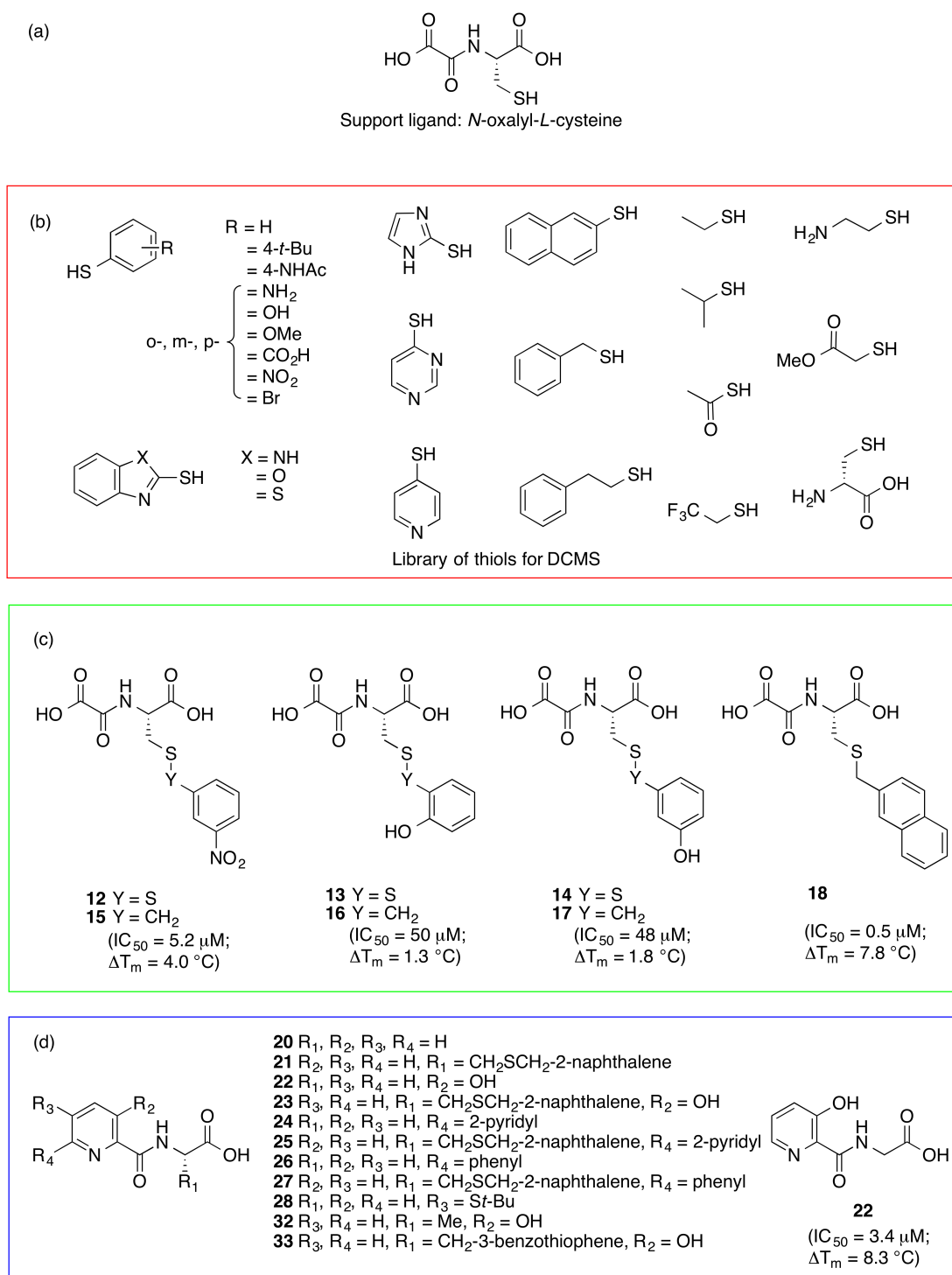


Figure 2 - 8. Chemical structures of selected compounds investigated as AlkB inhibitors. (a) Support ligand *N*-oxalyl-*L*-cysteine (LNOC); **(b)** library of thiols used in dynamic combinatorial mass spectrometry; **(c)** hits from DCMS and their corresponding stable analogues; **(d)** compounds with improved potency from structure-guided design. IC₅₀ values and melting temperature shift (ΔT_m) are shown in parenthesis.

The thiols and LNOC were incubated with AlkB in the presence of Fe(II). Non-denaturing electrospray ionisation-mass spectrometry (ESI-MS) carried out after 4 h of incubation recorded two main groups of peaks corresponding to thiols with the masses of ~124 and ~153 Da (Figure 2 - 9f). When AlkB was incubated with *N*-oxalyl *D*-cysteine (DNOC) as a support ligand and without support ligand, peaks corresponding to disulphide formation were not observed, indicating that LNOC is essential for the binding of disulphides in AlkB. Further ESI-MS binding experiments demonstrated disulphides **12** – **14** bound to AlkB most tightly, when tested as isolated compounds. Stable analogues (**15** – **17**) of **12** – **14** (Figure 2 - 8c) were subsequently synthesised. A pairwise competition study was then carried out among **15** – **17** using non-denaturing ESI-MS. The results revealed that **15** binds most strongly to AlkB, followed by **17** and **16**. These rankings were also in agreement with melting temperature shift (ΔT_m) shifts and IC_{50} values, obtained from DSF assays and capillary electrophoresis respectively: **15** (IC_{50} = 5.2 μ M, ΔT_m = 4.0 °C) > **17** (IC_{50} = 48 μ M, ΔT_m = 1.8 °C) > **16** (IC_{50} = 50 μ M, ΔT_m = 1.3 °C).

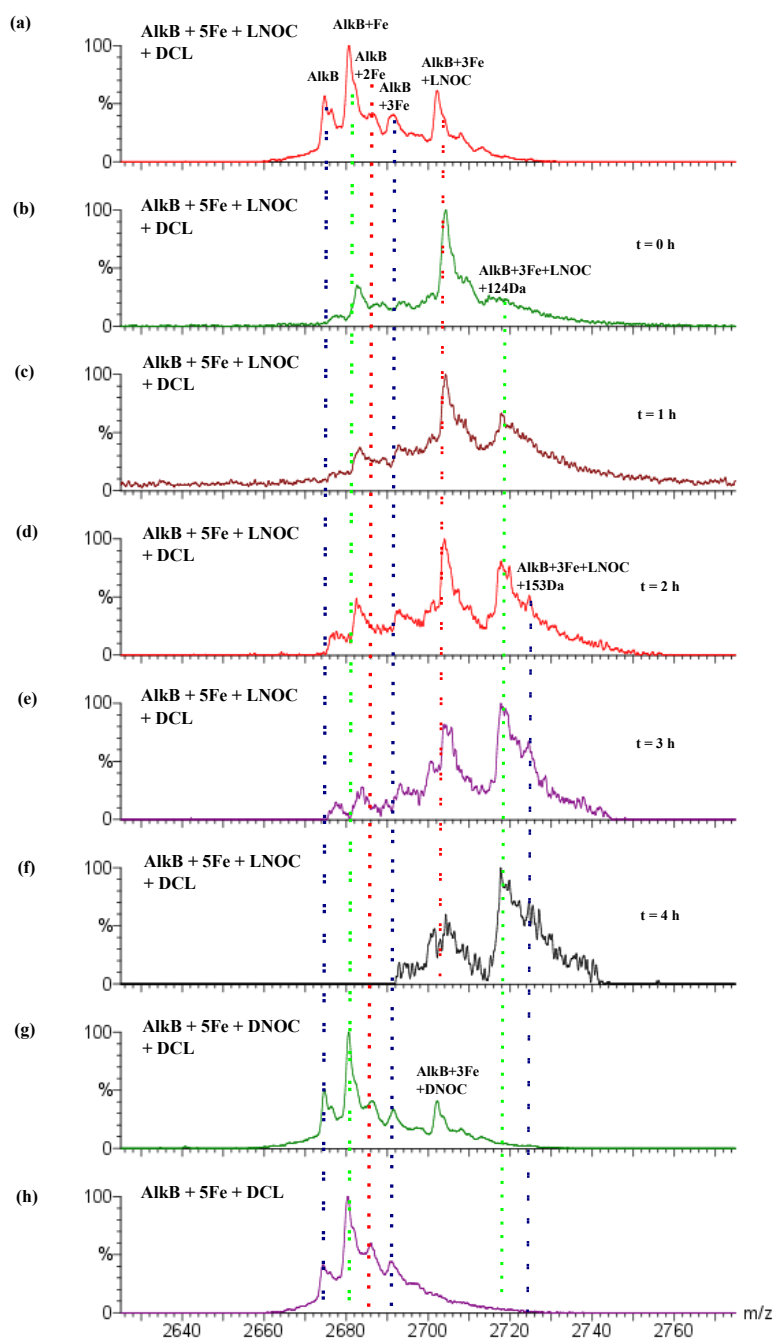


Figure 2 - 9. Non-denaturing mass spectra of DCMS assays with AlkB. (a – f) Time-dependent (0 – 4 h incubation) spectra of AlkB in the presence of Fe(II), support ligand *N*-oxalyl-*L*-cysteine (LNOC) and the library of thiols (DCL); spectra of AlkB in the presence of (g) Fe(II), *N*-oxalyl-*D*-cysteine (DNOC), and DCL; (h) Fe(II) and DCL. In (f), two major peaks corresponding to mass shifts of +124 Da and + 153 Da can be clearly observed after 4 h incubation. No peak corresponding to disulphide formation was observed when using (g) DNOC as support ligand and (h) without support ligand. The figure was originally provided by Dr Marina Demetriades and subsequently adapted.

Structural studies were then carried out to investigate the binding mode of compound **15** in AlkB (Dr Michael McDonough and Jerome Ma). A crystal structure of AlkB-**15** complex, which has a space group of *P1* and unit cell constants of $a = 36.66 \text{ \AA}$, $b = 38.83 \text{ \AA}$, and $c = 40.27 \text{ \AA}$, $\alpha = 77.64^\circ$, $\beta = 75.44^\circ$, and $\gamma = 66.42^\circ$, was determined to 1.6 $\text{\AA}$ resolution. Compound **15** was observed to be bound in the 2OG binding site chelating the iron in a bidentate manner using its oxalyl group (Figure 2 - 10). The other carboxylate of **15** that does not chelate iron forms electrostatic interactions with the side chains of Tyr122, Asn120, and Arg204. The side chain of **15** that bears the nitrophenyl group binds in the hydrophobic pocket at the back of the DSBH core, with the aromatic nitrophenyl forming π - π interactions with the side chains of Trp178 and His187. The π - π interactions are likely to contribute to the high affinity of **15** for AlkB.

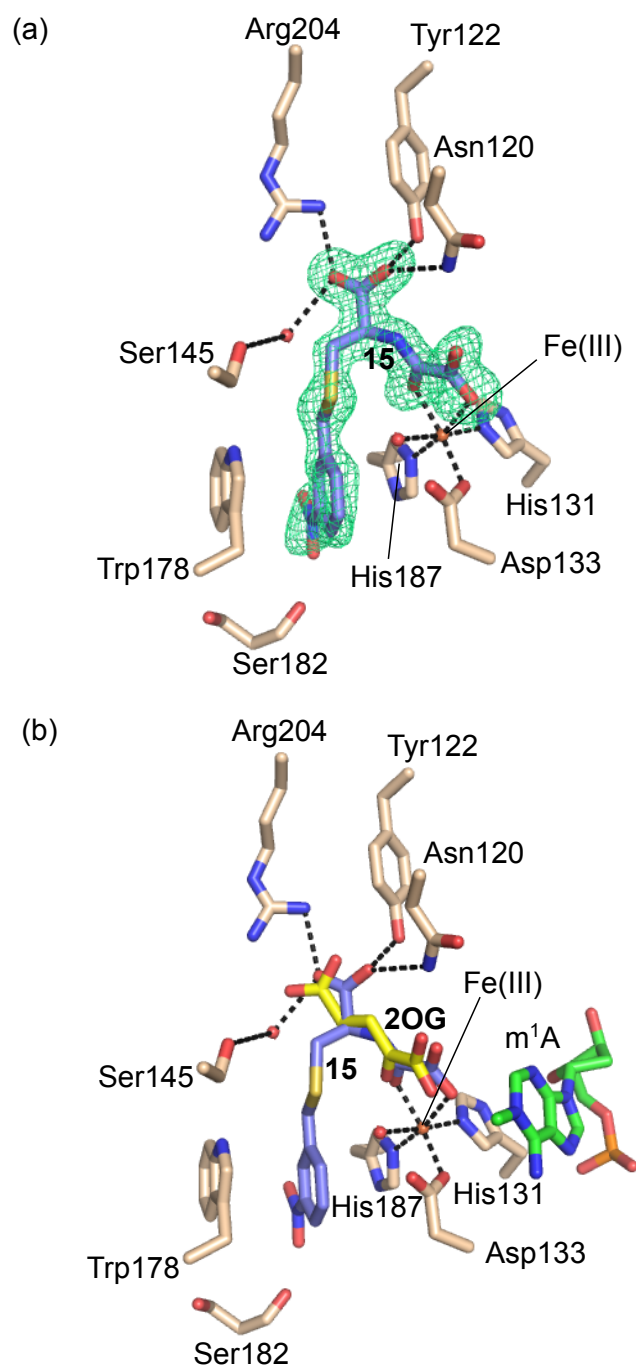


Figure 2 - 10. Views of a crystal structure of AlkB in complex with 15. (a) View from the active site of a crystal structure of the AlkB-15 complex. Electron density maps for 15 ($\sigma = 3.0$, mF_o-DF_c OMIT, green mesh) are shown. **(b)** Superimposition of 2OG and m¹A nucleotide substrate from a crystal structure of the AlkB-dsDNA complex (PDB ID 3BIE) onto the active site of a crystal structure of the AlkB-15 complex. Oxygen (red), nitrogen (blue), Fe(III) (orange sphere), water (red sphere), and electrostatic interactions (black dashes) are indicated. Protein structure models were produced using MacPyMOL v1.3 (27).

2.2.5 Structure-guided inhibitor design

Due to the presence of an apparently hydrophobic pocket in AlkB and the presence of a tryptophan residue (Trp178) that is likely to form π - π interactions with the aromatic group of inhibitors, further inhibitors were designed based on the structural knowledge from the AlkB-**15** complex crystal structure. These inhibitors contain a naphthalene group, a bulkier aromatic group compared to the nitrophenyl group. It was envisaged that the naphthalene might form stronger π - π , or hydrophobic, interactions with Trp178 and His187. Compound **18** ($IC_{50} = 0.5 \mu M$) was shown to be more potent than compound **15** ($IC_{50} = 5.3 \mu M$) (Figure 2 - 8c). Both compounds **15** and **18** had IC_{50} values greater than 1 mM when tested against hypoxia-inducible factor prolyl hydroxylase 2 (PHD2) and histone demethylase PHD Finger 8 (PHF8) (IC_{50} values were determined by Andrew Mun Chiang Chan and Dr Louise Walport), suggesting that **15** and **18** are selective inhibitors for AlkB, at least with respect to PHD2 and PHF8.

A pyridyl glycinamide series **20** – **28**, **32**, and **33** was also explored as AlkB inhibitors (Figure 2 - 8d). The substituted pyridine ring was predicted to occupy the nucleotide binding site. However, only compound **22** ($IC_{50} = 3.4 \mu M$) (Figure 2 - 8d) of the series had better potency than **15**.

Crystallographic studies were then carried out to investigate the binding modes of compound **18** and **22** in AlkB.

2.2.6 Crystal structure of the AlkB-18 complex

Crystallisation trials for AlkB Δ 11 were set up in the presence of diammonium iron (II) sulphate, and **18** using an Index screen (Hampton Research) and an

optimisation screen based on AlkB-**15** crystallisation conditions (see Table 2 - 2). An AlkB crystal in one of the wells in the Index screen containing 0.1 M bis-tris pH 6.5 and 25% w/v polyethylene glycol (PEG) 3350 diffracted to 1.7 Å. The space group was found to be *P1* with unit cell constants of $a = 37.07$ Å, $b = 39.23$ Å, $c = 40.26$ Å, $\alpha = 76.33^\circ$, $\beta = 74.56^\circ$, and $\gamma = 65.28^\circ$. The structure was solved by molecular replacement using the AlkB-**15** structure (PDB ID 3T4H; resolution: 1.60 Å) as a search model. The structure has the same *P1* space group as the AlkB-**15** structure. Compound **18** was observed in the active site (Figure 2 - 11a), chelating the active site iron with its oxalyl group, analogous to 2OG and **15** (Figure 2 - 10). The C5-carboxylate group of the *N*-oxalyl cysteine group of **18** is positioned to interact with the side chains of Tyr122, Asn120 and Arg204, similar to the C5-carboxylate of **15**. The cysteinyl group of **18** adopts a different conformation from **15**, most probably to align the bulkier 2-naphthalene side chain to form efficient π - π interactions with the side chains of His187 and Trp178. Overall, the binding mode of **18** was as predicted and the efficient π - π interactions likely contribute to its low IC₅₀ value (0.5 μ M), which is the lowest among the inhibitors explored in this study.

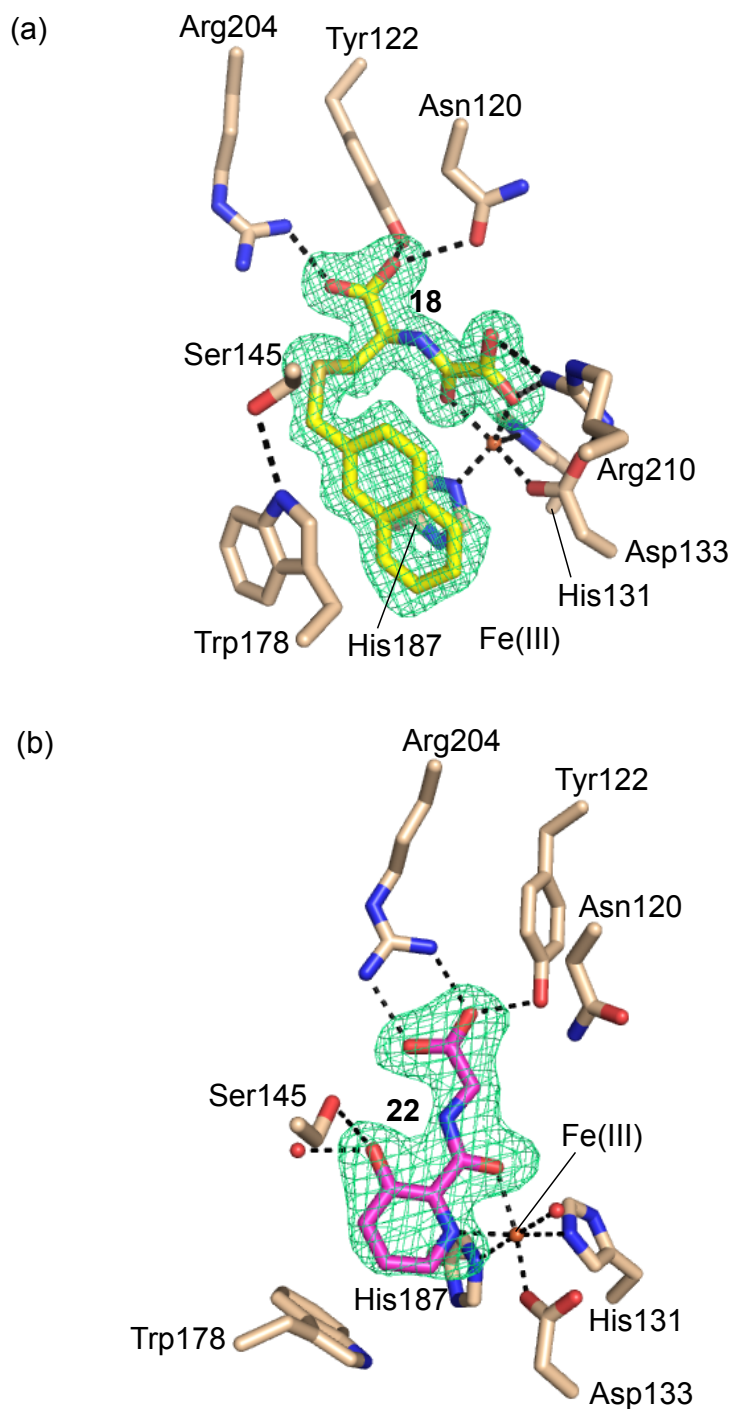


Figure 2 - 11. Views of crystal structures of AlkB in complex with 18 and 22. Views of the active site of crystal structures of **(a)** the AlkB-18 complex and **(b)** the AlkB-22 complex. Compound **18** (yellow), compound **22** (magenta), oxygen (red), nitrogen (blue), Fe(III) (orange sphere), water (red sphere), and electrostatic interactions (black dashes) are indicated. Electron density maps for **18** and **22** ($\sigma = 3.0$, $mF_o - DF_c$ OMIT, green mesh) are shown. Protein structure models were produced using MacPyMOL v1.3 (27).

2.2.7 Crystal structure of the AlkB-22 complex

Subsequently, AlkB was co-crystallised with the pyridyl-compound **22**, which was the only pyridyl series compound that showed a higher potency than compound **15**, but which was less potent than **18**. Crystallisation trials were set up for AlkB in the presence of **22** using a similar approach as for AlkB-**18**. However, no crystal formation was observed. During the crystallisation sample preparation, the protein solution turned yellow upon addition of compound **22**, suggesting that the compound might have been chelating with iron in solution instead of at the active site. Co-crystallisation experiments were thus performed again with a 1 : 1 protein to metal molar ratio to avoid excess metal in solution. Fe(II) was first added to the protein solution and incubated for 5 min on ice. Compound **22** was then added to the protein-metal solution. This approach resulted in a crystal that diffracted to 2.0 Å resolution. The space group of the crystal for the AlkB-**22** structure is *P*1 and the unit cell constants are $a = 36.81$ Å, $b = 38.59$ Å, and $c = 40.49$ Å, $\alpha = 71.25^\circ$, $\beta = 72.21^\circ$ and $\gamma = 66.70^\circ$. The crystal structure was solved by molecular replacement using the AlkB-**18** structure (PDB ID 3T4V; resolution: 1.70 Å) as a search model.

Compound **22** was observed to be in the AlkB active site but, surprisingly, its pyridine ring is positioned close to the hydrophobic pocket near Trp178 and Ser145, as opposed to the predicted position near the nucleotide binding site (Figure 2 - 11b). The pyridyl nitrogen of **22** is positioned *trans* to His131 and the amide carbonyl of **22** is positioned *trans* to Asp133, chelating the active site metal. Similarly to the C5-carboxylate of 20G, the carboxylate of **22** interacts with the side chains of Arg204 and Tyr122. The 3-hydroxyl group of **22** is positioned to participate in a hydrogen bond to the side chain of Ser145. The side

chain of Trp-178 is rotated (approximately 90° about the C_α-C_β bond and approximately 180° about the C_β-C_γ bond relative to Trp178 in AlkB-**18**) to form π-π interactions with the pyridine ring of **22**. The flexibility of Trp178 could be contributing to the observed binding mode of **22** by allowing the formation of a π stack that is energetically more favourable.

All the structures of AlkB (AlkB-**15**, AlkB-**18**, AlkB-**22**, and AlkB-IOX1) described in this chapter was refined using Phenix.Refine and Coot by performing iterative model fitting and refinement until the R and R_{free} values no longer converged (28,29). Parameters included in structural refinement were XYZ refinement, individual B factor refinement, X-ray/stereochemistry and X-ray/ADP weights refinement, secondary structure restraints and translation, libration, screw (TLS) refinement.

Table 2 - 2. Data collection and refinement details of structures of AlkB-inhibitor complexes.

	AlkB-15	AlkB-18	AlkB-22	AlkB-IOX1
PDB ID	3T4H	3T4V	3T3Y	4JHT
Crystal Size (Longest Dimension)	~100 μm	~100 μm	~100 μm	~100 μm
Data Collection				
Radiation Source	Rotating anode	Rotating anode	Rotating anode	Synchrotron Diamond Beamline I04
Detector	Saturn 944+ CCD	Saturn 944+ CCD	Saturn 944+ CCD	ADSC QUANTUM 315r
X-ray Wavelength ($\text{\AA}$)	1.5412	1.5412	1.5412	0.9795
Temperature (K)	100	100	100	100
Resolution Range ($\text{\AA}$)	21.14 - 1.60 (1.66 - 1.60) [*]	22.99 - 1.70 (1.76 - 1.70) [*]	22.91 - 2.00 (2.07 - 2.00) [*]	38.71 - 1.18 (1.22 - 1.18) [*]
Space Group	<i>P</i> 1	<i>P</i> 1	<i>P</i> 1	<i>P</i> 1
Unit Cell Dimensions (<i>a</i> $\text{\AA}$, <i>b</i> $\text{\AA}$, <i>c</i> $\text{\AA}$, α°, β°, γ°)	36.66, 38.83, 40.27, 77.64, 75.44, 66.42	37.07, 39.23, 40.26, 76.33, 74.56, 65.28	36.81, 38.59, 40.49, 71.25, 72.21, 66.70	36.84, 38.59, 40.45, 77.62, 75.01, 66.10
Total Number of Reflections Observed	89589	90015	46973	64400
Number of Unique Reflections	24608 (2161) [*]	20297 (1776) [*]	12688 (1260) [*]	60068 (5513) [*]
Multiplicity	3.6 (2.3) [*]	4.4 (1.9) [*]	3.7 (3.2) [*]	4.1 (3.1) [*]
Completeness (%)	95.6 (84.0) [*]	93.2 (81.2) [*]	99.2 (97.9) [*]	93.3 (85.7) [*]
<i>I</i>/σ(<i>I</i>)	14.2 (4.5) [*]	9.3 (2.9) [*]	25.2 (4.3) [*]	18.5 (3.8) [*]
[§]<i>R</i>_{merge} (%)	8.2 (22.6) [*]	10.7 (28.2) [*]	4.3 (21.8) [*]	8.0 (33.3) [*]
Wilson B Factor ($\text{\AA}^2$)	20.4	20.1	29.7	13.8
Software for Data Reduction	HKL-2000	HKL-2000	HKL-2000	HKL-2000
Refinement				
[*]<i>R</i>_{work} (%)	15.9	16.1	18.2	15.9
[†]<i>R</i>_{free} (%)	18.7	20.2	22.1	16.6
[¶]RMS Deviation	0.010 $\text{\AA}$ (1.44°)	0.011 $\text{\AA}$ (1.49°)	0.010 $\text{\AA}$ (1.25°)	0.009 $\text{\AA}$ (1.25°)
Ramachandran Outliers (%)	0.5	0.5	0.5	0
[‡]Average B Factor ($\text{\AA}^2$)	26.4 (31.6)	27.7 (22.5)	37.8 (31.6)	21.7 (15.1)
Number of Water Molecules	226	282	129	269
Software for refinement	PHENIX	PHENIX	PHENIX	PHENIX

$$\S R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle \times 100$$

$$* R_{\text{work}} = \sum ||\text{Fobs}| - |\text{Fcalc}|| / |\text{Fobs}| \times 100$$

† R_{free} , based on 10% (for AlkB-**15**, AlkB-**18**, AlkB-**22**) and 3.3% (for AlkB-IOX1) of the total reflections

‡ RMS deviation from ideality for bonds (followed by the value for angles)

‡ Overall average B factor (followed by average B factor of ligand)

‡ Outermost shell

2.3 Conclusions

In summary, the results of the structural and inhibition studies on AlkB reveal that an NAOX can be inhibited by broad spectrum 2OG oxygenase inhibitors. However, the NAOX AlkB showed differences in the crystallographically observed IOX1 binding mode compared to other 2OG oxygenases, particularly when compared to the JmjC oxygenases (20,26). IOX1 adopts an alternative binding mode wherein the positions of the 8-hydroxyl group and nitrogen of the quinoline of IOX1 are reversed in AlkB compared to those observed in JMJD2A, JMJD3 and the HIF hydroxylase FIH (20,26). Using *N*-oxalyl-*D*-cysteine as a support ligand and a library of thiols, the dynamic combinatorial mass spectrometry (DCMS) approach, combined with crystallography, successfully identified potent and selective JMJD2A/JMJD2E inhibitors that are *N*-oxalyl-*D*-tyrosine derivatives (30). Like JMJD2A, potent AlkB inhibitors can be developed using a similar DCMS-crystallography approach. The inhibition studies of AlkB described in this chapter formed a basis for the development of inhibitors for the human AlkB homologues FTO and ALKBH5 (Chapters 3 and 4).

2.4 Experimental Procedures

2.4.1 FDH-based Assays

Reactions (25 μ L) in triplicate containing final concentrations of 2 μ M AlkBdN11 (1 μ M AlkB dN11 for 2,4-PDCA), 0.1 U/100 μ L FDH, 10 μ M Fe(II), 500 μ M NAD⁺, 20 μ M 2OG, 100 μ M ascorbate, 20 μ M ssDNA with the sequence 5'-GC(m¹A)AGGTCCCGTAGTGCG-3' (ATD Bio, U.K.), and 50 mM HEPES pH 7.5 in the presence of varied concentrations of inhibitor (100 nM, 300 nM, 1 μ M, 3 μ M,

10 μ M, 30 μ M, 100 μ M, 300 μ M, or 1 mM) were set up. Reactions were started by mixing 12.5 μ L enzyme mix containing 2X final concentrations of AlkBdN11, FDH in 50 mM HEPES pH 7.5, with 12.5 μ L substrate mix containing 2X final concentrations of the rest of the components in 50 mM HEPES pH 7.5. Fluorescence readings (355 nm excitation, 460 nm emission) were recorded immediately after the start of the reaction at 30 s intervals over 15 min. Linear regions of the curves were used to calculate the initial reaction rates. The initial reaction rates were then plotted against inhibitor concentrations and the IC₅₀ values were calculated using GraphPad Prism®.

2.4.2 Protein Crystallography

Crystals of AlkB in complex with **15**, **18**, **22** and IOX1 (approximately 100 μ m in the longest dimension) were grown in sitting drops at 293 K using vapour diffusion methods. The ratio of protein to reservoir solution for the AlkB-**15** and AlkB-**18** co-crystallisation drops was 2:1 (300 nL total drop volume). For AlkB-**22** and AlkB-IOX1, the ratio of protein to reservoir solution was 1:1 (200 nL total drop volume). The reservoir volume was 80 μ L. The protein solution and reservoir solution conditions are listed in Table 2 - 2. The crystals were cryoprotected by soaking them in the well solution diluted with 25% v/v glycerol for approximately 30 s and they were then flash cooled in an Oxford Cryosystems nitrogen gas stream (AlkB-**15**, AlkB-**18**, and AlkB-**22**) or in liquid nitrogen (AlkB-IOX1). Data were collected from single crystals at 100 K using a Rigaku FR E+ Superbright diffractometer equipped with a copper rotating anode (X-ray wavelength: 1.5412 Å), Osmic HF optics, and a Saturn 944+ CCD detector for AlkB-**15**, AlkB-**18**, and AlkB-**22**. Data for AlkB-IOX1 were collected from a

single crystal at 100 K at Diamond Beamline I04 (X-ray wavelength: 0.9795 Å) equipped with an ADSC QUANTUM 315r detector. The data were indexed, integrated, and scaled using HKL2000 (31). The structure of AlkB-**15** was determined by molecular replacement using the AutoMR (PHASER) (32) subroutine in PHENIX (29) with a structure of AlkB (PDB ID 2FDJ) (8) as a search model. The AlkB-**15** model (PDB ID 3T4H) was in turn used as a search model for molecular replacement solution for AlkB-**18** and AlkB-**22**. The structure of AlkB-IOX1 was solved by molecular replacement using the AlkB-**18** structure (PDB ID: 3T4V; resolution 1.70 Å). Iterative rounds of model building and refinement using COOT (28) and PHENIX (29) were performed until the decreasing R and R_{free} no longer converged. (Methods were adapted from (33) and (20).)

Table 2 - 3. Crystallisation conditions of AlkB-inhibitor complexes

	AlkB- 15	AlkB- 18	AlkB- 22	AlkB-IOX1
PDB ID	3T4H	3T4V	3T3Y	4JHT
Protein Solution Conditions	~10.1 mg/ml AlkBΔ11, 2.2 mM ammonium iron(II) sulfate, 5.7 mM 15 , 50 mM HEPES pH 7.5	~10.1 mg/ml AlkBΔ11, 2.2 mM ammonium iron(II) sulfate, 1 mM 18 , 50 mM HEPES pH 7.5	~10.1 mg/ml AlkBΔ11, 0.44 mM ammonium iron(II) sulfate, 1 mM 22 , 50 mM HEPES pH 7.5	~10.1 mg/ml AlkBΔ11, 1 mM MnCl ₂ , 1 mM IOX1, 50 mM HEPES pH 7.5
Reservoir Solution Conditions	0.2 M sodium chloride, 0.1 M HEPES pH 7.5, and 25% (w/v) PEG 3350	0.2 M ammonium sulfate, 0.1 M tris-hydrochloride pH 8.5, and 25% (w/v) PEG 3350	0.1 M bis-tris, pH 6.5, and 25% (w/v) PEG 3350	0.264 M sodium chloride, 0.1 M HEPES pH 7.5, 23% (w/v) PEG 3350

Experimental procedures on AlkB expression and purification, non-denaturing ESI-MS studies, differential scanning fluorometry (DSF), inhibition assays for AlkB and inhibition assays for PHF8 and PHD2 can be found in the original

publication (Woon *et al. J. Med. Chem.* (2012), 55, 2173-2184) attached in the Appendix.

2.5 References

1. Falnes, P.O., Johansen, R.F. and Seeberg, E. (2002) AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature*, **419**, 178-182.
2. Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. and Sedgwick, B. (2002) Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature*, **419**, 174-178.
3. Aravind, L. and Koonin, E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol*, **2**, RESEARCH0007.
4. Begley, T.J. and Samson, L.D. (2003) AlkB mystery solved: oxidative demethylation of N1-methyladenine and N3-methylcytosine adducts by a direct reversal mechanism. *Trends Biochem Sci*, **28**, 2-5.
5. Li, D., Delaney, J.C., Page, C.M., Yang, X., Chen, A.S., Wong, C., Drennan, C.L. and Essigmann, J.M. (2012) Exocyclic carbons adjacent to the N6 of adenine are targets for oxidation by the *Escherichia coli* adaptive response protein AlkB. *Journal of the American Chemical Society*, **134**, 8896-8901.
6. Mishina, Y., Yang, C.G. and He, C. (2005) Direct repair of the exocyclic DNA adduct 1,N6-ethenoadenine by the DNA repair AlkB proteins. *Journal of the American Chemical Society*, **127**, 14594-14595.
7. Maciejewska, A.M., Ruszel, K.P., Nieminuszczy, J., Lewicka, J., Sokolowska, B., Grzesiuk, E. and Kusmierk, J.T. (2010) Chloroacetaldehyde-induced mutagenesis in *Escherichia coli*: the role of AlkB protein in repair of 3,N(4)-ethenocytosine and 3,N(4)-alpha-hydroxyethanocytosine. *Mutat Res*, **684**, 24-34.
8. Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R. and Hunt, J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879-884.
9. Yang, C.G., Yi, C., Duguid, E.M., Sullivan, C.T., Jian, X., Rice, P.A. and He, C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature*, **452**, 961-965.
10. Aas, P.A., Otterlei, M., Falnes, P.O., Vagbo, C.B., Skorpen, F., Akbari, M., Sundheim, O., Bjoras, M., Slupphaug, G., Seeberg, E. *et al.* (2003) Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature*, **421**, 859-863.

11. Zheng, G., Fu, Y. and He, C. (2014) Nucleic Acid Oxidation in DNA Damage Repair and Epigenetics. *Chem Rev.*
12. Kurowski, M.A., Bhagwat, A.S., Papaj, G. and Bujnicki, J.M. (2003) Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics*, **4**, 48.
13. Lee, D.H., Jin, S.G., Cai, S., Chen, Y., Pfeifer, G.P. and O'Connor, T.R. (2005) Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *J Biol Chem*, **280**, 39448-39459.
14. Duncan, T., Trewick, S.C., Koivisto, P., Bates, P.A., Lindahl, T. and Sedgwick, B. (2002) Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A*, **99**, 16660-16665.
15. Wei, Y.F., Carter, K.C., Wang, R.P. and Shell, B.K. (1996) Molecular cloning and functional analysis of a human cDNA encoding an Escherichia coli AlkB homolog, a protein involved in DNA alkylation damage repair. *Nucleic Acids Res*, **24**, 931-937.
16. Gerken, T., Girard, C.A., Tung, Y.C., Webby, C.J., Saudek, V., Hewitson, K.S., Yeo, G.S., McDonough, M.A., Cunliffe, S., McNeill, L.A. *et al.* (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*, **318**, 1469-1472.
17. Thalhammer, A., Aik, W., Bagg, E.A.L. and Schofield, C.J. (2012) The potential of 2-oxoglutarate oxygenases acting on nucleic acids as therapeutic targets. *Drug Discovery Today: Therapeutic Strategies*, **9**, e91-e100.
18. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. *et al.* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889-894.
19. Shimada, K., Nakamura, M., Anai, S., De Velasco, M., Tanaka, M., Tsujikawa, K., Oujii, Y. and Konishi, N. (2009) A Novel Human AlkB Homologue, ALKBH8, Contributes to Human Bladder Cancer Progression. *Cancer Research*, **69**, 3157-3164.
20. Hopkinson, R.J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K.H., Li, X.S., Kristensen, J.B.L., King, O.N.F., Chan, M.C. *et al.* (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chemical Science*, **4**, 3110.
21. Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C.G., Cui, Q. and He, C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*, **468**, 330-333.
22. Ando, M., Yoshimoto, T., Ogushi, S., Rikitake, K., Shibata, S. and Tsuru, D. (1979) Formaldehyde dehydrogenase from *Pseudomonas putida*. Purification and some properties. *J Biochem*, **85**, 1165-1172.
23. Lohmann, W., Lohmann, C. and Ibrahim, M. (1988) Fluorescence spectra of NADH/NAD, kynurenine, tryptophan, and tyrosine. *Naturwissenschaften*, **75**, 141-142.

24. Roy, T.W. and Bhagwat, A.S. (2007) Kinetic studies of Escherichia coli AlkB using a new fluorescence-based assay for DNA demethylation. *Nucleic Acids Res*, **35**, e147.
25. Rose, N.R., McDonough, M.A., King, O.N., Kawamura, A. and Schofield, C.J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev*, **40**, 4364-4397.
26. King, O.N., Li, X.S., Sakurai, M., Kawamura, A., Rose, N.R., Ng, S.S., Quinn, A.M., Rai, G., Mott, B.T., Beswick, P. *et al.* (2010) Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS ONE*, **5**, e15535.
27. Schrodinger, LLC. (2010).
28. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D*, **66**, 486-501.
29. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*, **66**, 213-221.
30. Rose, N.R., Woon, E.C., Kingham, G.L., King, O.N., Mecnovic, J., Clifton, I.J., Ng, S.S., Talib-Hardy, J., Oppermann, U., McDonough, M.A. *et al.* (2010) Selective inhibitors of the JMJD2 histone demethylases: combined nondenaturing mass spectrometric screening and crystallographic approaches. *Journal of Medicinal Chemistry*, **53**, 1810-1818.
31. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol*, **276**, 307-326.
32. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
33. Woon, E.C.Y., Demetriades, M., Bagg, E.A.L., Aik, W., Krylova, S.M., Ma, J.H.Y., Chan, M., Walport, L.J., Wegman, D.W., Dack, K.N. *et al.* (2012) Dynamic Combinatorial Mass Spectrometry Leads to Inhibitors of a 2-Oxoglutarate-Dependent Nucleic Acid Demethylase. *Journal of Medicinal Chemistry*, **55**, 2173-2184.

Chapter 3

Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO)

3.1 Introduction

The *FTO* gene was the first human gene found to be associated with body weight in genome-wide association studies (GWAS) (1,2). The *FTO* gene encodes for the fat mass and obesity associated protein (FTO) (1-4). The results of the GWAS led to the conclusion that single-nucleotide polymorphisms (SNPs) in the first intron of the *FTO* gene are associated with body mass index (BMI). Homozygous individuals carrying the risk allele A for the SNP rs9939609 are on average 3 kg heavier than homozygous individuals carrying a low-risk allele T (2). The results from GWAS were subsequently supported by several mouse model studies; *Fto* knockout mice show a higher frequency of postnatal death as well as reduced body weight, while overexpression of *Fto* led to obesity as a result of increased food intake (5,6). Subsequent studies found that the *FTO* gene encodes for a nucleic acid modifying Fe(II) and 2OG-dependent oxygenase that demethylates *N*³-methylthymine (m³T) and *N*³-methyluracil (m³U) in single-stranded DNA (ssDNA) (4). Further biochemical studies demonstrated that FTO has a higher activity on ssRNA than on ssDNA (7).

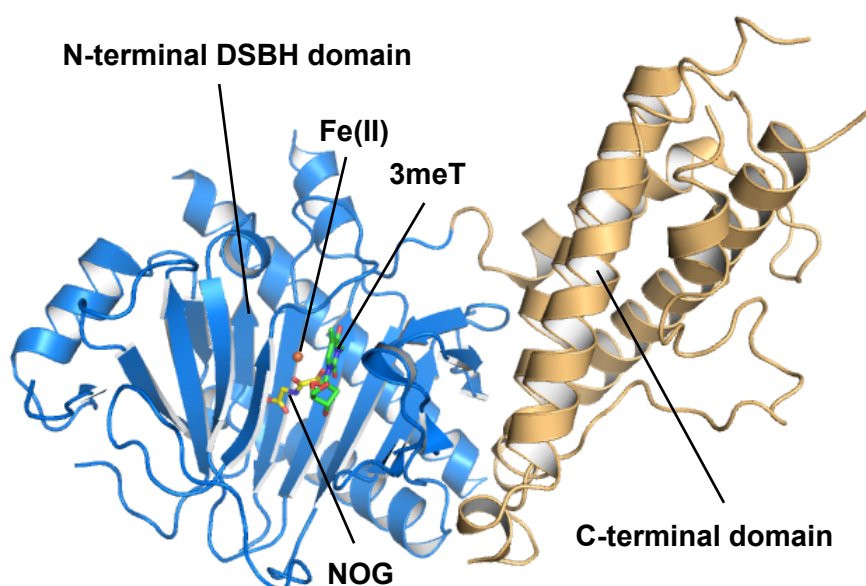
Precisely how the SNPs and the catalytic activity of FTO relate to an obese phenotype is still unknown. Recent developments have begun to shed light on possible molecular links between the nucleic acid demethylase activity of FTO and the obesity phenotype. In 2011, FTO was reported to demethylate *N*⁶-

methyladenine (m⁶A) in single-stranded RNA (ssRNA) (8). It was further demonstrated that *FTO* overexpression in cells reduced m⁶A levels in total mRNA, indicating that the m⁶A levels in mRNA are dynamic and thus m⁶A could be a regulatory marker (8). A recent report demonstrated that overexpression of *FTO* in cells also resulted in a reduction of m⁶A on ghrelin mRNA and a concomitant increase in ghrelin mRNA and peptide levels (9). Furthermore, homozygous individuals carrying the risk allele A for the SNP rs9939609 in the *FTO* gene have higher blood levels of ghrelin, a hunger stimulating peptide, compared to homozygous individuals carrying the non-risk allele T (9). The correlation between reduced m⁶A levels and increased ghrelin mRNA could be explained by the findings of an independent study demonstrating that m⁶A on mRNA is a marker for mRNA degradation (10). This study showed that m⁶A readers (proteins that bind specifically to m⁶A) such as YTHDF2 target hypermethylated mRNA for degradation (10). In summary, it is possible that *FTO* regulates levels of specific proteins associated with food intake and metabolism (e.g. ghrelin) through the demethylation of m⁶A on mRNAs that encode for those associated proteins.

Because *FTO* activity is linked to obesity, inhibition of *FTO* is a potential anti-obesity treatment. Developing small molecule inhibitors of *FTO* would be useful to validate *FTO* as an anti-obesity target. Structure-guided inhibitor design work on AlkB, the *E. coli* nucleic acid demethylase, is proof-of-principle that the AlkB nucleic acid demethylase subfamily is amenable to small-molecule inhibition (see Chapter 2) (11). In 2010, a crystal structure of *FTO* in complex with *N*-oxalylglycine and a substrate (3-methylthymidine nucleoside, 3meT) at 2.50 Å resolution was reported (12). This structure revealed that *FTO* consists of

two domains: an N-terminal AlkB-like oxygenase domain that contains the double-stranded beta helix (DSBH) fold common to all 2OG and Fe(II) dependent oxygenases, and a C-terminal 5-helical bundle domain of which the function is still unclear (Figure 3 - 1a). Iron is bound within the active site and is coordinated by the conserved 2OG oxygenase Fe(II) binding motif HXD...H (His231, Asp233, His307) (Figure 3 - 1b). In this structure, the broad-spectrum inhibitor, *N*-oxalylglycine (NOG) occupies the cosubstrate 2-oxoglutarate (2OG) binding site and a 3-methylthymidine (3meT) nucleoside is observed in the nucleotide substrate binding site (Figure 3 - 1b). While this structure may be useful for preliminary computational and manual docking studies for inhibitor design, it is not perfect with respect to predicting the binding modes of different classes of inhibitors. Therefore, crystal structures of FTO in complex with various classes of small molecule inhibitors will enhance our understanding of their binding modes, allowing for inhibitor design optimisation. This chapter presents structure-based inhibition studies on FTO.

(a)



(b)

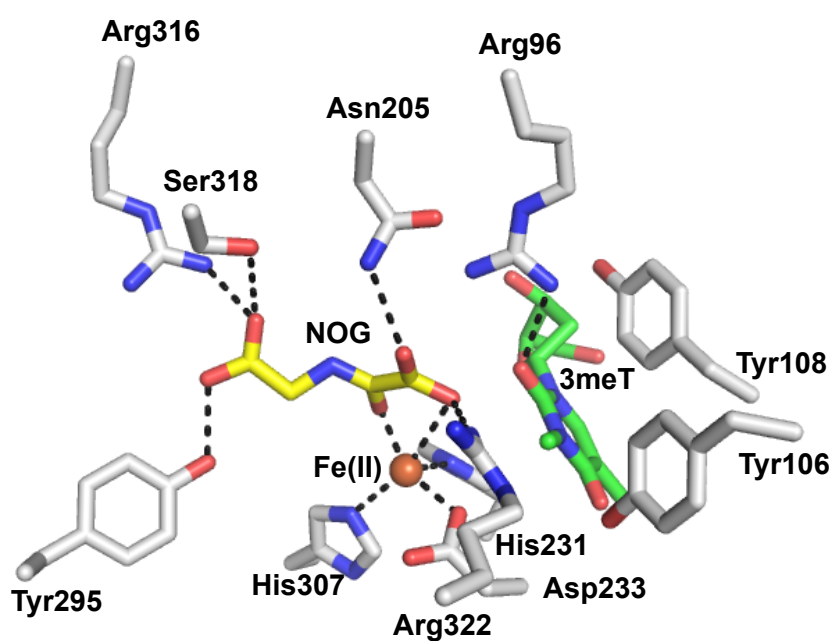


Figure 3 - 1. View of a crystal structure of FTO in complex with NOG and N³-methylthymidine (3meT). **(a)** Overall view of a crystal structure of FTO (PDB ID 3LFM; resolution: 2.50 Å) (12). N-terminal DSBH domain, blue; C-terminal domain, light orange. **(b)** View of the active site of a structure of the FTO-NOG-3meT complex (PDB ID 3LFM) (12). Active site residues, white sticks; NOG, yellow sticks; 3meT, green sticks; Fe(II) ion, orange sphere; electrostatic interactions, black dashes. Protein structure models were produced using MacPyMOL v1.3 (13).

3.2 Results and Discussions

Some of the results and associated experimental procedures described in this chapter are reprinted (adapted) with permission from Aik *et al.* (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, **56**, 3680-3688. (14) Copyright (2013) American Chemical Society.

3.2.1 Expression and purification of full-length and truncated ($\Delta 31$) FTO

Two FTO protein constructs, N-terminal hexahistidine-tagged full-length human FTO (hFTO) and N-terminal hexahistidine-tagged human FTO with an N-terminal 31 residue truncation (hFTO $\Delta 31$), were used in this study. hFTO was used for *in vitro* biochemical studies while hFTO $\Delta 31$ was used for crystallography. The previously reported hFTO $\Delta 31$ protein construct was used because it had resulted in the reported crystal structure of FTO in complex with NOG and the 3meT nucleoside (2.5 Å resolution) (12).

hFTO was recombinantly expressed in *Escherichia coli* BL21 (DE3) and purified using nickel-affinity chromatography followed by a Superdex 75 (S75) size exclusion chromatography (Figure 3 - 2). The S75 UV-chromatogram showed two overlapping peaks, both corresponding to hFTO according to SDS-PAGE gel analysis (Figure 3 - 2b,d). Based on previous reports, these are likely to represent monomeric (MW = 59.3 Da) and dimeric (MW = 118.6 Da) forms of hFTO (15). Fractions containing hFTO from both peaks were pooled together and concentrated for use in dynamic scanning fluorometry (DSF) and *in vitro* activity assays.

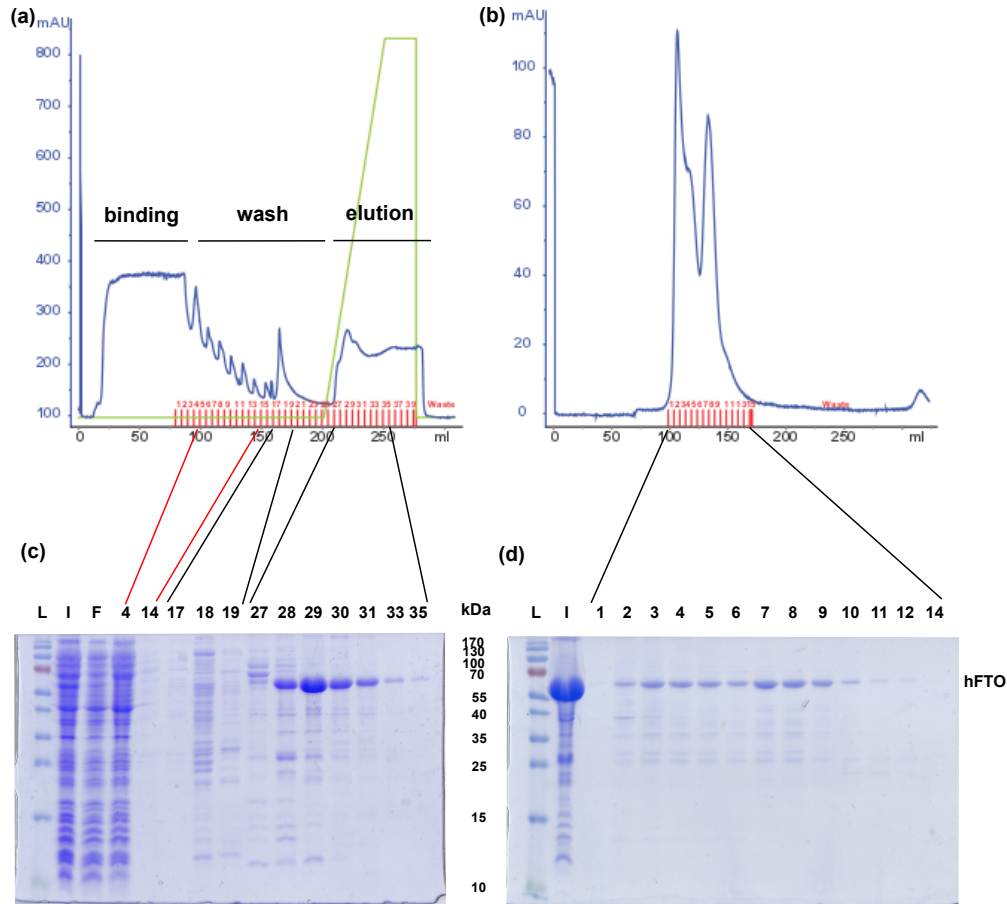


Figure 3 - 2. Full-length human FTO purification. AKTA FPLC UV trace (blue) of **(a)** nickel affinity purification and **(b)** Superdex 75 (S75) size exclusion chromatography. The green trace indicates elution buffer concentration; the red lines and numbers indicate fractions. SDS-PAGE analysis of fractions from **(c)** nickel affinity chromatography and **(d)** S75 size exclusion chromatography. The protein ladder as molecular weight marker (kDa) is labelled L, input material is labelled I, and flow through from HisTrap column is labelled F, and followed by fraction numbers.

In order to produce the hFTO Δ 31 protein, mutagenesis was performed on the pET28a-hFTO plasmid to obtain the pET28a-hFTO Δ 31 plasmid. Expression trials were then performed to determine the effect of varying isopropyl β -D-1-thiogalactopyranoside (IPTG) concentrations (0.5 – 1mM), induction temperature (18 – 37 °C) and time of incubation after induction (4 h or overnight). SDS-PAGE analysis (Figure 3 - 3a) of the soluble and insoluble fractions of the *E. coli* cell lysates revealed that the optimally determined

condition for hFTO Δ 31 expression is 18 °C overnight, after induction with 0.5 mM IPTG when the OD₆₀₀ reaches 0.6 – 0.8.

Large-scale production of hFTO Δ 31 was then performed using a 7.2 L culture media and the optimised expression/induction conditions. hFTO Δ 31 was purified by nickel-affinity chromatography (Figure 3 - 3b), followed by S75 size exclusion chromatography (Figure 3 - 3c). The S75 UV-chromatogram showed a broad and asymmetrical peak corresponding to hFTO Δ 31, suggesting that the hFTO Δ 31 protein purified is heterogenous (Figure 3 - 3c).

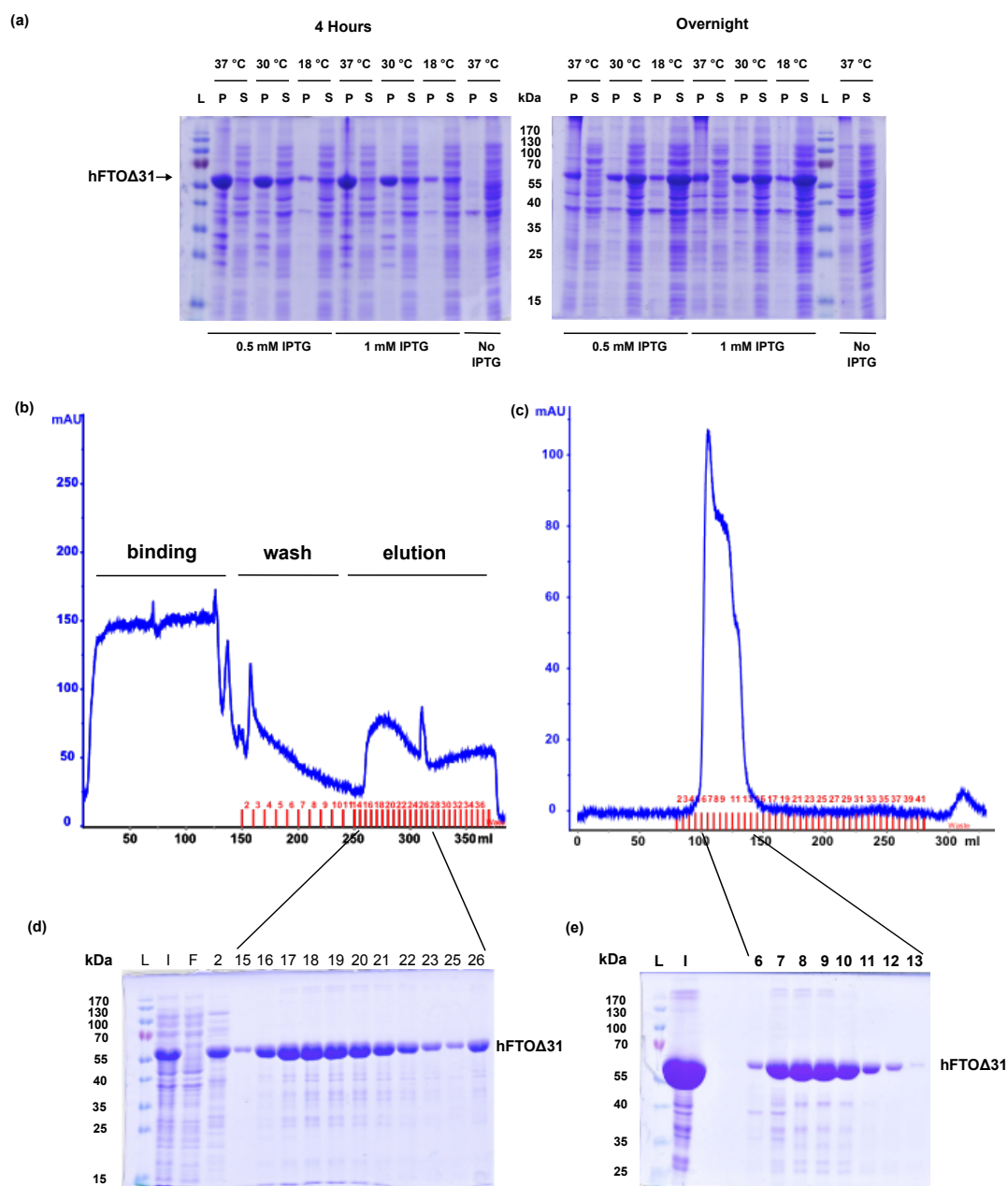


Figure 3 - 3. Expression and purification of hFTOΔ31. (a) SDS-PAGE analysis of expression trials of hFTOΔ31 in *E. coli* BL21 (DE3). Insoluble fractions are indicated by P and soluble fractions are indicated by S. The protein ladder as molecular weight marker is indicated by L. AKTA FPLC UV-chromatogram (blue) from (b) nickel affinity chromatography and (c) Superdex 75 (S75) size exclusion chromatography. The red lines and numbers indicate fractions. SDS-PAGE analysis of fractions from (d) nickel affinity purification and (e) S75 size exclusion chromatography. The protein molecular weight standards ladder (kDa) is labelled L, input material is labelled I, and flow through from HisTrap column is labelled F, and followed by fraction numbers.

3.2.2 Investigating the enzymatic activities of recombinant hFTO and hFTOΔ31

In order to determine whether the purified recombinant hFTO and hFTOΔ31 were catalytically active, a pilot *in vitro* demethylation assay was carried out. Given that FTO has been reported to demethylate *N*³-methylthymidine deoxyribonucleoside (3meT) (12), a liquid chromatography (LC)-based assay detecting 3meT and thymidine nucleosides was developed to test for FTO activity.

Firstly, the 3meT nucleoside was synthesized so that it could be tested as a substrate of FTO. Following the reported procedure (16) as shown in the reaction scheme (Figure 3 - 4), a reaction to methylate thymidine using methyl iodide was performed for 24 h at room temperature. The product, 3meT nucleoside, was purified using high-performance liquid chromatography (HPLC) to give a yield of 24%. The purity of the product was confirmed by proton and carbon NMR (see experimental Section 3.4.7). (Synthesis was carried out with the guidance of Marcus Sims and Dr Luc Henry.)

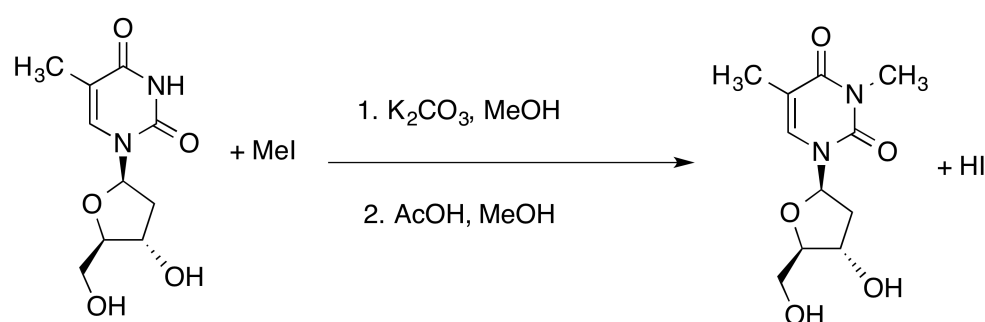


Figure 3 - 4. Reaction scheme for the synthesis of *N*³-methylthymidine from thymidine.

The purified 3meT nucleoside was then tested *in vitro* as a substrate for hFTO and hFTOΔ31 according to the reported conditions (12). The *in vitro*

demethylation reactions were carried out at room temperature and at 37 °C for 15 h. Pilot assays were performed with 5 μ M hFTO Δ 31, 80 μ M 3meT, 1 mM 2OG, 2 mM L-ascorbate and 500 μ M Fe(II) in 50 mM MES pH 6.3 buffer. 3meT (substrate) and thymidine (product) were separated by Ultra-Performance Liquid Chromatography (UPLC) using an oligonucleotide separating column (Waters Acquity OST C18, 130 Å X 1.7 μ m X 2.1 mm X 150 mm), and a methanol gradient (see experimental Section 3.4.4 for details). The results of the pilot assays showed that hFTO Δ 31 was active but the full-length hFTO was not, probably due to experimental error (full-length hFTO was subsequently re-tested by an Oxford Chemistry Part II student Muhammad Khairul Kamilin Hamdan and found to be active). As shown in Figure 3 - 5, peaks corresponding to thymidine were present in the samples containing hFTO Δ 31, but not in the no-enzyme control. Preliminary results showed that room temperature incubation resulted in more turnover of 3meT to thymidine compared to the reaction at 37 °C. Together, these results verify that hFTO Δ 31 is active and that 3meT nucleoside is a substrate for hFTO Δ 31. This method was further developed into an inhibition assay for full-length hFTO (see Section 3.2.3).

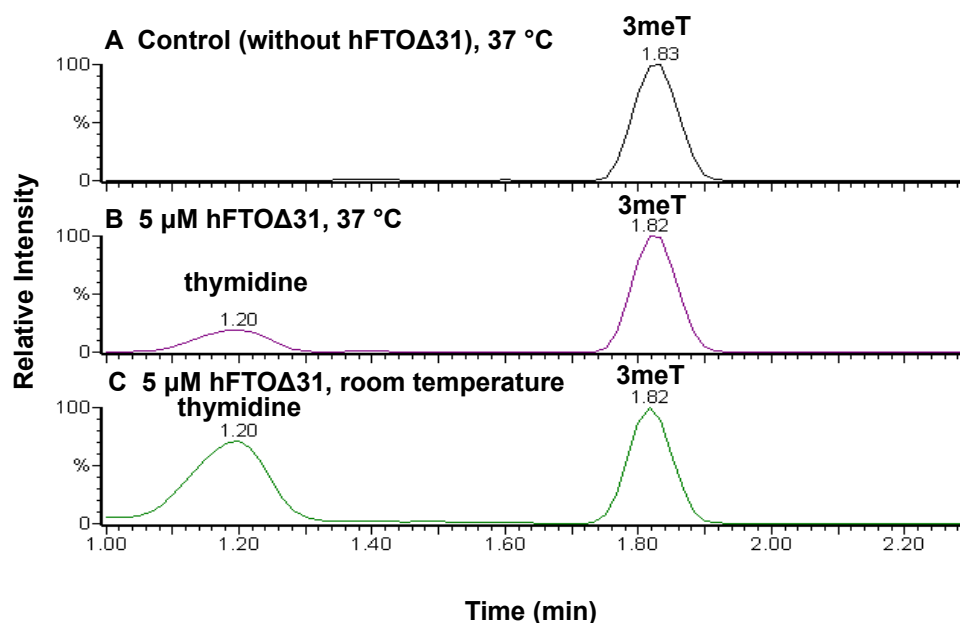


Figure 3 - 5. UPLC analysis of hFTO Δ 31 demethylation reactions showing UV chromatograms of separated reaction components. Peaks at retention time (Rt) = 1.82 min correspond to 3meT and peaks at Rt = 1.20 correspond to thymidine, the product of the demethylation reaction by hFTO Δ 31. Demethylation was observed for samples containing hFTO Δ 31 (B and C) but not in the control without enzyme (A).

3.2.3 Optimisation of *in vitro* demethylation assay for inhibition studies

The *in vitro* demethylation assay described above was then optimised to be suitable for quantification of inhibitory potency (experiments were carried out in collaboration with Muhammad Khairul Kamlin Hamdan). hFTO was used for the *in vitro* demethylation assay and the reactions were analysed by UPLC using a Waters ACQUITY UPLC BEH (Ethylated Bridged Hybrid) C18 column (130 Å, 1.7 μ m, 1 mm X 50 mm).

An optimal inhibition assay uses the lowest possible enzyme concentration that still allows for robust detection and quantification of enzymatic activity. The use of a lower quantity of enzyme not only saves on materials, but also enhances the lower limit in differentiating inhibitors with higher potency (lower IC₅₀ values). Therefore, the limit of detection for substrate

3meT and product thymidine was first investigated using a range of hFTO enzyme concentrations, i.e. 1 – 5 μM , in *in vitro* demethylation reactions. The minimum hFTO concentration that still allowed for detection of thymidine product was 3 μM . A time course assay was then conducted using 3 μM hFTO, 1 mM 2OG, 2 mM L-ascorbate, 500 μM Fe(II), and 80 μM 3meT. The time course curve for 3 μM hFTO remains linear up to 1 h (Figure 3 - 6a). Hence, a 1 h reaction time was used for subsequent reactions for the optimisation of the other assay components (i.e. 2OG, 3meT, Fe(II) and ascorbate).

Kinetic studies were then carried out for FTO by varying 2OG, 3meT, ascorbate, and Fe(II) concentrations. In each case, the initial rates of reaction were plotted against the concentration of the variable component (Figure 3 - 6b-e). The K_m values of 2OG and 3meT were determined as 151 μM and 63 μM , respectively. The concentration of Fe(II) cofactor for optimal activity was found to be >100 μM , whereas for ascorbate, it was > 500 μM . For subsequent FTO inhibition studies, 160 μM 2OG, 70 μM 3meT, 100 μM Fe(II), and 500 μM ascorbate, and 3 μM of FTO were used.

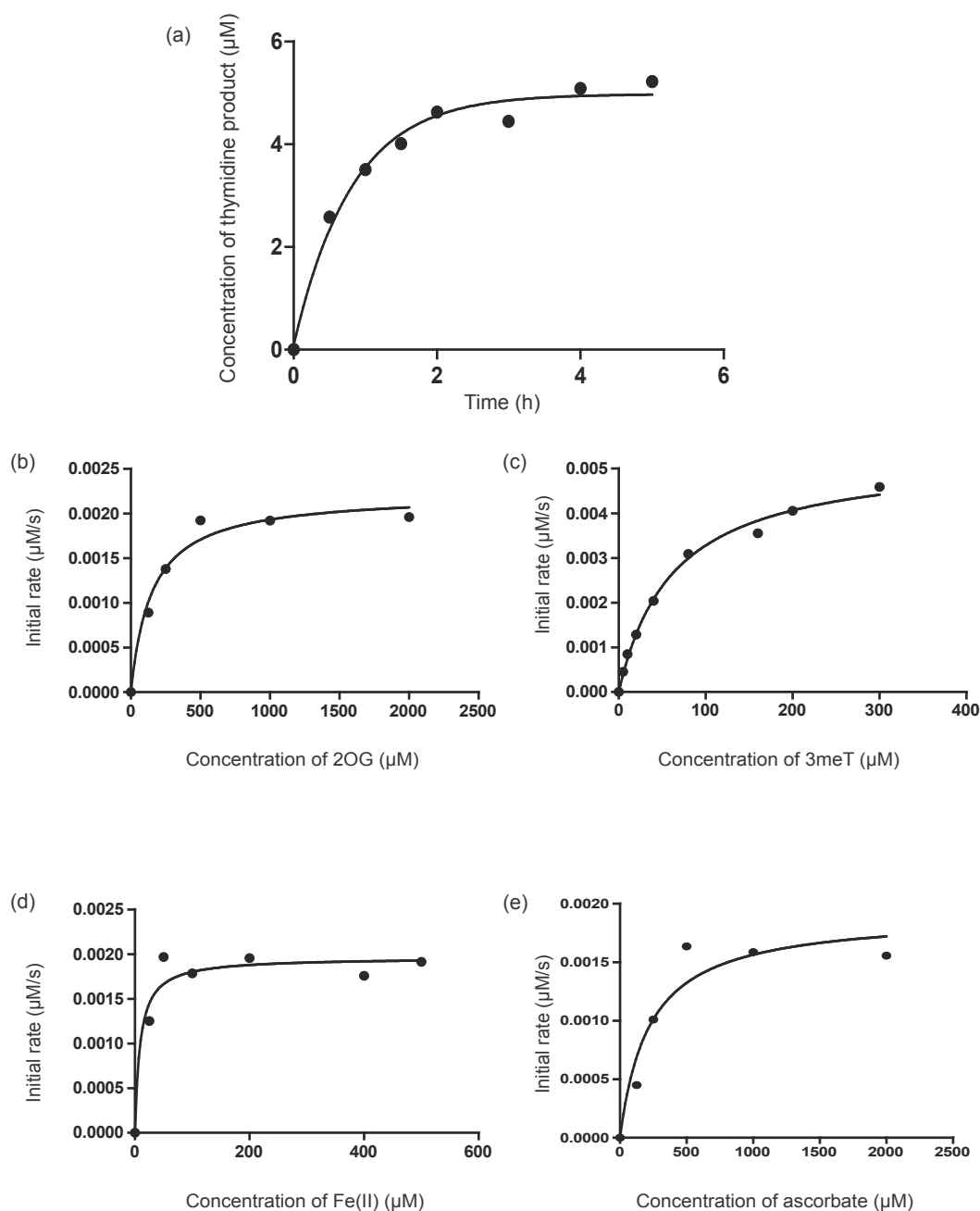


Figure 3 - 6. FTO kinetic studies. (a) FTO time course assay using 3 μM full-length human FTO. (b) Michaelis-Menten curve of initial rate ($\mu\text{M/s}$) versus concentration of 2-oxoglutarate (2OG) cosubstrate (μM). K_m of 2OG was determined to be 151 μM . (c) Michaelis-Menten curve of initial rate ($\mu\text{M/s}$) versus concentration of 3-methylthymidine (3meT) prime substrate (μM). K_m of 3meT was determined to be 63 μM . (d) Graph of initial rate ($\mu\text{M/s}$) versus concentration of Fe(II) (μM). (e) Graph of initial rate ($\mu\text{M/s}$) versus concentration of ascorbate (μM).

3.2.4 Inhibition studies of FTO

To identify potential FTO inhibitors, a single point inhibition assay (SPA) and a dynamic scanning fluorometry (DSF) assay were used to screen a library of 105 2OG analogues and metal chelators. (SPA was performed by Dr Marina Demetriades while DSF was performed by Muhammad Khairul Kamlin Hamdan)

To carry out the SPA assay, 200 μ M potential inhibitor was added to the *in vitro* demethylation reactions; compounds that gave a residual hFTO activity of 25% or less relative to no inhibitor control, were considered as hits, i.e. worthy of further analyses.

In order to screen the library of 2OG mimetic compounds in an efficient manner, thermal denaturation experiments were performed using DSF (17), which measures the melting temperature (T_m) of a protein by monitoring changes in the fluorescence signal of a dye (Sypro Orange) when it interacts with hydrophobic regions of a protein, indicating unfolding of the protein. The assay is conducted by slowly increasing the temperature of the protein solution, causing the protein to unfold and thus, the fluorescence signal to increase. T_m is assumed to be the temperature when half the protein molecules are unfolded (Figure 3 - 7). At this temperature, it is proposed that the Gibbs Free Energy of unfolding is equal to zero ($\Delta G_u=0$). The stability of the protein can be increased in the presence of inhibitors, and the magnitude of the change in T_m (ΔT_m) is often indicative of the binding affinity of the compound (i.e. higher T_m indicates higher affinity). Therefore, compounds that increase the T_m of hFTO are more likely to be strong binders and could potentially be inhibitors.

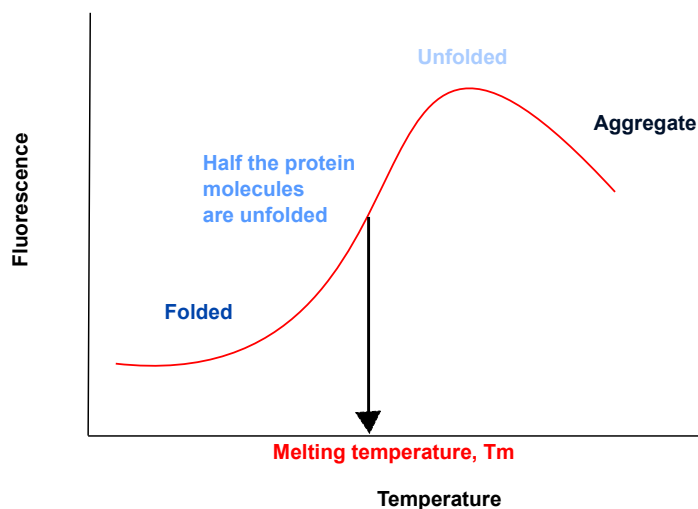


Figure 3 - 7. A typical DSF curve of fluorescence intensity versus temperature. The temperature corresponding to the inflection point of the sigmoidal portion of the curve is taken as the melting temperature, T_m , of the protein.

Template 'hits' obtained from the SPA and DSF assays include hydroxyquinolines, pyridyl glycinamides, isoquinolines and hydroxythiazoles (Table 3 - 1), which are various types of 2OG oxygenase inhibitors (18). Compounds from these chemical families were selected for half maximal response concentration (IC_{50}) determination. In addition to the inhibitors identified from the screening assays, the IC_{50} values of broad spectrum 2OG oxygenase inhibitors including *N*-oxalylglycine (**1**), 2,4-pyridine dicarboxylate (**2**) and 5-carboxy-8-hydroxyquinoline (**3**) were also determined (19). The IC_{50} and ΔT_m values of all the compounds tested are listed in Table 3 - 1.

(Compounds were synthesised by Dr Marina Demetriades, Dr Kar Kheng Yeoh, Dr Clarisse Lejeune, and Jacob Bush.)

Table 3 - 1. Structures, T_m shift values ($\Delta T_m/^\circ\text{C}$), residual activity at 200 μM concentration (SPA/ $^\circ\text{C}$), and IC₅₀ values of selected FTO inhibitors. Hydroxyquinolines are shaded light blue, pyridyl glycinamides are shaded light red, isoquinolines are shaded light green and hydroxythiazoles are shaded light purple.

No.	Chemical Structure	ΔT_m ($^\circ\text{C}$)/ SPA (%)	IC ₅₀ (μM)	No.	Chemical Structure	ΔT_m ($^\circ\text{C}$)/ SPA (%)	IC ₅₀ (μM)
1		NA	44 ± 1	11		12 $^\circ\text{C}$	2.8 ± 1
2		NA	8.3 ± 1	12		28 $^\circ\text{C}$	38 ± 1
3		NA	3.3 ± 1	13		26 $^\circ\text{C}$	60 ± 1
4		14%	23 ± 1	14		23 $^\circ\text{C}$	120 ± 1
5		17%	18 ± 1	15		29 $^\circ\text{C}$	84 ± 1
6		25%	15 ± 1	16		29 $^\circ\text{C}$	110 ± 1
7		NA	120 ± 1	17		17 $^\circ\text{C}$	160 ± 1
8		9 $^\circ\text{C}$	92 ± 1	18		NA	9.8 ± 1
9		16%	96 ± 1	25		10 $^\circ\text{C}$	130 ± 1
10		19%	43 ± 1	26		14 $^\circ\text{C}$	6.7 ± 1

The most potent inhibitors of FTO identified were compounds **3** and **11**, which had IC_{50} values of 3.3 and 2.8 μ M, respectively. These values were at least 10 fold lower than that of the broad spectrum 2OG oxygenase inhibitor, NOG (**1**) (IC_{50} 44 μ M). The most potent compounds from each class of inhibitor in Table 3 - 1 (compounds **1**, **2**, **3**, **6** and **11**) were chosen for co-crystallisation trials with FTO to investigate their mode of binding.

At approximately the same time as the inhibition studies on FTO described above were being carried out, several anthraquinones were reported as FTO inhibitors that were shown to be active in cells (20). Compounds **19** (rhein) and **20** were the two most potent anthraquinone compounds reported (20). Based on the results from our inhibition assay, compounds **19** and **20** had IC_{50} values of 9.0 ± 1 μ M and 11 ± 1 μ M respectively, thus confirming that they are FTO inhibitors. As there was no reported crystal structure of FTO (nor indeed any other 2OG oxygenase) in complex with anthraquinones, the precise binding mode of **19** and **20** in FTO was not known. Therefore, co-crystallisation trials of **19** and **20** with FTO were also carried out alongside compounds **1**, **2**, **3**, **6** and **11**.

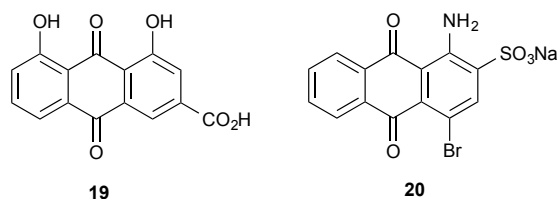


Figure 3 - 8. Chemical structures of anthraquinone compounds that have been reported as FTO inhibitors.

3.2.5 Initial Crystallisation Attempts of hFTOΔ31

Crystallisation trials were set up with the purified hFTOΔ31 protein (see Section 3.2.1) in the presence of NOG, nickel chloride, and 3meT using various commercial crystallisation broad screens and an optimisation screen based on the published crystallisation conditions (100 mM trisodium citrate pH 5.6, 10% (w/v) polyethylene glycol 3350 (PEG3350)) (12). The commercial crystallisation screens employed included JCSG+ (Molecular Dimensions, MD), PACT Premier (MD), Structure (MD)/Crystal (Hampton Research, HR), Salt Rx (HR), PEG/Ion (HR), and Wizard (Emerald Biosystem). However, no crystal formation was observed. Various attempts to optimise the purification of hFTOΔ31 were subsequently carried out (summarised in Figure 3 - 9).

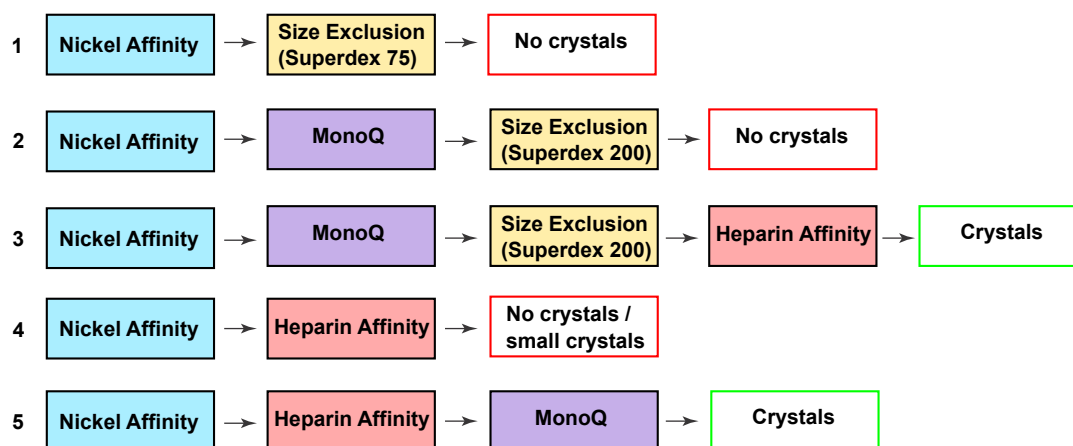


Figure 3 - 9. Flow charts summarising attempts to optimise the purification hFTOΔ31 to obtain crystallisation quality protein.

Because the first attempt in purifying hFTOΔ31 did not result in protein suitable for crystallisation, most likely due to impurities and heterogeneity of the protein, a second attempt to purify hFTOΔ31 was carried out following methods reported by Han *et al.* who determined the first crystal structure of human FTO,

which involved purification using nickel affinity chromatography, MonoQ anion exchange and then finally by S200 size exclusion chromatography (12). The UV-chromatogram from the MonoQ anion exchange showed two main peaks, each containing a few subpeaks (Figure 3 - 10a). Fractions 17 – 20, corresponding to the first major peak, were further purified using Superdex 200 size exclusion chromatography (Figure 3 - 10b). SDS-PAGE analysis showed that all fractions from the MonoQ and S200 chromatography still contained impurities or degradation products (Figure 3 - 10c). Subsequent crystallisation trials using the resulting protein from this purification method did not produce any crystals.

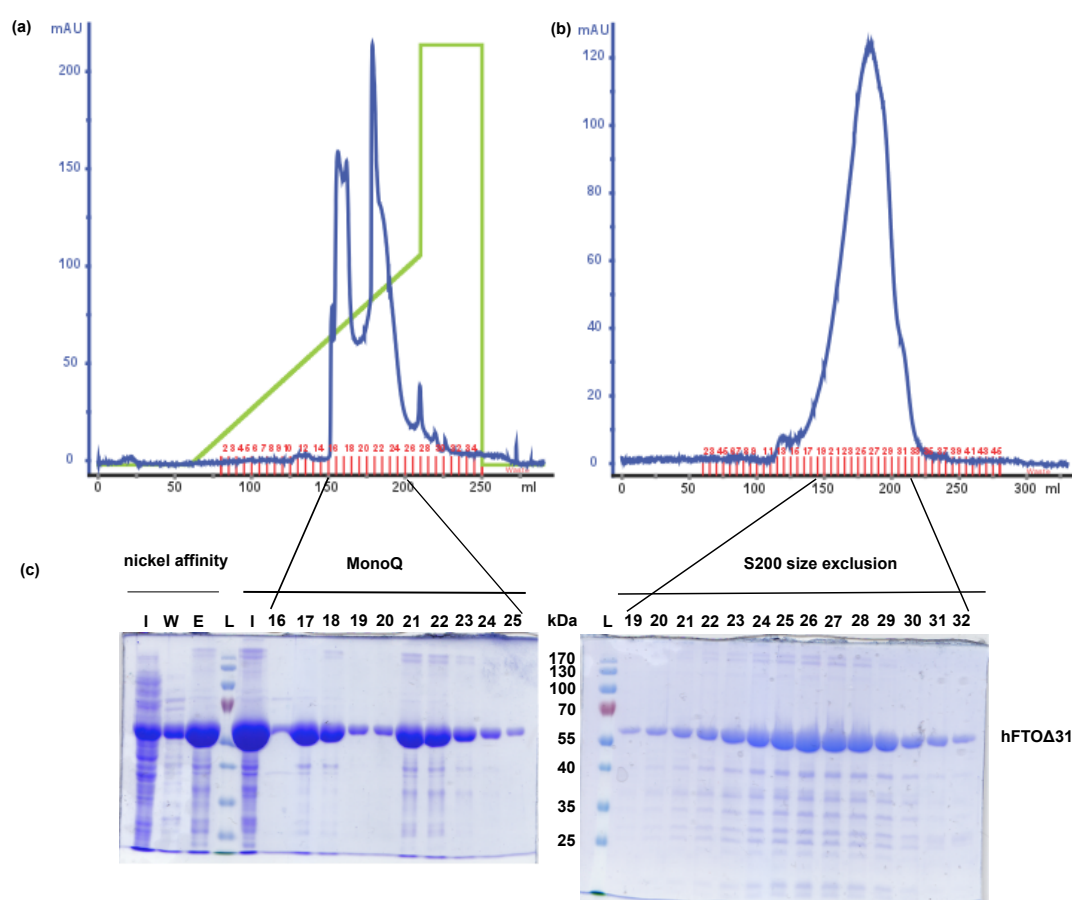


Figure 3 - 10. Attempt 2 in purifying hFTOΔ31. AKTA FPLC UV trace (blue) of (a) MonoQ ion exchange and (b) Superdex 200 (S200) size exclusion chromatography. The green line in (a) indicates NaCl concentration between 0 – 1 M. (c) SDS-PAGE analysis of fractions from nickel affinity chromatography, MonoQ anion exchange and S200 size exclusion chromatography. The protein molecular weight standards ladder is labelled L, input material is labelled I, and wash fraction using 40 mM imidazole is labelled W, elution fraction with 375 mM imidazole is labelled E, and is followed by fraction numbers.

The broad and asymmetrical peaks observed from the MonoQ and S200 chromatograms (Figure 3 - 10a,b) for hFTO Δ 31 protein suggested that the hFTO Δ 31 protein was still not homogenous. This observation subsequently led to the speculation that the purified hFTO Δ 31 protein could be interacting with nucleic acids since FTO is a nucleic acid demethylase (4). One method used to separate nucleic acids from protein employs cation exchange chromatography, a purification method that has previously been used in our laboratory used to purify AlkB protein (Dr Eleanor Bagg) that was subsequently used for crystallographic studies (see Chapter 2) (11). Cation exchange columns, such as the S Sepharose column, contain negatively charged resin that has a high affinity for nucleic acid-binding proteins (21,22). However, a general requirement for cation exchange chromatography is that the protein must carry an overall positive charge in order to bind to the resin, i.e. the pH of the buffer needs to be below the isoelectric point (pI) of the protein. The pI of hFTO Δ 31 protein is 5.4 (calculated using ExPASy Compute pI/MW (23,24)). Hence, hFTO Δ 31 protein was buffer exchanged into trisodium citrate buffer pH 5.0. However, the protein precipitated upon buffer exchange, possibly due to the very low pH. This method was therefore deemed unsuitable for purification of FTO.

3.2.6 Recombinant FTO Co-purifies with Nucleic Acids

Because S Sepharose cation exchange was unsuitable for purification of hFTO Δ 31, an alternative method to separate nucleic acid from hFTO Δ 31, which was by heparin chromatography, was attempted. Heparin affinity

chromatography is a protein purification technique capable of separating nucleic acids from protein (21,22). Heparin, a highly sulfated glucosaminoglycan, can mimic the negative charge of nucleic acid phosphate backbone and compete with nucleic acids for binding to protein. Any protein bound to the resin is then free of nucleic acids and can be eluted using a salt gradient.

hFTO Δ 31 protein that had already been purified by nickel affinity, MonoQ anion exchange, and S200 size exclusion chromatographies, was loaded onto a heparin column (GE Healthcare) and eluted using a salt gradient at pH 7.5. Fortunately, hFTO Δ 31 bound to the heparin resin even at pH 7.5 despite its much lower pI of 5.4. The eluted hFTO protein was concentrated and desalted for crystallisation trials using an optimisation screen based on the reported crystallisation conditions (12), in the presence of 3 mM NOG and 5 mM *N*⁶-methyladenosine nucleoside (6meA) (at that time, m⁶A has just been reported as a substrate of FTO (8)). Crystals ranging from 25 – 50 μ m appeared overnight. This result suggests that nucleic acid co-purification may have been contributing to the heterogeneity of the hFTO Δ 31 protein prior to the use of heparin chromatography.

The above speculation was further supported by a 2% agarose gel analysis of denatured hFTO protein purified using only nickel affinity. This analysis revealed a band of approximately 600-800 bp (Figure 3 - 11a), which was not present in the crystallisable hFTO Δ 31 sample purified with additional heparin and MonoQ columns (Figure 3 - 11a). The same analysis was later performed on a non-nucleic acid modifying oxygenase, *Pseudomonas putida* prolyl-4-hydroxylase (PPHD) (provided by John Scotti), purified using only a nickel affinity chromatography. No nucleic acid was observed on 2% agarose gel

for PPHD (Figure 3 - 11a). This result supports the earlier proposal that FTO co-purifies with nucleic acids.

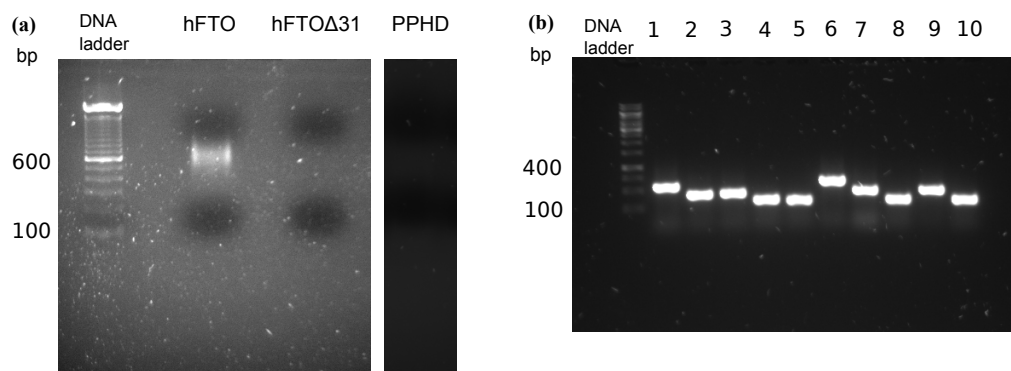


Figure 3 - 11. Recombinant FTO co-purifies with nucleic acids. (a) 2% Agarose gel analysis of the supernatant fractions of heat denatured full-length FTO (hFTO) purified using only a nickel affinity column, hFTOΔ31 purified to crystallisable quality, and *Pseudomonas putida* prolyl-4-hydroxylase purified using nickel affinity only. A smeared band at 600-800 bp can be observed on the lane for hFTO. (b) 2% Agarose gel analysis of cDNA inserts cloned into pGEMT-Easy plasmids. The inserts were amplified using PCR colony screening of *E. coli* XL10 Gold competent cells. Inserts of 100-300 bp are present in 10 selected white colonies.

Further analysis of the associated nucleic acid through reverse transcription, ligation into pGEM-T Easy vectors, polymerase chain reaction (PCR) and DNA sequencing revealed that the copurified nucleic acid was *E. coli* 23S ribosomal RNA (Figure 3 - 11b, Appendix Table A - 1). Although this result does not exclude the possibility of the presence of other types of nucleic acid, it does show that FTO interacts with 23S ribosomal RNA.

To further investigate potential RNA species from human cells that bind to FTO, full-length recombinant FTO that had been purified using nickel affinity, heparin affinity, and MonoQ ion exchange chromatographies, was incubated with HEK293T cell lysates for 3 h on ice. The mixture was then centrifuged and the hFTO protein from the soluble fraction was repurified using nickel affinity chromatography and heat denatured to release bound nucleic acids. The co-

purified nucleic acids were reverse transcribed into DNA, amplified by PCR, and sequenced. The sequencing results revealed fragments of sequences corresponding to several RNA including 45S pre-ribosomal RNA, mitochondrial ribosomal RNA, small nuclear RNA 7SK, RNA U4, and ATXN2L mRNA transcript (Appendix Table A - 2).

Co-purification of hFTO with RNA U4, which is a small nuclear RNA (snRNA) that is part of the spliceosome complex (25), suggests that FTO may play a role in pre-mRNA splicing. These results support the findings of Jia *et al.* (2011) showing that FTO partially colocalises with splicing factor SC35 (serine/arginine-rich splicing factor 2) and splicing-related speckle factors SART1 (U4/U6.U5 tri-snRNP-associated protein 1) (8). Based on the pull-down results, which showed that ribosomal RNA and 7SK snRNA were co-purified with FTO, FTO may also be involved in transcription and processing of pre-ribosomal RNA. It is notable that 7SK snRNA plays a role in transcription elongation as well as in alternative splicing (26).

To date, most evidence points towards FTO as a RNA demethylase and results from this study showed that it interacts with several types of RNA. These RNAs are part of the splicing or transcription machineries, some of which may be co-purified due to non-specific binding. Further investigations to determine the native substrate of FTO in cells will need to be carried out in order to explain the connection between the RNA m⁶A demethylase activity of FTO and the associated obesity phenotype.

3.2.7 Initial Structural Analysis of FTO

After one week, one of the crystals, grown using hFTO Δ 31 purified with heparin chromatography, had grown to $\sim 100\ \mu\text{m}$ in the longest dimension (Figure 3 - 13a). This crystal was harvested and a complete dataset was collected in-house using a Rigaku FR-E+ Superbright rotating anode X-ray source. The space group was found to be *R*3 and the unit cell constants were similar to the structure deposited in the PDB (3LFM; resolution: 2.50 Å) (12). The crystal structure was solved to a resolution of 3.3 Å by molecular replacement using the first published FTO structure (PDB ID 3LFM; resolution: 2.50 Å) as a search model. An NOG molecule was observed in the 2OG binding site while there was some residual electron density in the substrate binding site (Figure 3 - 12), suggesting that the substrate *N*⁶-methyladenosine (6meA) might be present. However, the electron density was not clear enough to confirm the presence of 6meA. Hence, a higher resolution structure was desirable.

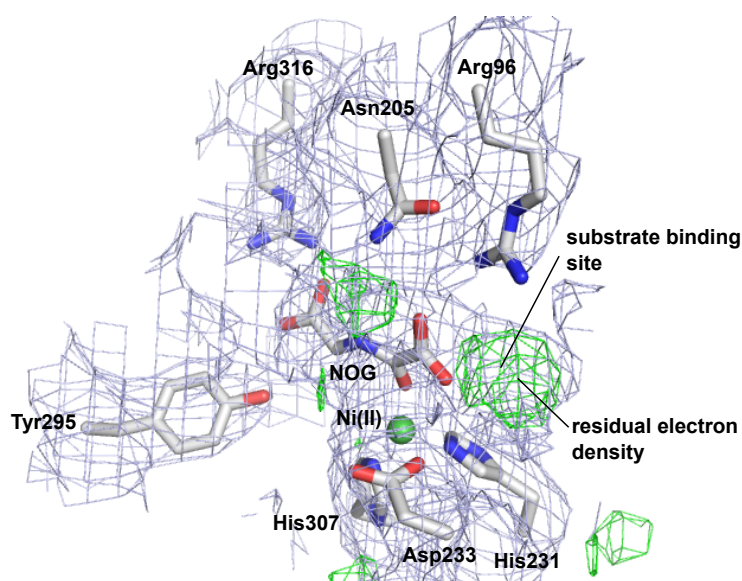


Figure 3 - 12. View of the active site of a low resolution (3.3 Å) crystal structure of FTO grown in the presence of NOG and 6meA. Electron density maps: $2mF_o-DF_c$, $\sigma = 1.0$, light blue mesh; mF_o-DF_c , $\sigma = 3.0$, green mesh. Note the residual electron density in the substrate binding site. The structure is not fully refined. Protein structure models were produced using MacPyMOL v1.3 (13).

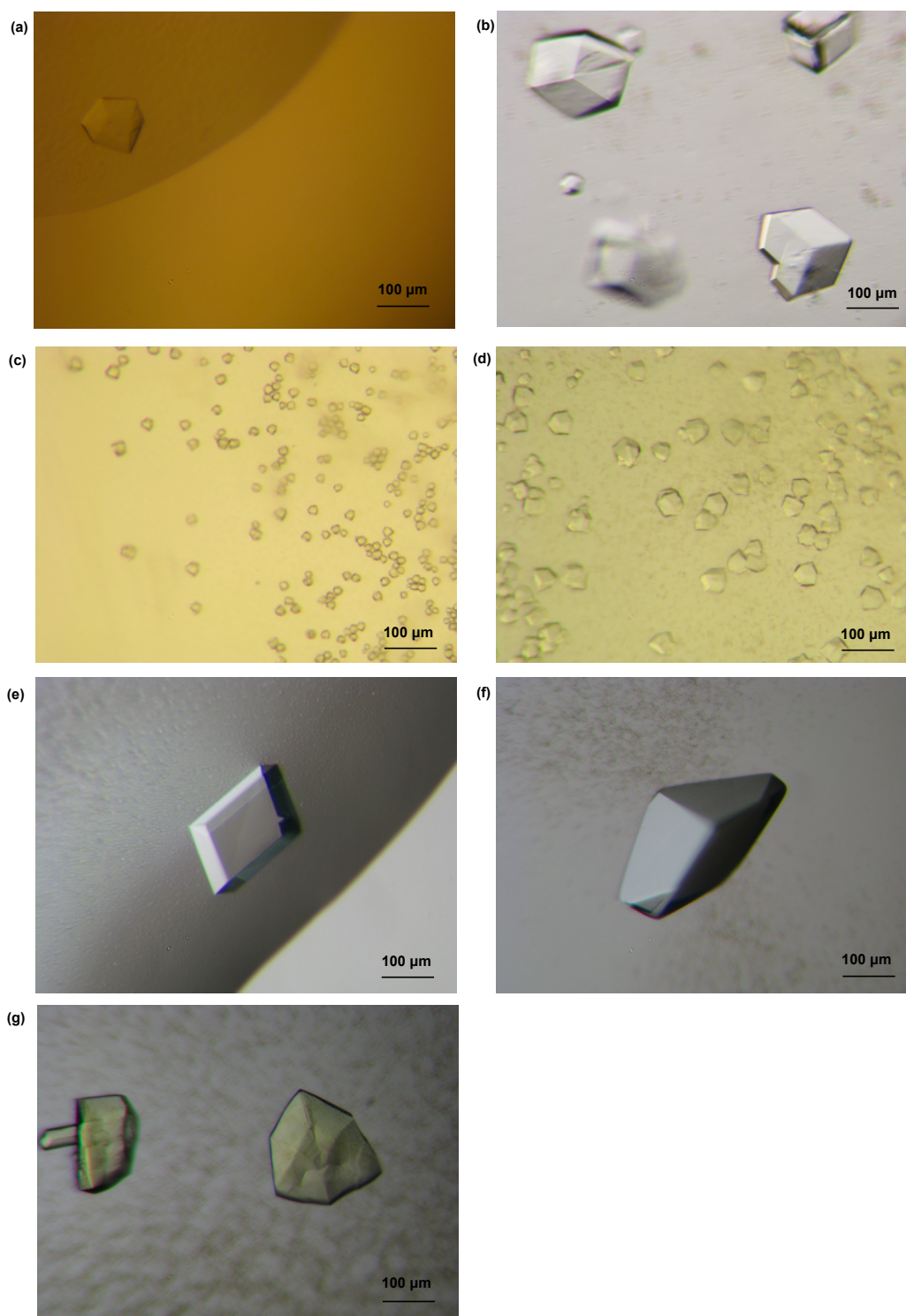


Figure 3 - 13. Images of hFTOΔ31 crystals grown in the presence of different inhibitors. (a) Initial crystal grown in the presence of Ni(II) and NOG that diffracted to 3.3 Å (in-house), **(b)** optimised crystals grown in the presence of Ni(II) and NOG, **(c)** initial crystals grown in the presence of Ni(II) and 2,4-PDCA, **(d)** crystals grown after seeding in the presence of Ni(II) and 2,4-PDCA, **(e)** crystal grown in the presence of Zn(II) and compound **6**, **(f)** crystal grown in the presence of Zn(II) and 2,4-PDCA, and **(g)** crystals grown in the presence of Zn(II) and compound **19**.

A new batch of hFTO Δ 31 was then purified using nickel affinity chromatography, followed by heparin chromatography in order to optimise the initial FTO crystal conditions to obtain higher resolution structures. Crystallisation trials were repeated with previous conditions that yielded crystals. However, only small crystals (10 μ m in the longest dimension) were observed. These crystals did not grow to a usable size for diffraction. Therefore, optimisation of crystal growth was needed, which led to an attempt to use various additives and detergents.

3.2.8 Additive and Detergent Screens

Under certain conditions, additives or detergents added to crystallisation drops can assist in crystal growth (27). Thus, additives and detergent screens (Hampton Research) were used with the existing crystallisation conditions in an attempt to optimise hFTO Δ 31 crystal growth. A number of additives including trifluoroethanol (TFE), isopropanol and ethanol helped FTO crystals to grow to larger (~50 μ m in the longest dimension). Crystals in drops containing 4% TFE and 3% isopropanol gave the best morphology. These crystals were obtained from sitting nanodrops (dropsize of 0.2 – 0.3 μ L) set up using an Art Robbins Phoenix Rigaku Edition crystallisation automation robot. For further optimisation, manual hanging drops on plastic coverslips were set up (dropsize of 3 – 6 μ L) using 1 – 4% TFE and isopropanol, and crystals of 100 μ m in the longest dimension were obtained. However, these crystals only diffracted to about 4 Å resolution in-house, which was still insufficient for ligand assignment.

Given the unsuccessful attempts to optimise crystal growth using additives, attention was turned back to protein purification. Earlier attempts at purifying hFTO Δ 31 described in Section 3.2.6 in this chapter suggested that a combination of heparin affinity and MonoQ anion exchange for protein purification might be necessary to produce usefully crystallisable hFTO Δ 31. As a result, a new batch of hFTO Δ 31 was purified sequentially using nickel affinity, heparin affinity, and MonoQ anion exchange chromatographies (Figure 3 - 14). In this attempt, the MonoQ UV-chromatogram revealed a sharp peak followed by two overlapping broad peaks (Figure 3 - 14c). Fractions corresponding to all three peaks were collected separately and concentrated for crystallisation. Only hFTO Δ 31 from the sharp peak crystallised (crystal size approximately 50 μ m in the longest dimension) in the published conditions, without the presence of TFE and isopropanol as additives, suggesting that the combination of heparin and MonoQ chromatographies was key to producing hFTO Δ 31 protein that was able to crystallise without the use of additives.

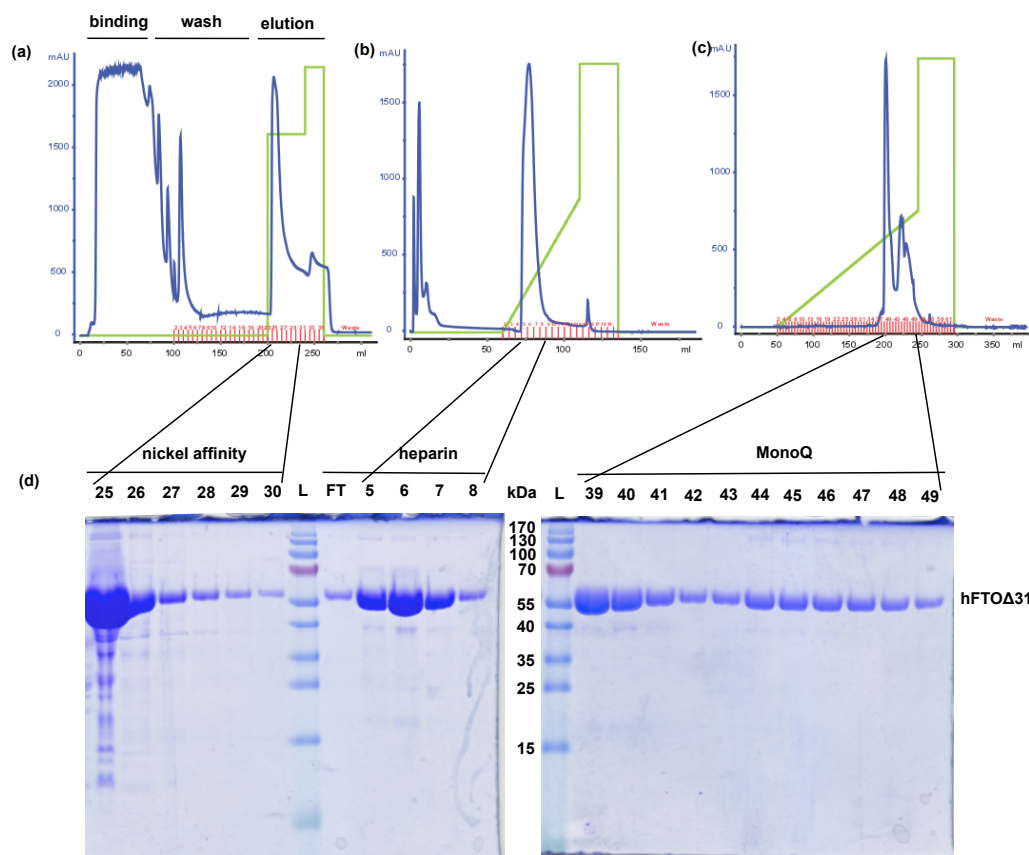


Figure 3 - 14. Attempt 5 at purifying hFTOΔ31. AKTA FPLC UV trace (blue) of **(a)** nickel affinity chromatography, **(b)** heparin affinity chromatography and **(c)** MonoQ anion exchange chromatography. In **(a)**, the green line indicates elution buffer concentration; in **(b)** and **(c)**, the green lines indicate NaCl concentration; the red lines and numbers indicate fractions. **(d)** SDS-PAGE analysis of fractions from nickel affinity, heparin, and MonoQ anion exchange chromatographies. Protein molecular weight standards (kDa) ladder is labelled L, and is followed by fraction numbers.

Although a method to purify the hFTOΔ31 protein that was crystallisable had been developed, initial hFTOΔ31 crystals did not grow beyond 100 μm (in the longest dimension). To optimise crystal growth, manual hanging drops, previously shown to give bigger crystals than sitting nanodrops, were set up to crystallise hFTOΔ31 in the presence of Ni(II), NOG and 6meA nucleoside. Poorly formed crystals that were often small and clustered together were observed, most likely due to a high equilibration and nucleation rate. To reduce the nucleation and equilibration rate, the well solution volume was reduced from 500 μL to 200 μL and 5- 25 % glycerol was added to the well solution (85 mM

trisodium citrate pH 5.6, 14% PEG 3350). Under these conditions, crystals of approximately 250 – 300 μm (in the longest dimension) were obtained, which diffracted to 2.4 Å in-house. The structure of FTO in complex with NOG was solved by molecular replacement using a structure of the FTO-NOG-3meT complex (PDB ID 3LFM; resolution: 2.50 Å) (12) as a search model. (see Section 3.2.13 – 16 for data collection, structure solution, refinement and structure description).

3.2.9 Co-crystallisation of FTO with 2,4-Pyridine Dicarboxylate

2,4-Pyridine dicarboxylate (compound **2**) is another broad spectrum inhibitor of 2OG oxygenases (19) that was found to inhibit FTO ($\text{IC}_{50} = 8.3 \mu\text{M}$, see Section 3.2.5). Co-crystallisation of hFTO Δ 31 was carried out in the presence of compound **2** and Ni(II) using the same conditions as those employed for the FTO.Ni.NOG complex. However, no crystal formation was observed.

New conditions are sometimes required to grow crystals for protein-ligand complexes with different ligands (28). Thus, crystallisation trials of hFTO Δ 31 were performed using a NATRIX screen (Hampton Research) in the presence of Ni(II) and **2**. Crystals formed in the presence of **2** in the well containing 200 mM sodium cacodylate pH 6.0 and 20% 2-methyl-2,4-pentanediol (MPD). To co-crystallise hFTO Δ 31 with Ni and **2**, optimisation was carried out using manual hanging drops in 100 – 300 mM sodium cacodylate pH 6.5 and 10 – 35% MPD. Small crystals appeared within minutes after the drops were set up, a clear sign of overnucleation was evident. Further optimisation steps were carried out, including a reduction in the MPD concentration, addition

of glycerol and varying the pH of the sodium cacodylate buffer (i.e. pH 5.0, pH 5.6, pH 6.0) but there was no improvement in crystal growth.

Due to the rapid formation of hFTO Δ 31 crystals (within 10 min), repeated sequential seeding of fast growing crystals was attempted. Crystals grown in the initial crystallisation drop were sequentially transferred into new drops for 3 – 4 cycles in the hope that the crystal seeded would grow to sufficient size for diffraction. This method resulted in an improvement of approximately 7-fold in crystal size from 10 μ m to about 75 μ m in the longest dimension. However, the larger FTO crystals of 75 μ m (in the longest dimension) did not diffract beyond 4 Å resolution.

3.2.10 Soaking FTO Crystals Grown in NOG with 2,4-PDCA and FG-2216

Since crystals grown in the presence of **2** did not diffract beyond 4 Å resolution, hFTO Δ 31 crystals grown to a size of ~300 μ m (longest dimension) in the presence of Ni(II) and NOG were soaked with **2** and IOX3 (compound **11**) respectively (5 and 10 mM, respectively), for 1 h. These crystals diffracted to about 2.5 Å resolution at the synchrotron, and the structures were solved by molecular replacement using a FTO-NOG-3meT structure (PDB ID 3LFM; resolution: 2.50 Å) as a search model (12). Both datasets revealed NOG and glycerol in the active site, indicating that soaking crystals with **2** and **11** at these concentrations (5 and 10 mM) of **11** does not displace NOG from the 2OG binding pocket under these conditions.

The unsuccessful efforts to co-crystallise FTO with **2** and **11** could be due to Ni(II) not being the ideal metal as an Fe(II) substitute. Although Fe(II) can be

easily oxidised to Fe(III) under aerobic crystallisation conditions, using Fe(II) for co-crystallisation has previously been a successful strategy for AlkB (Chapter 2) (11). Therefore, FTO co-crystallisation trials were performed in the presence of Fe(II) and either **2** or **11**. Well-formed rhombohedral FTO crystals were observed in conditions containing 100 mM trisodium citrate pH 5.6, 14% PEG 3350 and with 4% tert-butanol as an additive. As a result, hanging drops were set up manually for optimisation of crystal growth of FTO in the presence of Fe(II) and **11**. Crystals of 300-400 μm (longest dimension) were formed and diffracted to 2.0 Å in-house. The structure was solved by molecular replacement but neither Fe(II) nor **11** were present (Figure 3 - 15a). Instead, a citrate molecule from the crystallisation buffer was observed in the active site.

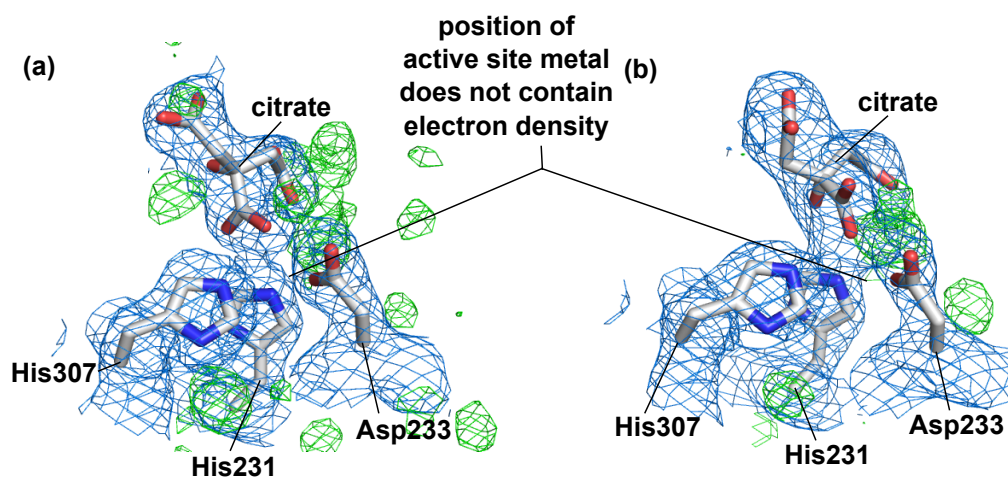


Figure 3 - 15. Electron density maps of the active sites of structures of FTO without metal binding. (a) Structure (resolution = 2.0 Å) determined from a hFTOΔ31 crystal grown in the presence of Fe(II) and compound **11**. (b) Structure (resolution = 3.0 Å) determined from a hFTOΔ31 crystal grown in the presence of Mn(II) and compound **11**. Note that the position where active site metal is normally found does not contain electron density that corresponds to Fe(II) (for (a)) and Mn(II) (for (b)). A citrate molecule can be observed in the active site of both (a) and (b). Electron density map: $2mF_o - DF_c$, $\sigma = 1.0$, blue mesh; $mF_o - DF_c$, $\sigma = 3.0$, green mesh. Both structures were not fully refined. Protein structure models were produced using MacPyMOL v1.3 (13).

3.2.11 Active site metal substitutes

The inhibitors chosen (compounds **1**, **2**, **3**, **6** and **11**; Table 3 - 1) for structural investigation were either competitors of 2OG or metal chelators. Therefore, it was deemed imperative to have a metal in the active site in order to achieve stable inhibitor binding. Because Fe(II) was not observed in the crystal structure described above (Figure 3 - 15a), FTO co-crystallisation trials were carried out in the presence of either 0.5 mM Mn(II) or Zn(II), together with 1 mM inhibitor (**2** or **11**). A metal : hFTOΔ31 molar ratio of almost 2 : 1 was used because higher metal : hFTOΔ31 ratio might lead to non-specific binding of the metal to the protein at sites other than the active site. X-ray diffraction data were collected for crystals from drops containing Mn(II) or Zn(II). The electron density maps (resolution 3.0 Å) revealed that Mn(II) was not present in the active site (Figure 3 - 15b). In contrast, Zn(II) was observed, suggesting that FTO has a higher affinity for Zn(II) under these conditions. To enhance the occupancy of Zn(II) in the active site, 1mM Zn(II) was used in subsequent crystallisation trials.

3.2.12 Inhibitor Concentration Optimisation for Crystallographic Studies

The data collected thus far showed active sites either without metal or with both Zn(II) and a citrate molecule. In an attempt to increase the inhibitor occupancy, the concentration of each inhibitor was increased incrementally in the hope that higher concentration of inhibitor would outcompete citrate for binding at the active site. However, too high inhibitor concentration can result in poor crystal formation. After several trials, the optimal concentration of **2** was found to be 1 – 2 mM. This condition allowed cocrystallisation of FTO with **2**, resulting in

crystals of up to 400 μm (longest dimension). One of the crystals diffracted to 2.5 Å resolution and resulted in the FTO-2 crystal structure complex (see Section 3.2.17).

The optimised crystallisation conditions for co-crystallisation of hFTOΔ31 with **2** subsequently allowed co-crystallisation of hFTOΔ31 with other inhibitors including compounds **3**, **6**, **11** and **19** (see Section 3.2.30-35, Table 3 - 1).

3.2.13 Crystallographic Data Collection

The space group of the hFTOΔ31 crystals was found to be $R3$. The unit cell constants are $a = \sim 140$ Å, $b = \sim 140$ Å, $c = \sim 83$ Å, $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$ (see Table 3 - 2 and Table 3 - 3 for details on data collection and refinement statistics). The calculated Matthews Coefficient ($V_m = 2.8$ Å³), with 55% solvent content, predicted one molecule per asymmetric unit related by a three-fold symmetry within a unit cell (29-31). A typical data collection strategy using the home source Rigaku FR-E+ Superbright X-ray diffractometer (Cu K_α 1.5401 Å) was 0.5° rotation per image for 60 s for a total rotation of 90°, at 100 K. In some cases, due to the high background in the low resolution bins, potentially due to a large amount of solvent surrounding the crystal, an additional 90° of data collection at low X-ray exposure (5 – 10 s per image) was carried out to achieve near 100% completeness.

3.2.14 Data Processing and Structure Solution

Data were integrated and scaled using HKL2000 (32). Error models were adjusted to obtain a χ^2 near to 1 for all resolution bins. The highest resolution bin

was usually cut off at $I/\sigma(I)$ of at least 2.0. Molecular replacement by Phaser (33) was used to solve the structures of hFTO Δ 31 in complex with various inhibitors using the first published FTO structure (PDB ID: 3LFM; resolution: 2.50 Å) (12) as the search model. All searches resulted in a Translation Function Z Score (TFZ) greater than 8.0 and a positive log-likelihood gain (LLG), indicating correct solutions.

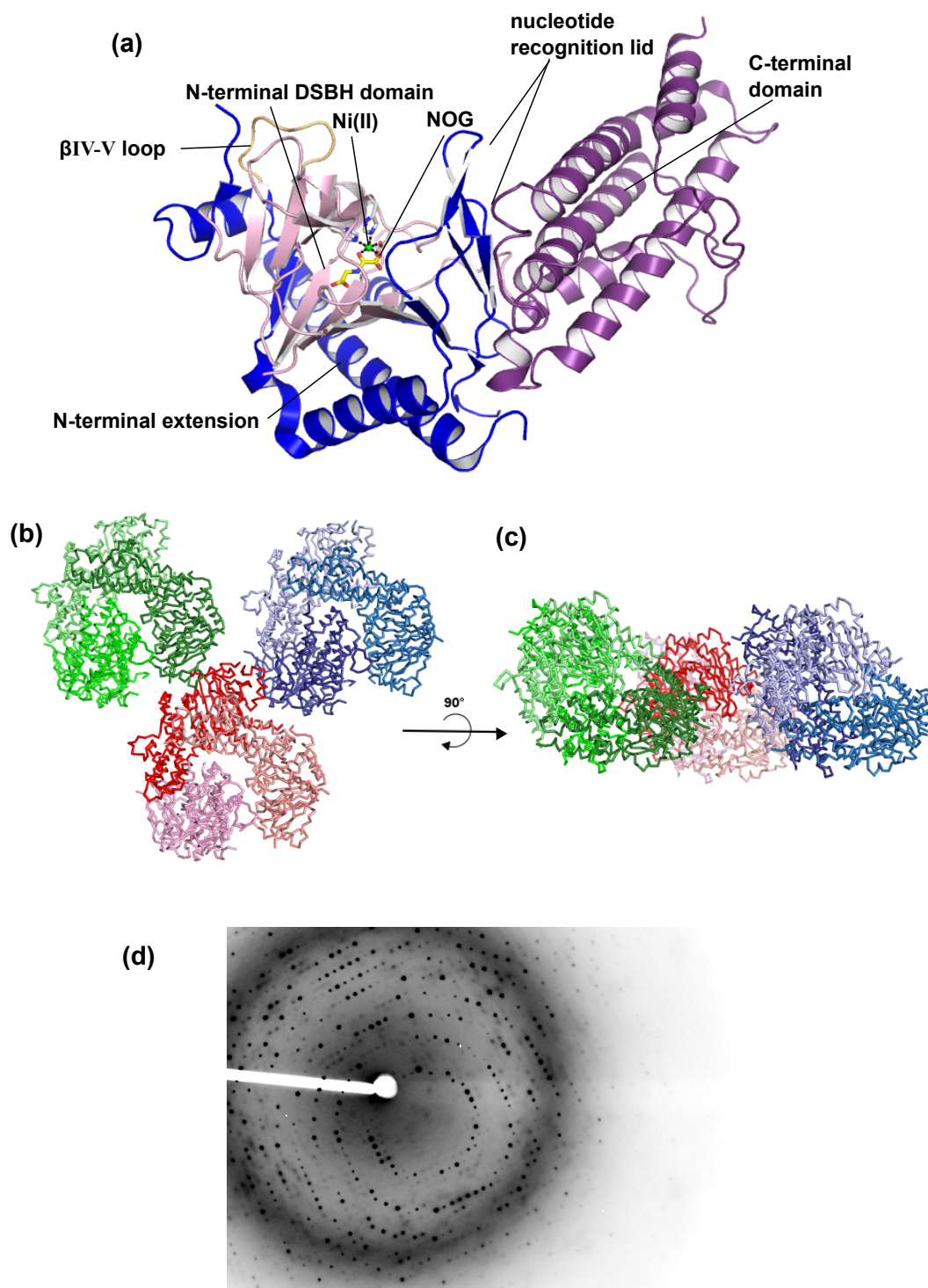


Figure 3 - 16. Crystallographic studies on FTO. (a) Overall view of a structure of the FTO-NOG complex. (b) Crystal packing of FTO in two orientations rotated at 90° relative to each other. There is one molecule per asymmetric unit. (d) A representative diffraction pattern of FTO crystals. Protein structure models were produced using MacPyMOL v1.3 (13).

3.2.15 Structure Refinement

Phenix.Refine and Coot were used to perform iterative model fitting and refinement until the R and R_{free} values no longer converged (34,35). Initial rounds of refinement parameters included XYZ refinement, individual B factor refinement, X-ray/stereochemistry and X-ray/ADP weights refinement, and secondary structure restraints. Translation, libration, screw (TLS) refinement was added as a parameter in later rounds of refinement.

Table 3 - 2. FTO Crystal Structures Data Collection and Refinement Statistics.

	FTO-1	FTO-2	FTO-3	FTO-6	FTO-11	FTO-19
PDB ID	4IDZ	4IE0	4IE4	4IE5	4IE6	4IE7
Crystal Size (Longest Dimension)	~300 μm	~300 μm	~250 μm	~300 μm	~200 μm	~200 μm
Data Collection						
Radiation Source	Rotating anode	Synchrotron Diamond Beamline I03	Rotating anode	Rotating anode	Synchrotron Diamond Beamline I03	Synchrotron Diamond Beamline I03
Detector	Saturn 944+ CCD	Pilatus 6M-F	Saturn 944+ CCD	Saturn 944+ CCD	Pilatus 6M-F	Pilatus 6M-F
X-ray Wavelength ($\text{\AA}$)	1.5412	0.9763	1.5412	1.5412	0.9763	0.9763
Temperature (K)	100	100	100	100	100	100
Resolution Range ($\text{\AA}$) (‡)	50.00-2.46 (2.55-2.46)	50.00-2.53 (2.62-2.53)	50.00-2.50 (2.59-2.50)	50.00-1.96 (2.03-1.96)	50.00-2.50 (2.59-2.50)	50.00-2.60 (2.69-2.60)
Space Group	<i>R</i> 3	<i>R</i> 3	<i>R</i> 3	<i>R</i> 3	<i>R</i> 3	<i>R</i> 3
Unit Cell Dimensions (<i>a</i> $\text{\AA}$, <i>b</i> $\text{\AA}$, <i>c</i> $\text{\AA}$, α° , β° , γ°)	140.41, 140.41, 83.45, 90°, 90°, 120°	142.17, 142.17, 83.71, 90°, 90°, 120°	141.63, 141.63, 84.59, 90°, 90°, 120°	141.39, 141.39, 84.18, 90°, 90°, 120°	141.46, 141.46, 83.81, 90°, 90°, 120°	141.62, 141.62, 85.60, 90°, 90°, 120°
Total Number of Reflections Observed	113209	99502	107229	274777	102807	96516
Number of Unique Reflections	22286	21499	21717	45555	21417	19675
Multiplicity (‡)	5.1 (5.1)	4.6 (4.6)	4.9 (5.5)	6.0 (6.4)	4.8 (4.4)	4.9 (4.8)
Completeness (%) (‡)	98.6 (99.9)	99.6 (100.0)	99.8 (100.0)	99.8 (100.0)	100.0 (100.0)	99.9 (99.9)
<i>I</i> / σ (<i>I</i>) (‡)	15.6 (2.2)	9.0 (2.1)	6.6 (2.1)	14.5 (2.1)	16.2 (3.52)	12.7 (4.4)
R_{merge} (%) (‡)	8.0 (69.3)	12.3 (63.2)	17.7 (91.9)	9.7 (ND)	7.3 (45.5)	9.8 (35.5)
Wilson B Factor ($\text{\AA}^2$)	51.11	80.46	52.28	36.10	71.00	74.65
Software for data reduction	HKL-2000	HKL-2000	HKL-2000	HKL-2000	HKL-2000	HKL-2000
Refinement						
R_{work} (%)	19.36	17.61	20.55	18.30	16.95	17.37
R_{free} (%)	24.47	21.29	24.25	20.31	21.04	21.80
σ RMS Deviation	0.009 $\text{\AA}$ (1.178°)	0.011 $\text{\AA}$ (1.184°)	0.010 $\text{\AA}$ (1.105°)	0.010 $\text{\AA}$ (1.171°)	0.010 $\text{\AA}$ (1.203°)	0.012 (1.250°)
Ramachandran Outliers (%)	0.2	0	0	0	0.2	0.2
σ Average B Factor ($\text{\AA}^2$)	60.21	83.50	47.87	46.27	70.76	79.17
Number of Water Molecules	92	21	71	324	21	21
Software for refinement	PHENIX	PHENIX	PHENIX	PHENIX	PHENIX	PHENIX

(‡) Numbers in brackets indicate outermost shell

$$\S R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle \times 100$$

$$* R_{\text{work}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}| \times 100$$

† R_{free} , based on 10% of the total reflections, except for FTO-6 that is 5% of total reflections

¶ RMS deviation from ideality for bonds (followed by the values for angles)

3.2.16 A Crystal Structure of hFTOΔ31 in Complex with NOG

The structure of FTO determined for a crystal grown in the presence of NOG (compound **1**) and 6meA (2.4 Å resolution) is similar to the previously reported FTO-1-3meT structure (12) (root mean square deviation, rmsd = 0.536 for 367 C_{α} atoms), consisting of an N-terminal DSBH domain and a C-terminal five helical bundle domain (Figure 3 - 16a). The active site of the determined FTO structure contains the conserved 2OG oxygenase Fe(II) coordinating residues His231, Asp233 and His307 (Figure 3 - 17a). In this structure, Ni(II) as an inactive Fe(II) substitute is coordinated by the facial triad. Electron density is observed in the 2OG binding site, apparently correlating with a species that chelates the active site metal (Figure 3 - 17a).

There is another piece of continuous density in the nucleotide binding site. Compound **1** was assigned to occupy the density in the 2OG binding site with its C5-carboxylate positioned to form hydrogen bonds with the side chain of Arg316. The C1-carboxylate of **1** is positioned *trans* to the His307 imidazole and the amide carbonyl is *trans* to the carboxylate side chain of Asp233. Refinement led to assignment of a glycerol molecule in the prime substrate binding site, instead of a 6meA nucleoside. This observation suggests that 6meA may also be absent in the lower resolution structure (3.3 Å) that was determined earlier (see Section 3.2.7).

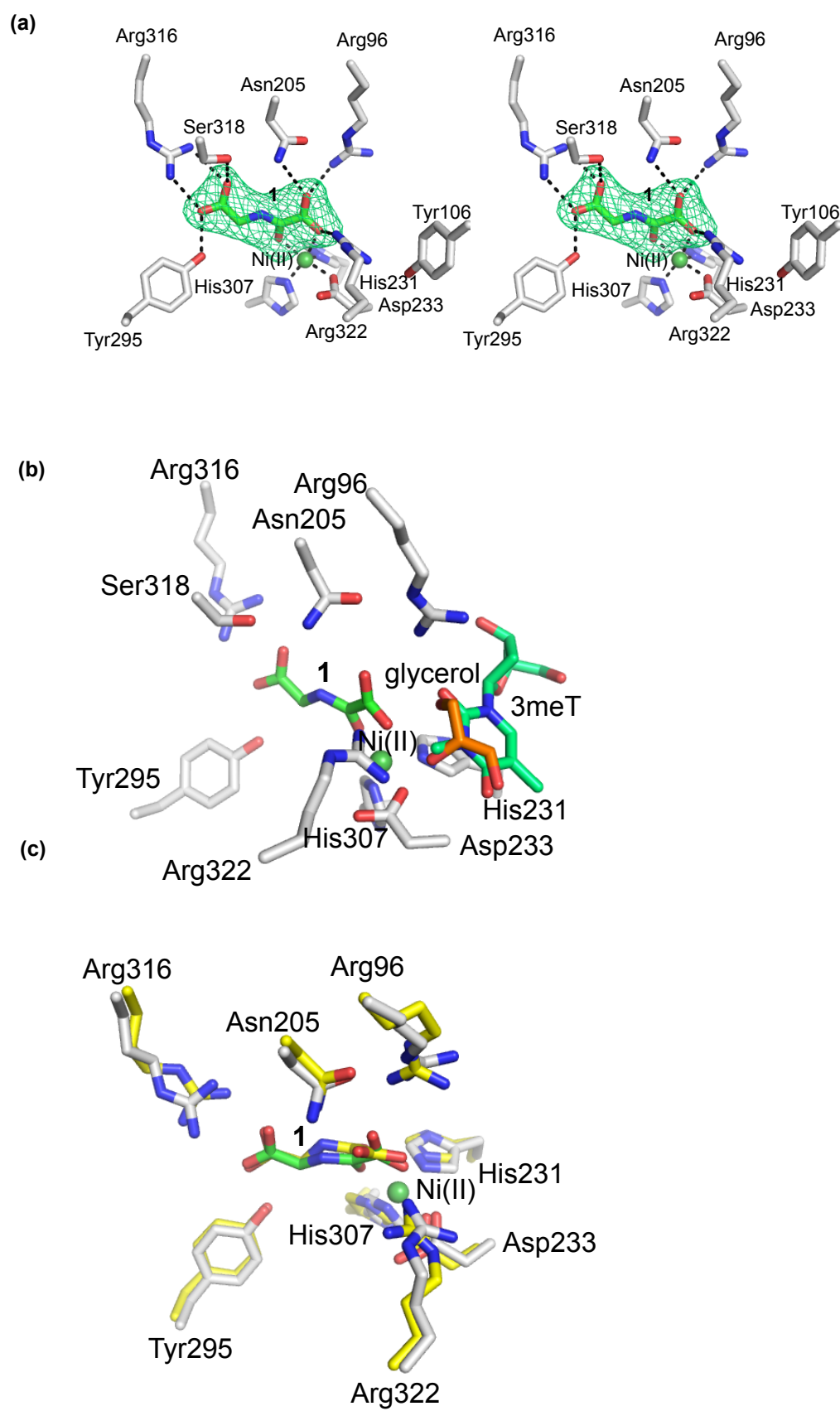


Figure 3 - 17. Views of a crystal structure of FTO in complex with NOG. (a) Wall-eyed stereo views of the active site from a crystal structure of FTO in complex with **1** (green) (resolution: 2.46 Å). (b) Superimposition of 3meT (lime green) from a FTO-**1**-3meT structure (PDB ID: 3LFM; resolution: 2.50 Å) (12) on a structure of the FTO-**1** complex (PDB ID: 4IDZ; resolution: 2.46 Å), glycerol (orange) occupies the 3meT binding site. (c) Superimposition of the active site residues and **1** of the FTO-**1**-3meT structure (PDB ID 3LFM; resolution: 2.50 Å) (yellow) previously determined of the FTO-**1** structure from this study (PDB ID 4IDZ; resolution: 2.46 Å) (residues, white sticks; compound **1**, green sticks). FTO residue side chains C (white), O (red), N (blue), Ni(II) (green sphere), and electrostatic interactions are indicated as black dashes. mF_o-DF_c ligand omit maps (contour level, $\sigma = 3.0$) are indicated as lime green mesh. Figure and figure legend are reprinted (adapted) with permission from Aik *et al.* (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, 56, 3680-3688. Copyright 2013 American Chemical Society. Protein structure models were produced using MacPyMOL v1.3 (13).

Attempts to co-crystallise hFTO Δ 31 with 6meA nucleoside substrate and to soak FTO crystals grown in the presence of **1** with 6meA were unsuccessful in producing a structure of the FTO-6meA complex. A comparison of compound **1** of the FTO-**1** complex structure with that of a previously determined FTO-**1**-3meT structure (PDB ID 3LFM) (12) revealed that **1** in the structure determined in this study appears more planar than that of 3LFM. The side chains of Arg96, Arg316 and Arg322 of 3LFM adopt different rotamer conformations (Figure 3 - 17c) compared to their positions in the determined 4IDZ structure. Overall, this new model has lower R_{work} (0.194) and R_{free} (0.245) values, and a higher percentage of residues in Ramachandran favoured regions (98%), compared to the reported FTO-**1**-3meT structure ($R_{work} = 0.239$; $R_{free} = 0.285$; Ramachandran favoured regions = 91%).

3.2.17 Crystal Structure of FTO in Complex with 2,4-PDCA

A crystal structure of FTO in complex with 2,4-PDCA (**2**) was determined to 2.5 Å resolution. This structure reveals that **2** binds in the 2OG binding pocket chelating the active site metal (Zn(II) in this case) in a bidentate manner (Figure 3 - 18). The C4-carboxylate is positioned to form electrostatic and hydrogen bonding interactions with side chains of Arg316, Ser318, and Tyr295. The 2-

carboxylate is positioned *trans* to the His307 imidazole and the pyridyl nitrogen of **2** is *trans* to the Asp233 carboxylate. Compound **2** also binds in a similar manner in several other 2OG oxygenases such as hypoxia inducible factor (HIF) asparaginyl hydroxylase (FIH), histone demethylase Jumonji domain-containing protein 2A (JMJD2A) and *Chlamydomonas rheinhardtii* collagen prolyl hydroxylase (CrCPH) (36-38).

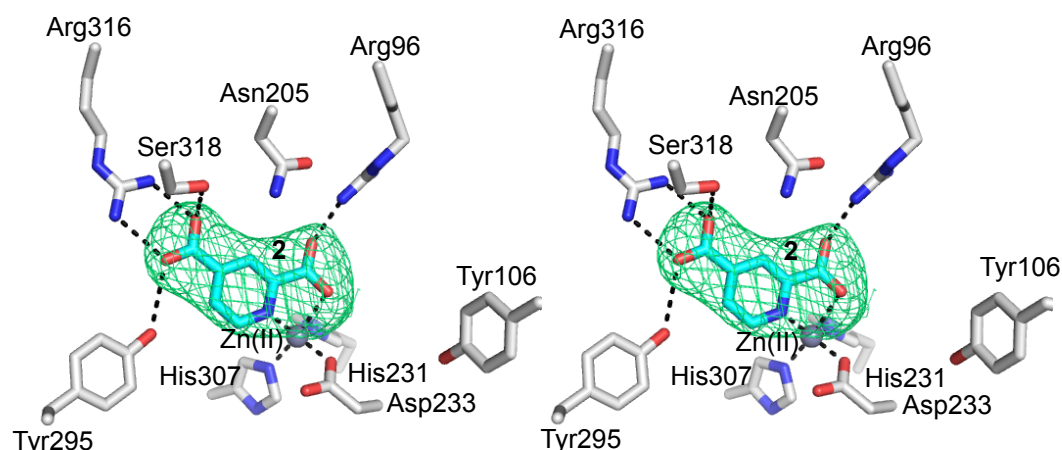


Figure 3 - 18. Wall-eyed stereo views of the active site of a crystal structure of FTO in complex with **2 (cyan) (resolution: 2.53 Å).** FTO residue side chains C (white), O (red), N (blue), Zn(II) (grey sphere), and electrostatic interactions are indicated as black dashes. mF_o-DF_c ligand omit maps (contour level, $\sigma = 3.0$) are indicated as lime green mesh. Figure and figure legend are reprinted (adapted) with permission from Aik, W. *et al* (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J. Med. Chem.*, 56, 3680-3688. Copyright 2013 American Chemical Society. Protein structure models were produced using MacPyMOL v1.3 (13).

3.2.18 Crystal Structure of FTO in Complex with 5-Carboxy-8-hydroxyquinoline

5-Carboxy-8-hydroxyquinoline (**3**) is a broad spectrum 2OG oxygenase inhibitor (19). Several 8-hydroxyquinoline derivatives (compounds **3**, **4** and **5**) have been shown to inhibit FTO (Section 3.2.4, Table 3 - 1) but their binding mode in FTO was unknown. A crystal structure of the FTO-**3** complex was determined to 2.5 Å resolution (Figure 3 - 19a). The structure reveals that the 5-carboxylate of **3** is in

a position similar to that observed for the C5-carboxylate of NOG and 4-carboxylate of **2**, interacting with the side chain of Arg316 through hydrogen bonds. **3** chelates Zn(II) in a bidentate manner with the nitrogen from the quinoline rings (*trans* to Asp233) and oxygen from the 8-hydroxyl group (*trans* to His307). Interestingly, this binding mode is different from the binding mode of **3** observed in the co-crystal structures with JMJD2A (39) and FIH (19). In these structures, the positions of the metal chelating oxygen and nitrogen were swapped relative to FTO, leading to an inverse orientation of **3**. The JMJD2A-**3** (Figure 3 - 19c) and FIH-**3** complex structures also show a shift in the metal position (~ 1.5 Å from its usual position) upon compound **3** binding (see Chapter 2, Section 2.2.2) (19,39), but such a difference in metal position was not observed for FTO. A similar binding mode without metal translocation was also observed in AlkB (Figure 3 - 19b) (see Chapter 2), suggesting that this alternative binding mode of compound **3** could be specific to the AlkB subfamily of 2OG oxygenases.

The position of the 5-carboxylate of **3** (IC_{50} 3.3 μ M) in the active site of FTO also explains the reduced potencies of two derivatives of **3**, in which the carboxylate is substituted with a sulfonamide (IC_{50} 23 μ M for **4** and 18 μ M for **5**) (Table 3 - 1). Replacing the carboxylate with a sulfonamide might be preventing effective hydrogen bonding and/or electrostatic interactions with Arg316, thus reducing the potencies of **4** and **5** against FTO.

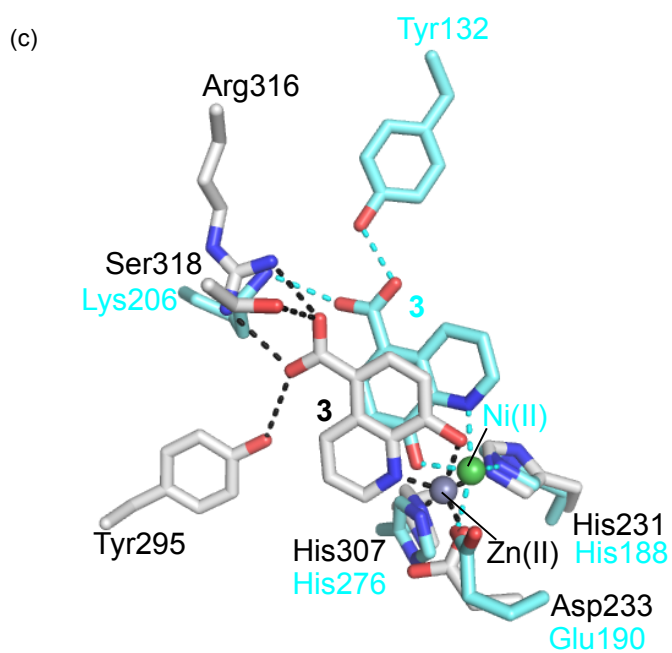
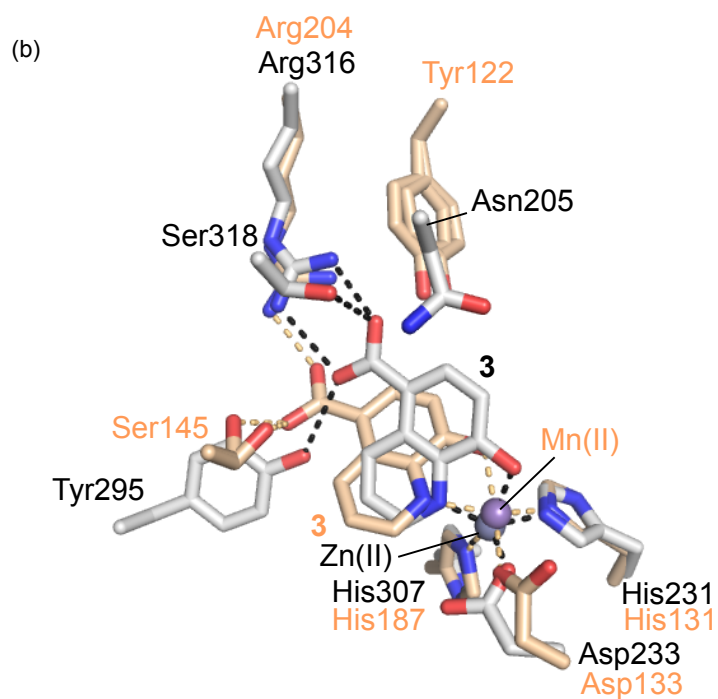
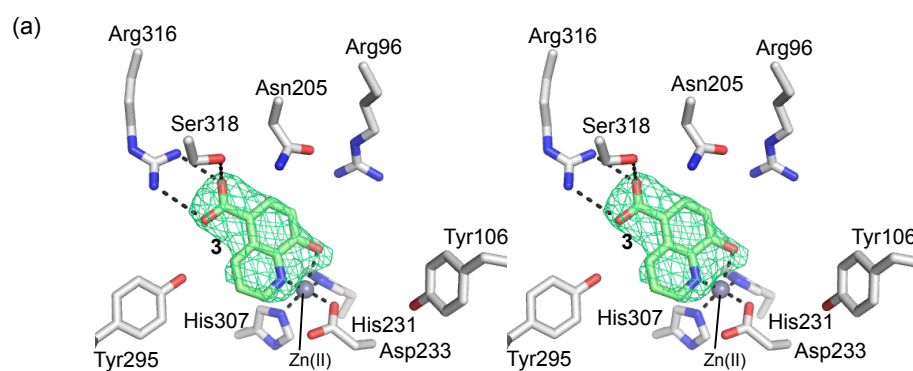


Figure 3 - 19. Views of a crystal structure of the FTO-3 complex. (a) Wall-eyed stereo views of the active site from a crystal structure of FTO in complex with **3** (light green) (resolution: 2.50 Å). **(b)** Superimposition of a structure of the FTO-3 complex (white, labelled in black; resolution: 2.50 Å) with a structure of the AlkB-3 complex (light orange, labelled in orange) (PDB ID 4JHT; resolution: 1.18 Å) (19). **(c)** Superimposition of a structure of the FTO-3 complex (white, labelled in black; resolution: 2.50 Å) with a structure of the JMJD2A-3 complex (light cyan, labelled in cyan) (PDB ID 3NJY; resolution: 2.60 Å) (39). FTO residue side chains C (white), O (red), N(blue), Ni(II) (green sphere), Zn(II) (grey sphere), Mn(II) (purple sphere) and electrostatic interactions of FTO are indicated as black dashes, of AlkB in bright orange dashes, and of JMJD2A in cyan dashes. mF_o-DF_c ligand omit maps (contour level, $\sigma = 3.0$) are indicated as lime green mesh. Figure and figure legend are reprinted (adapted) with permission from Aik, W. et al (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, 56, 3680-3688. Copyright 2013 American Chemical Society. Protein structure models were produced using MacPyMOL v1.3 (13).

3.2.19 Crystal Structure of FTO in Complex with Compound 6

Pyridyl glycinamide compounds have previously been explored as inhibitors of AlkB (11) (see Chapter 2 Section 2.2.5). Compound **6** (see Table 3 - 1) was one of the most potent inhibitors of AlkB (IC₅₀ 3.4 µM) (see Chapter 2 Section 2.2.5). Thus, a crystal structure of AlkB in complex with Fe(II) and compound **6** was determined (see Chapter 2, Section 2.2.7). The AlkB-**6** complex structure reveals that **6** is bound in the 2OG binding site with its pyridine ring positioned in a hydrophobic pocket buried within the core of the double-stranded β helix (DSBH) fold (11), leading to the prediction by Chen *et al.* (2013) that **6** may bind to FTO in a similar manner i.e. without disrupting nucleotide binding (20).

To investigate the binding mode of **6** and pyridyl glycinamides in general, an FTO-**6** structure was determined to 2.0 Å resolution (Figure 3 - 20a). Unexpectedly, the structure showed that **6** binds to FTO differently to that observed in AlkB (Figure 3 - 20c). In both FTO and AlkB, the side chain carboxylate of **6** is positioned to bind in the 2OG binding pocket. However, the hydrophobic pocket of AlkB in which the pyridine ring of **6** binds is larger and more flexible (Trp178 can adopt multiple conformations) and thus, more accommodating than that of FTO. As a result, it appears that the pyridine ring of

6 is positioned closer to the nucleotide substrate binding site in FTO such that the nitrogen of the pyridine ring chelates the active site metal *trans* to His307 (His₂ of His₁XAsp...His₂) whereas for AlkB, the nitrogen of the pyridine ring is positioned *trans* to His-131 (His₁) (Figure 3 - 20c). Superimposition of the FTO-**6** and FTO-**1**-3meT structures shows that C5 of the pyridine ring of **6** is positioned ~1.4 Å from the 3meT carbonyl oxygen (Figure 3 - 20b), suggesting that **6** may sterically hinder binding of FTO's nucleic acid primary substrate.

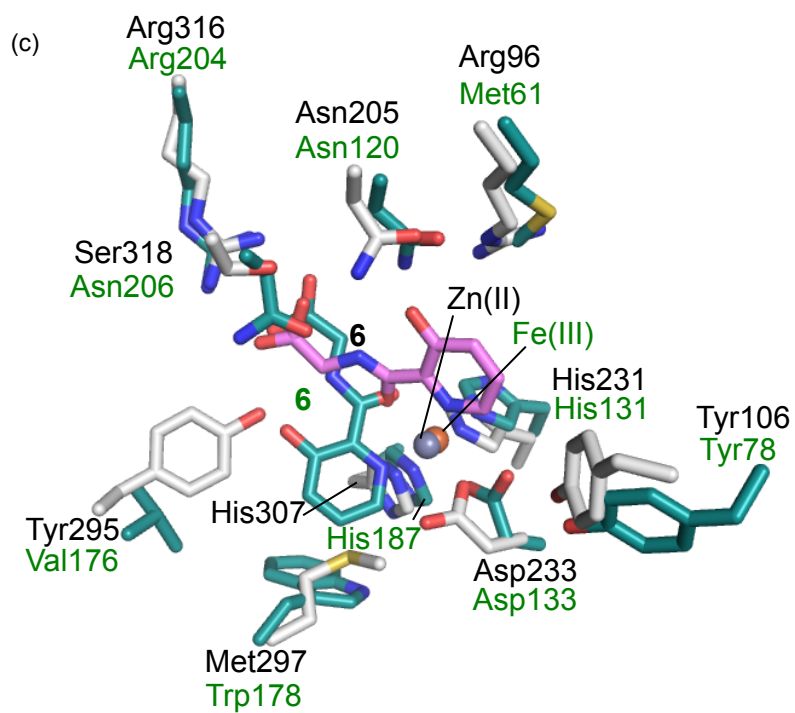
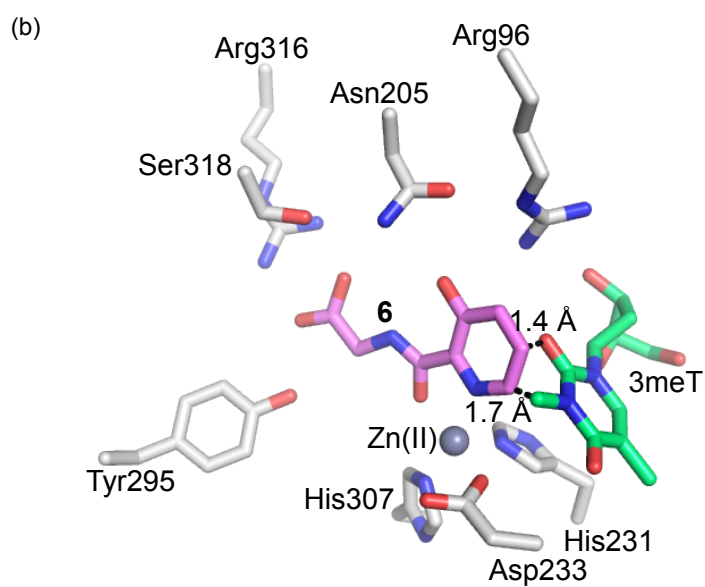
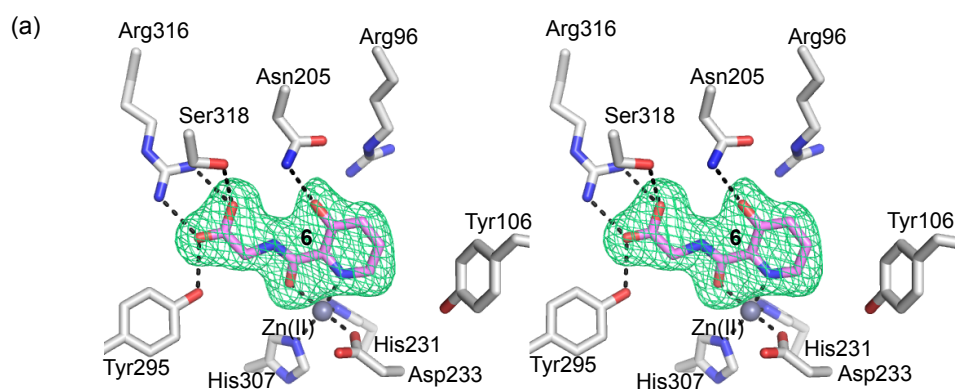


Figure 3 - 20. Views from a crystal structure of the FTO-6 complex. (a) Wall-eyed stereo views of the active site from the crystal structure of FTO in complex with **6** (magenta) (resolution: 1.96 Å). FTO residue side chains C (white), O (red), N (blue), Fe(III) (orange sphere), Zn(II) (grey sphere), and electrostatic interactions are indicated as black dashes. mF_o-DF_c ligand omit maps (contour level, $\sigma = 3.0$) are indicated as lime green mesh. (b) 3meT from the FTO-1·3meT structure (PDB ID: 3LFM; resolution: 2.50 Å) superimposed on the structure of FTO-Zn·**6** (PDB ID: 4IE5; resolution: 1.96 Å); Note that **6** may sterically hinder binding of substrate. (c) Superimposition of FTO-Zn·**6** (PDB ID: 4IE5; resolution: 1.96 Å) with AlkB·Fe·**6** (dark green; labeled in green) (PDB ID: 3T3Y; resolution: 2.00 Å) shows that **6** has different binding modes in FTO and in AlkB. Figure and figure legend are reprinted (adapted) with permission from Aik, W. *et al* (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, 56, 3680-3688. Copyright 2013 American Chemical Society. Protein structure models were produced using MacPyMOL v1.3 (13).

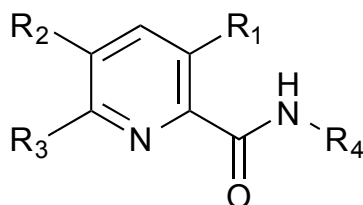


Figure 3 - 21. General chemical structure of pyridyl glycinamide compounds.

Based on the FTO-**6** structure, the lower potencies of the other pyridyl glycinamides (compounds **7** – **10**) (Table 3 - 1) can be attributed to substitutions on the R₂ and R₄ side chains, as shown in Figure 3 - 21, leading to steric clashes: the presence of a bulky group at R₂ (compounds **7**, **9** and **10**) could potentially cause a steric clash with the side chain of Arg96; likewise, substitution with a bulky group at R₄ (compound **8**) could lead to clashes with residues in the hydrophobic pocket such as Val244 and Tyr295. However, substitution at position R₃ of the pyridine ring (adjacent to the nitrogen) may lead to greater occupancy of the substrate binding site, thus improving binding affinity and inhibitory potency (see Section 3.2.22).

3.2.20 Crystal Structure of FTO in Complex with FG-2216

FG-2216/IOX3 (**11**) is a known inhibitor of the HIF-prolyl hydroxylases (PHD1 – 3) that is active both *in vitro* and *in vivo* (18,40-42). A crystal structure of FTO in complex with **11** was determined to 2.5 Å resolution (Figure 3 - 22). The occupancy of **11** was refined to 90%. The glycine side chain of **11** binds in the 2OG binding site and its chlorine is positioned in the nucleic acid substrate binding site. The side chain amide carbonyl deviates from planarity with respect to the isoquinoline ring by ~60°, skewing the isoquinoline ring nitrogen-to-metal coordination. This occurrence could be explained by the position of the side chain of Arg96 with respect to the bicyclic rings, whereby Arg96 sterically forces a non-ideal metal coordination by the compound. In comparison, compound **11** binds similarly in PHD2 as in FTO, but **11** adopts a more planar conformation in PHD2 (40).

Similar to the pyridyl glycine series, all the tested derivatives of **11** (compounds **12** – **18**) did not give lower IC₅₀ values than **11** (Table 3 - 1). This is because most of the substitutions were on the glycine side chain (compound **12** – **17**), which may clash with the side chains of Val244, Tyr295 and Val309 in the hydrophobic pocket. Compound **18** (identified by SelleckBio as FG-4592), a clinical trial drug targeting PHD2 for the treatment of anaemia (43), has a reasonably good potency as an inhibitor of FTO (IC₅₀ 9.8 µM) but the presence of a phenoxy group may be sterically hindered by the side chain of Tyr106, thus making **18** a less potent inhibitor than **11** (IC₅₀ 2.8 µM).

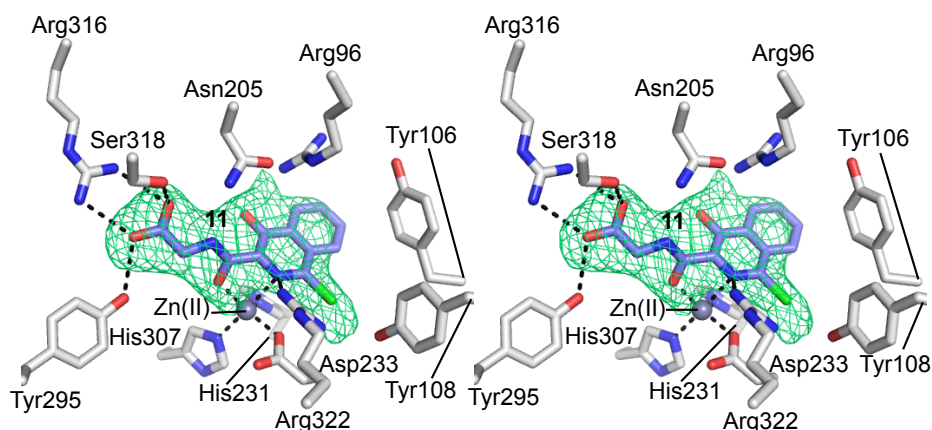


Figure 3 - 22. Wall-eyed stereo views from the active site of a crystal structure of FTO in complex with **11 (blue) (resolution: 2.50 Å).** FTO residue side chains C (white), O (red), N (blue), Zn(II) (grey sphere), and electrostatic interactions are indicated as black dashes. mF_o-DF_c ligand omit maps (contour level, $\sigma = 3.0$) are indicated as lime green mesh. Figure and figure legend are reprinted (adapted) with permission from Aik, W. et al (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, 56, 3680-3688. Copyright 2013 American Chemical Society. Protein structure models were produced using MacPyMOL v1.3 (13).

3.2.21 Crystal Structure of FTO in Complex with Rhein and Citrate

Rhein (**19**) and compound **20** (Figure 3 - 8) have been reported to be FTO inhibitors (20) and the inhibition assay described in Section 3.2.4 confirmed so (IC₅₀ values 9.0 μ M and 11 μ M, respectively). Compound **19** has been previously predicted by docking studies to bind across the 2OG, Fe(II), and 3meT binding sites (20). To verify the predicted binding mode, structural studies were carried out to investigate the binding modes of **19** and **20**. Co-crystallisation of rhein (**19**) and **20** with FTO were carried out using the optimised conditions for crystallising the other FTO-inhibitor complexes. Crystals grown in the presence of **19** were yellow (Section 3.2.7, Figure 3 - 13g), while those grown in the presence of **20** were red, consistent with the respective colours of the compounds. However the red crystals grown in the presence of **20** were clustered and appeared twinned. Therefore, only the structure of FTO in complex with **19** could be determined. Initially, a structure of the FTO-**19** complex was

determined to 2.8 Å resolution but the structure showed incomplete mF_o-DF_c omit density for compound **19** in the nucleic acid substrate binding site, possibly due to incomplete occupancy of the compound. However, when refined with the model of **19**, 2mF_o-DF_c electron density for **19** could be observed. Glycerol was used as a cryoprotectant during crystal cooling and was likely to compete with **19** for the substrate binding site. To overcome this hurdle, a final concentration of 1mM **19** was added to the cryoprotectant solution to minimise competition from glycerol. Crystals of FTO grown in the presence of **19** were soaked in the cryoprotectant for approximately 30 s before being flash-cooled. This strategy successfully yielded a structure with a clear mF_o-DF_c omit density for **19** (Figure 3 - 23a), which was refined to full occupancy.

The obtained FTO-**19** structure shows that compound **19** binds in the nucleotide substrate binding site, in contrast with predictions that it would bind across the 2OG, Fe(II) and nucleotide binding sites (20). The 2-carboxylate of **19** is positioned to hydrogen bond with the side chain of Ser229 (distance = 2.7 Å). An aromatic ring of the tricyclic **19** adjacent to the carboxylate group is positioned to form partial π - π interactions with His231. The 4-hydroxyl group of this ring forms a hydrogen bond with the side chain of Arg96. Superimposition of the FTO-**19** structure with the FTO-**1**-3meT structure shows that **19** binds in a similar position to 3meT (Figure 3 - 23b), validating the proposal by Chen *et al.* that **19** is a substrate competitor (20), but **19** is not both a substrate and 2OG competitor.

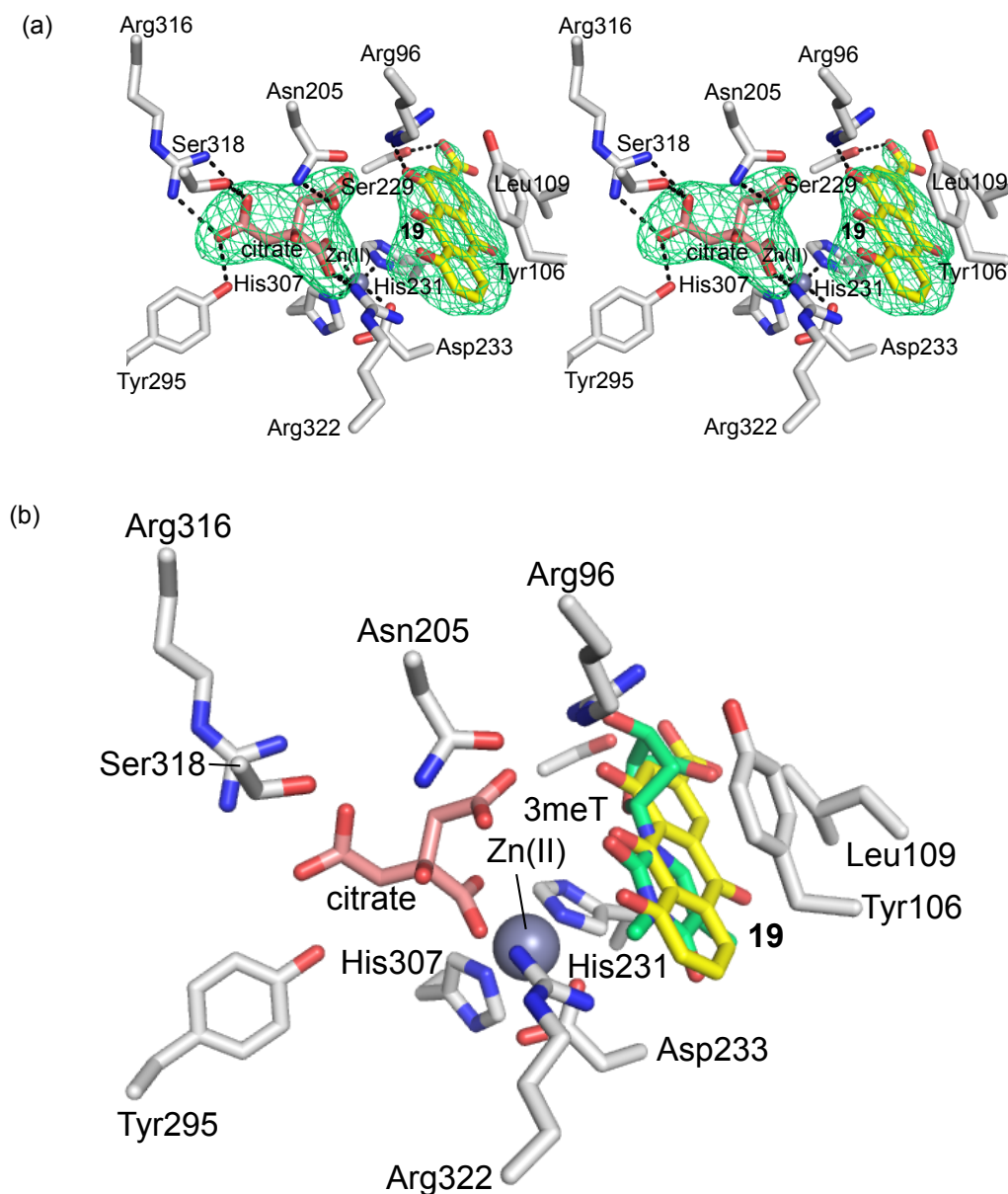


Figure 3 - 23. Views of a crystal structure of the FTO-19 complex. **(a)** Wall-eyed stereo views of the active site of a crystal structure of FTO in complex with citrate (light red) and **19** (yellow) (resolution: 2.60 Å). FTO residue side chains C (white), O (red), N (blue), Zn(II) (grey sphere), and electrostatic interactions are indicated as black dashes. mF_o-DF_c ligand omit maps (contour level, $\sigma = 3.0$) are indicated as lime green mesh. **(b)** Superimposition of 3meT (lime green) from the FTO-1-3meT structure (PDB ID 3LFM; resolution: 2.5 Å) (12) on the FTO-Zn-**19** structure (resolution: 2.60 Å). Figure and figure legend are reprinted (adapted) with permission from Aik, W. *et al* (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, 56, 3680-3688. Copyright 2013 American Chemical Society. Protein structure models were produced using MacPyMOL v1.3 (13).

A citrate molecule, probably derived from the crystallisation buffer, was also observed in the 2OG binding site in the FTO-**19** structure. One of the citrate carboxylates is positioned to interact with the active site metal, similar to the succinate binding observed in the other 2OG oxygenase structures (44), and another carboxylate interacts with the side chain of Arg316. The third citrate carboxylate is positioned to hydrogen-bond to the side chain of Asn205.

Because citrate was observed to bind in the 2OG binding site of FTO, it was tested by Dr Marina Demetriades as an inhibitor of FTO using the developed LC-based inhibition assay and was found to be a weak inhibitor with an IC₅₀ value of 300 µM. Although citrate was found to be a weak inhibitor of FTO, it does raise a possibility that FTO could be inhibited by tricarboxylic acid (TCA) cycle intermediates and related compounds such as *L*-2-hydroxyglutarate (2HG), which has been reported to be pro-oncogenic (45,46).

3.2.22 Second Generation FTO Inhibitors

The structures of FTO in complex with compound **6** and **19** were then used as templates to design second generation inhibitors. This new series of compounds was designed to occupy both the 2OG and the nucleotide binding pockets. The designed compounds should have a higher binding affinity than compounds that are either 2OG competitors or prime substrate competitors. The approach of linking a 2-oxoglutarate mimetic (*N*-oxalyl *D*-cysteine) with a T11C mutant (i.e. Thr11 mutated to Cys) substrate fragment (histone H3 trimethylated lysine 9, H3K9me3) has previously been demonstrated to successfully create a potent and selective inhibitor for the histone demethylase JMJD2A (47). To apply this

approach on FTO, an aromatic group was substituted at R₃ (see Section 3.2.19, Figure 3 - 21) of compound **6** so that the aromatic group would most likely occupy the nucleotide binding site. Compounds **21-24** were synthesised and their IC₅₀ values were determined using the LC-based *in vitro* assay described in Section 3.2.4 (Figure 3 - 24). Compound **23** (IC₅₀ 4 µM) was the most potent inhibitor of the four second generation compounds, followed by **22** (IC₅₀ 15 µM), **24** (IC₅₀ 25 µM), and **21** (IC₅₀ 35 µM). **23** was the only compound that showed an improvement in potency (3.75 fold) relative to the parent compound **6** (IC₅₀ = 15 µM).

(Organic synthesis and inhibition assays were carried out by Caitlin Clunnie-O'Connor and Dr Marina Demetriades. Second generation inhibitors were designed in collaboration with Dr Marina Demetriades.)

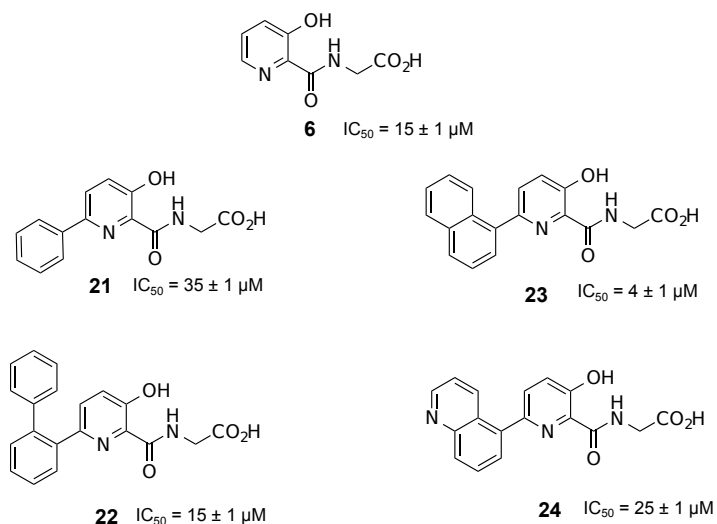


Figure 3 - 24. Structures and IC₅₀ values of pyridyl glycinamide compounds designed based on crystal structures of the FTO-6 and FTO-19 complexes. Compound 6 is also included for comparison.

Crystallographic studies were then carried out to investigate the binding mode of compound **23**. hFTO Δ 31 was crystallised in the presence of Zn(II) and compound **23** (crystallisation experiments were performed by Caitlin Clunnie-O'Connor), and the crystal structure of FTO-**23** complex was refined to 2.4 Å resolution. As predicted, the resultant structure revealed that the binding mode of **23** is similar to that of compound **6** (Figure 3 - 25a), with the parent pyridine glycinamide binding in the 2OG binding site. Moreover, the 6-naphthalene group bound in the nucleotide binding pocket almost perpendicular to the pyridine ring. Superimposition of FTO structures in complex with **23**, **6** and **19** shows a good overlap between the pyridine glycinamide side chains of **23** and compound **6** and between the naphthalene group of **23** and compound **19** (Figure 3 - 25b).

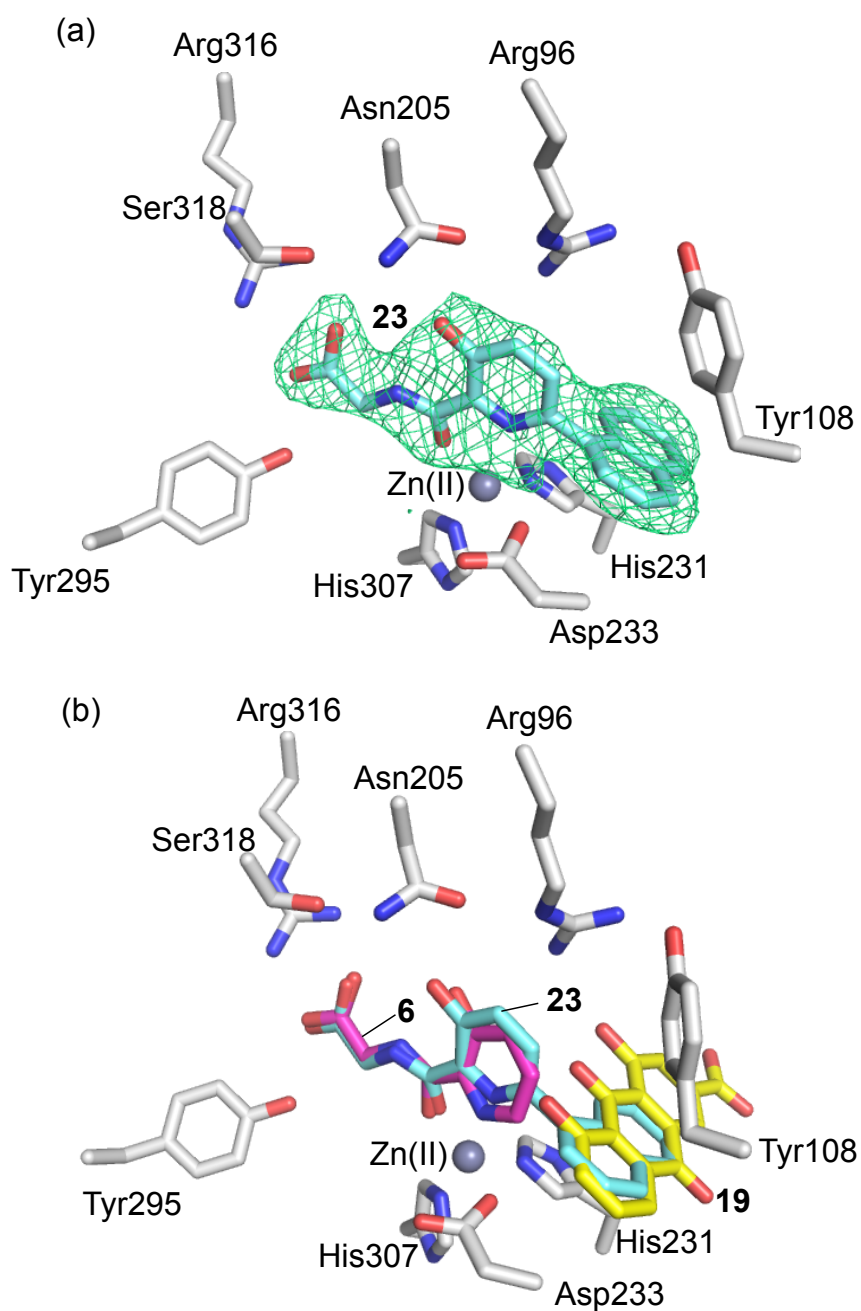


Figure 3 - 25. Views of a crystal structure of FTO in complex with 23. (a) Active site view of a structure of FTO in complex with 23 (cyan) (resolution: 2.32 Å). **(b)** Superimposition of compounds 6 and 19 from structures of the FTO-6 (resolution: 1.96 Å) and FTO-19 (resolution: 2.60 Å) complexes onto the active site of a structure of the FTO-23 complex (resolution: 2.32 Å). FTO residue side chains C (white), O (red), N (blue), Zn(II) (grey sphere), and electrostatic interactions are indicated as black dashes. mF_o-DF_c ligand omit maps (contour level, $\sigma = 3.0$) are indicated as lime green mesh. Protein structure models were produced using MacPyMOL v1.3 (13).

Table 3 - 3. Data collection and refinement statistics of the crystal structure of FTO-23 complex.

	FTO-23
PDB ID	4QHO
Crystal Size (Longest Dimension)	~200 μm
Data Collection	
Radiation Source	Synchrotron Diamond Beamline I04
Detector	Pilatus 6M-F
X-ray Wavelength ($\text{\AA}$)	1.2716
Temperature (K)	100
Resolution Range ($\text{\AA}$) $\text{\textcircled{R}}$	50.00-2.32 (2.40-2.32)
Space Group	<i>R</i> 3
Unit Cell Dimensions (<i>a</i> $\text{\AA}$, <i>b</i> $\text{\AA}$, <i>c</i> $\text{\AA}$, α° , β° , γ°)	141.38, 141.38, 83.95, 90°, 90°, 120°
Total Number of Reflections Observed	193273
Number of Unique Reflections	26707
Multiplicity $\text{\textcircled{R}}$	7.2 (4.3)
Completeness (%) $\text{\textcircled{R}}$	100.0 (100.0)
<i>I</i> / σ (<i>I</i>) $\text{\textcircled{R}}$	11.5 (2.0)
R_{merge} (%)	20.4
Wilson B Factor ($\text{\AA}^2$)	67.42
Software for Data Reduction	HKL-2000
Refinement	
R_{work} (%)	17.69
R_{free} (%)	20.76
$\text{\textcircled{R}}$ RMS Deviation (Bonds/Angle)	0.011 $\text{\AA}$ / 1.216°
Ramachandran Outliers (%)	0.2
$\text{\textcircled{R}}$ Average <i>B</i> Factor ($\text{\AA}^2$)	86.97
Number of Water Molecules	28
Software for Refinement	PHENIX

$\text{\textcircled{R}}$ Numbers in brackets indicate outermost shell

$$\S R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle \times 100$$

$$* R_{\text{work}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / |F_{\text{obs}}| \times 100$$

$\text{\textcircled{R}}$ R_{free} , based on 7.45% of the total reflections

Note: Refinement was cut off at 2.38 $\text{\AA}$ resolution to achieve >90% completeness for all resolution shells.

The combined biochemical and structural results of this study demonstrate that structure-based inhibitor design for FTO is achievable. Therefore, further development of more potent and selective inhibitors with the aid of crystal structures of FTO-ligand complexes is a realistic goal.

3.3 Conclusions

The roles of FTO in obesity and RNA demethylation have generated immense interest in validating FTO as an anti-obesity target (2,8,48,49). The results described in this chapter have uncovered a potential way of identifying potent and selective inhibitors of FTO. These inhibitors could potentially serve as starting point for the development of a new anti-obesity therapy or as probes for functional studies of FTO. An *in vitro* assay to quantify FTO enzyme activity has been developed and demonstrated to be useful for inhibition studies. Notably, two relatively potent inhibitors of FTO (compounds **11** and **18**) identified in this study are in clinical trials as prolyl-hydroxylase inhibitors to treat anaemia. Because these inhibitors also inhibit FTO, their off-target physiological effects should be carefully studied.

Importantly, a method to purify FTO protein that is of crystallisation quality has also been developed and resulted in 7 crystal structures of FTO in complex with inhibitors of different classes. The crystallographic studies have laid the foundation for structure-based inhibition studies on FTO. Modification of a pyridyl glycineamide scaffold demonstrates that FTO inhibitors can be improved through rational design with the aid of crystal structures. Future studies can now be directed towards developing more selective and cell permeable FTO

inhibitors, which will benefit from the biochemical and crystallographic work on FTO described in this study.

3.4 Experimental Procedures

3.4.1 Mutagenesis

To generate a pET28a plasmid that encodes for hFTO Δ 1-31, a mutagenesis protocol adapted from the QuikChange™ (Stratagene) site-directed mutagenesis protocol was carried out to delete DNA sequences encoding for amino acids 2-31 in the plasmid pET28a-hFTO, which encodes for full-length FTO. A 50 μ L polymerase chain reaction (PCR) mixture containing 5 μ L 10X Pfu Buffer, 1.75 μ L 14.3 ng/ μ L pET28a-hFTO plasmid, 1.25 μ L 100 ng/ μ L forward primer, 100 ng/ μ L reverse primer, 5 μ L dNTP mix (2.5 mM for each base A,T,G and C), 1 μ L 2.5U/ μ L Pfu polymerase, and 34.75 μ L miliQ water was prepared. The sequence for forward primer is

5'-TGTCGACATATGACCCCGAAAGATGATGAATTTTACCAGC-3' and for reverse primer, 5'-TGTCGAGAATTCTTACGGTTTCGCTTCCAGCAGC-3'. The mixture was then subjected to 'thermo-cycling', 95 °C for 30 s, then 18 cycles of 95 °C for 30 s, then 60 °C for 1 min, then 68 °C for 8 min 30 s, then 68 °C for 10 min. The mixture was then cooled to 4 °C. 1 μ L of DpnI restriction enzyme (New England Biolabs, NEB) was added to the PCR product mixture to digest the parent template plasmid. Subsequently, 1 μ L of the mixture was transformed into *E. coli* XL1 Blue competent cells and plated on LB agar containing 30 μ g/mL kanamycin antibiotics and left to grow overnight at 37 °C. Four colonies were picked and cultured in 5 mL 2TY cultures containing 30 μ g/mL kanamycin. Cells were then

harvested from the cultures from which plasmids were purified using a Qiagen Plasmid Miniprep Kit. The plasmids were then sent for sequencing to confirm that the planned mutagenesis had been successful.

3.4.2 Protein expression and purification

Protein purification methods were modified from methods previously reported for full-length mouse Fto (4). *E. coli* BL21 (DE3) transformed with the pET28a_hFTO2 plasmid (encoding for N-terminally hexahistidine-tagged full-length human FTO without the thrombin cleavage site between the hexahistidine tag and the protein) or pET28a_hFTOΔ31 (encoding for N-terminally hexahistidine-tagged human FTO with residues 2–31 deleted from the N-terminus) were grown (37 °C; 180 rpm) to an OD₆₀₀ of 0.6. FTO production was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and growth was continued at 16 °C for 16 h (hFTO) or 18 °C for 8 h (hFTOΔ31). The resultant cell pellets were stored at -80 °C. The cell pellets were thawed and resuspended in lysis buffer containing 20 mM Tris-HCl, pH 7.5, 500 mM NaCl and 10 mM imidazole, and an addition of 1 mg DnaseI, 10 mM MgCl₂, and a Roche cOmplete protease inhibitor tablet, then lysed by sonication on ice. The lysates were cleared by centrifugation and the supernatant was loaded onto a 5 mL HisTrap FF column (GE Healthcare) and purified using an AKTA FPLC system. The column was treated with 20 mM Tris-HCl pH 7.5, 500 mM NaCl, and 40 mM imidazole, then eluted with 20 mM Tris-HCl pH 7.5, 500 mM NaCl 500 mM imidazole. The eluted hFTO protein was further purified with S75 size exclusion chromatography pre-equilibrated with 25 mM Tris-HCl pH 7.5.

For crystallography, the eluted hFTO Δ 31 protein solution was then treated with 200 mM of ethylenediaminetetraacetic acid (EDTA). hFTO Δ 31 was further purified using a (5 mL) HiTrap Heparin column (GE Healthcare) followed by a (20 mL) MonoQ column (Pharmacia); in both cases a gradient of buffer A (25 mM Tris-HCl pH 7.5) and 50% of buffer B (25 mM Tris-HCl pH 7.5, 1 M NaCl) was used. The purified hFTO Δ 31 protein was then buffer exchanged into 25 mM Tris-HCl pH7.5. Both hFTO and hFTO Δ 31 were concentrated to 20 mg/mL for storage.

3.4.3 Dynamic scanning fluorimetry assays (17)

For dynamic scanning fluorimetry (DSF) assays, a mixture (50 μ L) containing 2 μ M full-length FTO protein, 1x SYPRO orange dye (Invitrogen) (from commercial stock that is 5000x), 50 μ M diammonium iron (II) sulfate complex, 200 μ M compound (or 2% DMSO control), and 50 mM HEPES, pH 7.5 was prepared. Assays were performed using a MiniOpticon™ Real-Time PCR Detection System (Bio-Rad). The sample plate was inserted into the machine and fluorescence readings were taken between 25 – 90 °C, increasing the temperature linearly over 1 h 20 min. Excitation and emission wavelength were 470 nm and 570 nm respectively. Melting curves were analysed using GraphPad Prism®.

3.4.4 *In vitro* demethylation assays

For pilot *in vitro* demethylation assays, a 100 μ L reaction mixture containing 5 μ M FTO Δ 1-31, 80 μ M N-3-methylthymidine, 500 μ M ammonium iron (II) sulphate, 1 mM 2OG, 2 mM sodium L-ascorbate and 50 mM MES, pH 6.25 was

incubated at either 16 °C, room temperature, or 37 °C for 15 h. A control reaction without FTOΔ1-31 was incubated at 37 °C for 15 h. Reactions were quenched with 200 µl methanol and samples were centrifuged with a Sigma 1-14 centrifuge at 24000 rpm for 10 min. Supernatant was evaporated and reconstituted in 50 µl water. Product (thymidine) and substrate (*N*³-methylthymidine) were separated by Waters Acquity Ultra Performance Liquid Chromatography (UPLC) oligonucleotide separation column (Waters Acquity OST C18, 1.7 µm, 2.1 x 100 mm) with a gradient of 0.1% formic acid in water (A) to 0.1% formic acid in methanol (B): 10–20 % B in A over 5 min, 20–100 % B in A over 0.5 min, 100 % B over 2.5 min, 100–10 % B in A over 0.5 min, 10 % B over 2 min. Rt (thymidine) = 1.20 min, Rt (*N*-3-methylthymidine) = 1.82 min. The identities of the peaks were identified by positive ion mode electrospray ionization – time of flight (ESI-ToF) mass spectrometry (Waters LCT Premier XE).

For inhibition assays, a 50 µL reaction mixture containing final concentrations of 3 µM FTO, 70 µM 3meT nucleoside, 160 µM 2OG, 500 µM L-ascorbate, 100 µM diammonium iron(II) sulfate complex, 200 µM test compound (2% DMSO in control reaction) and 50 mM 2-(*N*-morpholino)ethanesulfonic acid (MES), pH 6.3 was incubated at room temperature for 1 h. The reaction was quenched with methanol (50 µl) and centrifuged to remove precipitated protein. The supernatant was then dried using an Eppendorf Speedvac concentrator and reconstituted with (50 µl) water. The product (thymidine) and substrate (3meT) were separated using Waters Acquity Ultra Performance Liquid Chromatography (UPLC) BEH C18 column, 130 Å, 1.7 µm, 2.1 mm x 50 mm with a gradient of 95% A (H₂O with 1% formic

acid) to 80% B (methanol with 0.1% formic acid) over 5 min. UV detection wavelength was set to 266 nm room temperature (thymidine) = 0.95 min, room temperature (3meT) = 1.48 min. (The liquid chromatography method was modified from previously reported (12).) The peaks were identified using a positive ion mode in the ESI-ToF mass spectrometry (Waters LCT Premier XE machine). Peaks were integrated using MassLynx software (Agilent) and the percentage conversion of 3meT to thymidine was used to quantify the activity of FTO in the presence of inhibitors. For IC₅₀ determination, assays were carried out in triplicate with a range of compound concentration (0 μ M, 10 μ M, 30 μ M, 100 μ M, 300 μ M, 1 mM, 3 mM, and 10 mM). IC₅₀ values were calculated from dose-response curve plotted using GraphPad Prism®.

3.4.6 Crystallography

Crystals of hFTO Δ 31 in complex with **1**, **2**, **3**, **6**, **11** or **19** were grown in hanging drops at 293 K by the vapor diffusion method using 24 well pre-greased VDX plates and round plastic coverslips (HR8-082, Hampton Research). Protein solution and reservoir solution (200 μ L) conditions are listed in Table 3 - 4 (see below). Hanging drops were set up by adding protein solution to reservoir solution with a 2:1 ratio (total volume 3 μ L). The resultant crystals cryo-protected by soaking them for approximately 30 s in well-solution diluted with 15 – 20% (v/v) glycerol (cryoprotectant solution for the FTO-**19** crystal contains an additional 1 mM **19**) then flash cooled in liquid nitrogen. Data were collected at 100 K for single crystals of the FTO-**1**, FTO-**3** and FTO-**6** complexes using a Rigaku FR-E+ Superbright diffractometer (X-ray wavelength: 1.5412 Å) equipped with a copper rotating anode, osmic HF optics, and a Saturn 944+ CCD

detector. Data for FTO-2, FTO-11, and FTO-19 were collected at Diamond Beamline I03 (X-ray wavelength: 0.9763 Å), while data for FTO-23 were collected at Diamond Beamline I04 (X-ray wavelength: 1.2716 Å). Both beamlines were equipped with a Pilatus 6M-F. All data were collected at 100 K using single crystals.

All data were indexed, integrated, and scaled using HKL2000 (32). Structures were solved by molecular replacement using the MR-PHASER (33) subroutine in PHENIX (35) using a structure of the FTO:Fe:NOG:3meT complex (PDB ID: 3LFM; resolution: 2.50 Å) as a search model, with the exception of FTO-23, the structure of which was solved using a structure of the FTO-6 complex (PDB ID 4IE5; resolution: 1.96 Å) as a search model. Iterative cycles of model building and refinement were performed using COOT (34) and PHENIX until decreasing R and R_{free} no longer converged. Data collection and refinement statistics table can be found in Table 3 - 2 and Table 3 - 3 (see Sections 3.2.15 and 3.2.22).

Table 3 - 4. Crystallisation conditions for protein and reservoir solutions.

	FTO-1	FTO-2	FTO-3	FTO-6	FTO-11	FTO-19	FTO-23
PDB ID	4IDZ	4IE0	4IE4	4IE5	4IE6	4IE7	4QHO
Protein Solution Conditions	~8 mg/ml hFTOΔ31, 1 mM NiCl ₂ , 3 mM 1 , 5 mM 6meA	~8 mg/ml hFTOΔ31, 1 mM ZnSO ₄ , 1 mM 2	~8 mg/ml hFTOΔ31, 1 mM ZnSO ₄ , 2 mM 3	~8 mg/ml hFTOΔ31, 1 mM ZnSO ₄ , 1 mM 6	~8 mg/ml hFTOΔ31, 1 mM ZnSO ₄ , 1.5 mM 11	~8 mg/ml hFTOΔ31, 1 mM ZnSO ₄ , 2 mM 19	~8 mg/ml hFTOΔ31, 1 mM ZnSO ₄ , 1 mM 23
Reservoir Solution Conditions	85 mM trisodium citrate pH 5.6, 14% PEG3350, 10% glycerol	100 mM trisodium citrate pH 5.6, 11% PEG3350, 4% tert-butanol	100 mM trisodium citrate pH 5.6, 13% PEG3350, 4% tert-butanol	100 mM trisodium citrate pH 5.6, 9% PEG3350, 4% tert-butanol	100 mM trisodium citrate pH 5.6, 11.5% PEG3350, 4% tert-butanol	100 mM trisodium citrate pH 5.6, 11.5% PEG3350, 4% tert-butanol	100 mM trisodium citrate pH 5.6, 8% PEG3350, 4% tert-butanol

3.4.7 Synthesis of *N*³-methylthymidine (16)

The method used for the synthesis of *N*³-methylthymidine was modified from a previously reported method (16). To a solution of thymidine (0.79 g, 3.26 mmol) and potassium carbonate (0.83 g, 6.01 mmol) in MeOH (50 mL), methyl iodide (0.38 mL, 6.1 mmol) was added dropwise at room temperature. The reaction was heated at 37 °C for 24 h, then 5 mL of acetic acid was added to neutralise the mixture. The solvent was evaporated *in vacuo* and the crude compound was redissolved in water. The product was purified using preparative reverse-phase HPLC (t_R = 25 min) (Vydac 218TP C18, 250 mm x 22 mm, 10-15 μ m, Grace Davison, USA) using a gradient of: 2% B for 5 min, 2-98% B in A over 30 min, 98% B for 5 min. A: 0.1% formic acid v/v water, B: 0.1% formic acid v/v acetonitrile. Product (24%; 0.19 g) mp 129 – 133 °C; IR (neat) ν/cm^{-1} 3500 (OH), 3412 (OH), 3300 (OH), 1683 (C=O), 1665 (C=O), 1622 (C=O); ¹H NMR (400 MHz, DMSO) δ 7.77 (1H, s, 6-H), 6.20 (1H, t, J = 6.5, 1'-H), 5.04 (1H, *br* s, 5'-OH), 4.24 (1H, m, 3'-H), 3.78 (1H, app. q., J = 4.0, 4'-H), 3.58 (2H, m, 5'-H), 3.16 (3H, s, N-CH₃), 2.09 (2H, m, 2'-H), 1.82 (3H, s, CH₃); ¹³C NMR (400 MHz, DMSO) δ 162.9 (C=O), 150.6 (C=O), 134.5 (6C), 108.3 (5C), 87.4 (4'C), 84.9 (1'C), 70.3 (3'C), 61.2 (5'C), 27.5 (N-CH₃), 13.0 (CH₃); HRMS calcd. for C₁₁H₁₆N₂NaO₅ [M+Na]⁺ 279.0951; Found 279.0951.

3.4.8 RNA-pulldown

Approximately 2 g of an HEK293T cell pellet was resuspended with 4 mL of buffer containing 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 10 mM imidazole and 0.5% NP-40, which was pre-added with 100 μ L RNase-free DNase I (Qiagen), 100 μ L 1M MgCl₂, and 40 μ L protease inhibitor cocktail (Sigma-Aldrich).

Recombinant hexahistidine-tagged hFTO protein (16 mg), purified by nickel affinity, heparin chromatography and MonoQ ion exchange, was added to the resuspended HEK293T cell mixture. The mixture was then incubated in RT for 5 min before being incubated on ice for 3 hours. Lysate was then centrifuged at 21000 g at 4 °C. Supernatant was then loaded onto a HisTrap HP column (GE) and hFTO was purified using standard nickel affinity purification protocol described above (see Section 3.4.2). hFTO protein was concentrated to about 30 mg/mL and stored at -80 °C.

3.4.9 RNA Sequencing

Approximately 25 µL of concentrated FTO protein (10 – 30 mg/mL) was heat-treated at 90 °C for 10 min to denature the protein. The heat-treated protein was then centrifuged at 24000 rpm for 10 min to pellet the precipitated protein. To synthesise the first strand of cDNA from RNA, a reaction mixture containing 10 µL of 2X First Strand Master Mix (Agilent), 3 µL 0.1 µg/µL random primers (Agilent), 1 µL AffinityScript RT/RNase Block Enzyme Mixture (Agilent), and 6 µL RNA obtained from the supernatant fraction of the boiled FTO protein was prepared. The reaction was carried out at 25 °C for 5 min followed by 42 °C for 1 h. The sample was then heat inactivated on a heat block (90 °C) for 5 min. The second strand of the cDNA was synthesised using materials and protocol from NEB. A reaction mixture containing 20 µL first strand cDNA, 48 µL nuclease-free water, 8 µL 10X Second Strand Synthesis Buffer Mix (NEB), and 4 µL Second Strand Synthesis Enzyme Mix (NEB) was prepared. The reaction was carried out at 16 °C for 2.5 h. The synthesised cDNA was purified using a PCR purification kit (Thermo Scientific) and eluted in 25 µL of miliQ water. The DNA solution was

concentrated using a vacuum concentrator (Eppendorf Concentrator Plus 230 V/50-60 Hz) to approximately 5 μ L. The cDNA was then A-tailed with Taq polymerase (Promega) by preparing a reaction mixture containing 5 μ L cDNA, 1 μ L Taq 10X Reaction Buffer with $MgCl_2$ (Promega), 1 μ L 2 mM deoxyadenosine triphosphate (dATP), 1 μ L Taq polymerase (Promega), 2 μ L MiliQ water was incubated at 70 $^{\circ}C$ for 30 min. The A-tailed cDNA inserts were ligated into pGEM-Teasy vector. The ligation mixture contained 5 μ L 2X Rapid Ligation Buffer, 1 μ L pGEM-T Easy Vector (Promega), 3 μ L cDNA from the previous reaction, and 1 μ L T4 DNA ligase. The ligation mixture was incubated at 4 $^{\circ}C$ overnight before being transformed into *Escherichia coli* XL1 Blue or XL10 Gold competent cells using standard transformation protocol. The transformed cells were plated onto LB agar containing 100 μ g/mL ampicillin and pre-spread with 50 μ L 20 mg/mL X-Gal, 20 μ L 0.5 M IPTG, and 50 μ L 2TY media. White colonies were then picked for PCR colony screening and plasmid purification. The plasmids were finally sent for DNA sequencing to determine the sequence of the inserts.

Experimental procedures for the synthesis of compounds used in this study can be found in the original publication (Aik *et al.* (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, 56, 3680-3688.) attached in the Appendix.

3.5 References

1. Dina, C., Meyre, D., Gallina, S., Durand, E., Korner, A., Jacobson, P., Carlsson, L.M., Kiess, W., Vatin, V., Lecoecur, C. *et al.* (2007) Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet*, **39**, 724-726.

2. Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. *et al.* (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889-894.
3. Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. *et al.* (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*, **316**, 1341-1345.
4. Gerken, T., Girard, C.A., Tung, Y.C., Webby, C.J., Saudek, V., Hewitson, K.S., Yeo, G.S., McDonough, M.A., Cunliffe, S., McNeill, L.A. *et al.* (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*, **318**, 1469-1472.
5. Church, C., Moir, L., McMurray, F., Girard, C., Banks, G.T., Teboul, L., Wells, S., Bruning, J.C., Nolan, P.M., Ashcroft, F.M. *et al.* (2010) Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet*, **42**, 1086-1092.
6. Fischer, J., Koch, L., Emmerling, C., Vierkotten, J., Peters, T., Bruning, J.C. and Ruther, U. (2009) Inactivation of the Fto gene protects from obesity. *Nature*, **458**, 894-898.
7. Jia, G., Yang, C.-G., Yang, S., Jian, X., Yi, C., Zhou, Z. and He, C. (2008) Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Letters*, **582**, 3313-3319.
8. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.-G. *et al.* (2011) N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nature Chemical Biology*, **7**, 885-887.
9. Karra, E., O'Daly, O.G., Choudhury, A.I., Yousseif, A., Millership, S., Neary, M.T., Scott, W.R., Chandarana, K., Manning, S., Hess, M.E. *et al.* (2013) A link between FTO, ghrelin, and impaired brain food-cue responsivity. *Journal of Clinical Investigation*, **123**, 3539-3551.
10. Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G. *et al.* (2013) N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature*, **505**, 117-120.
11. Woon, E.C.Y., Demetriades, M., Bagg, E.A.L., Aik, W., Krylova, S.M., Ma, J.H.Y., Chan, M., Walport, L.J., Wegman, D.W., Dack, K.N. *et al.* (2012) Dynamic Combinatorial Mass Spectrometry Leads to Inhibitors of a 2-Oxoglutarate-Dependent Nucleic Acid Demethylase. *Journal of Medicinal Chemistry*, **55**, 2173-2184.
12. Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y. and Chai, J. (2010) Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature*, **464**, 1205-1209.
13. Schrodinger, LLC. (2010).
14. Aik, W., Demetriades, M., Hamdan, M.K.K., Bagg, E.A.L., Yeoh, K.K., Lejeune, C., Zhang, Z., McDonough, M.A. and Schofield, C.J. (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *Journal of Medicinal Chemistry*, **56**, 3680-3688.

15. Church, C., Lee, S., Bagg, E.A., McTaggart, J.S., Deacon, R., Gerken, T., Lee, A., Moir, L., Mecinovic, J., Quwailid, M.M. *et al.* (2009) A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet*, **5**, e1000599.
16. Balzarini, J., Celen, S., Karlsson, A., de Groot, T., Verbruggen, A. and Bormans, G. (2006) The effect of a methyl or 2-fluoroethyl substituent at the N-3 position of thymidine, 3'-fluoro-3'-deoxythymidine and 1-beta-D-arabinosylthymine on their antiviral and cytostatic activity in cell culture. *Antivir Chem Chemother*, **17**, 17-23.
17. Niesen, F.H., Berglund, H. and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc*, **2**, 2212-2221.
18. Rose, N.R., McDonough, M.A., King, O.N.F., Kawamura, A. and Schofield, C.J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev*, **40**, 4364-4397.
19. Hopkinson, R.J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K.H., Li, X.S., Kristensen, J.B.L., King, O.N.F., Chan, M.C. *et al.* (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chemical Science*, **4**, 3110.
20. Chen, B., Ye, F., Yu, L., Jia, G., Huang, X., Zhang, X., Peng, S., Chen, K., Wang, M., Gong, S. *et al.* (2012) Development of Cell-Active N6-Methyladenosine RNA Demethylase FTO Inhibitor. *Journal of the American Chemical Society*, **134**, 17963-17971.
21. Gadgil, H., Oak, S.A. and Jarrett, H.W. (2001) Affinity purification of DNA-binding proteins. *J Biochem Bioph Meth*, **49**, 607-624.
22. Hornby, D., Ford, K. and Shore, P. (1992) Purification of DNA binding proteins by affinity and ion exchange chromatography. *Methods Mol Biol*, **11**, 273-286.
23. Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S. and Hochstrasser, D. (1993) The Focusing Positions of Polypeptides in Immobilized Ph Gradients Can Be Predicted from Their Amino-Acid-Sequences. *Electrophoresis*, **14**, 1023-1031.
24. Bjellqvist, B., Basse, B., Olsen, E. and Celis, J.E. (1994) Reference Points for Comparisons of 2-Dimensional Maps of Proteins from Different Human Cell-Types Defined in a Ph Scale Where Isoelectric Points Correlate with Polypeptide Compositions. *Electrophoresis*, **15**, 529-539.
25. Berget, S.M. and Robberson, B.L. (1986) U1, U2, and U4/U6 small nuclear ribonucleoproteins are required for in vitro splicing but not polyadenylation. *Cell*, **46**, 691-696.
26. Barboric, M., Lenasi, T., Chen, H., Johansen, E.B., Guo, S. and Peterlin, B.M. (2009) 7SK snRNP/P-TEFb couples transcription elongation with alternative splicing and is essential for vertebrate development. *Proc Natl Acad Sci U S A*, **106**, 7798-7803.
27. Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994) Screening and optimization strategies for macromolecular crystal growth. *Acta Crystallogr D Biol Crystallogr*, **50**, 414-423.

28. Sleeman, M.C., Sorensen, J.L., Batchelar, E.T., McDonough, M.A. and Schofield, C.J. (2005) Structural and mechanistic studies on carboxymethylproline synthase (CarB), a unique member of the crotonase superfamily catalyzing the first step in carbapenem biosynthesis. *J Biol Chem*, **280**, 34956-34965.
29. Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.
30. Matthews, B.W. (1976) X-Ray Crystallographic Studies of Proteins. *Annu Rev Phys Chem*, **27**, 493-523.
31. Kantardjieff, K.A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci*, **12**, 1865-1871.
32. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol*, **276**, 307-326.
33. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
34. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D*, **66**, 486-501.
35. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*, **66**, 213-221.
36. Conejo-Garcia, A., McDonough, M.A., Loenarz, C., McNeill, L.A., Hewitson, K.S., Ge, W., Lienard, B.M., Schofield, C.J. and Clifton, I.J. (2010) Structural basis for binding of cyclic 2-oxoglutarate analogues to factor-inhibiting hypoxia-inducible factor. *Bioorg Med Chem Lett*, **20**, 6125-6128.
37. Rose, N.R., Ng, S.S., Mecinovic, J., Lienard, B.M., Bello, S.H., Sun, Z., McDonough, M.A., Oppermann, U. and Schofield, C.J. (2008) Inhibitor scaffolds for 2-oxoglutarate-dependent histone lysine demethylases. *Journal of Medicinal Chemistry*, **51**, 7053-7056.
38. Koski, M.K., Hieta, R., Bollner, C., Kivirikko, K.I., Myllyharju, J. and Wierenga, R.K. (2007) The active site of an algal prolyl 4-hydroxylase has a large structural plasticity. *Journal of Biological Chemistry*, **282**, 37112-37123.
39. King, O.N., Li, X.S., Sakurai, M., Kawamura, A., Rose, N.R., Ng, S.S., Quinn, A.M., Rai, G., Mott, B.T., Beswick, P. *et al.* (2010) Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS ONE*, **5**, e15535.
40. Chowdhury, R., Candela-Lena, J.I., Chan, M.C., Greenald, D.J., Yeoh, K.K., Tian, Y.M., McDonough, M.A., Tumber, A., Rose, N.R., Conejo-Garcia, A. *et al.* (2013) Selective Small Molecule Probes for the Hypoxia Inducible Factor (HIF) Prolyl Hydroxylases. *ACS Chem Biol*.
41. Yan, L., Colandrea, V.J. and Hale, J.J. (2010) Prolyl hydroxylase domain-containing protein inhibitors as stabilizers of hypoxia-inducible factor: small molecule-based therapeutics for anemia. *Expert Opin Ther Pat*, **20**, 1219-1245.

42. Warshakoon, N.C., Wu, S.D., Boyer, A., Kawamoto, R., Sheville, J., Renock, S., Xu, K., Pokross, M., Zhou, S., Winter, C. *et al.* (2006) Structure-based design, synthesis, and SAR evaluation of a new series of 8-hydroxyquinolines as HIF-1 α prolyl hydroxylase inhibitors. *Bioorg Med Chem Lett*, **16**, 5517-5522.
43. Macdougall, I.C. (2012) New anemia therapies: translating novel strategies from bench to bedside. *Am J Kidney Dis*, **59**, 444-451.
44. Hausinger, R.P. (2004) FeII/ α -ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol*, **39**, 21-68.
45. Chowdhury, R., Yeoh, K.K., Tian, Y.M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K., Li, X.S., Woon, E.C., Yang, M. *et al.* (2011) The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep*, **12**, 463-469.
46. Frezza, C., Pollard, P.J. and Gottlieb, E. (2011) Inborn and acquired metabolic defects in cancer. *J Mol Med (Berl)*, **89**, 213-220.
47. Woon, E.C.Y., Tumber, A., Kawamura, A., Hillringhaus, L., Ge, W., Rose, N.R., Ma, J.H.Y., Chan, M.C., Walport, L.J., Che, K.H. *et al.* (2012) Linking of 2-Oxoglutarate and Substrate Binding Sites Enables Potent and Highly Selective Inhibition of JmJc Histone Demethylases. *Angewandte Chemie International Edition*, **51**, 1631-1634.
48. Jia, G., Fu, Y. and He, C. (2013) Reversible RNA adenosine methylation in biological regulation. *Trends Genet*, **29**, 108-115.
49. Thalhammer, A., Aik, W., Bagg, E.A.L. and Schofield, C.J. (2012) The potential of 2-oxoglutarate oxygenases acting on nucleic acids as therapeutic targets. *Drug Discovery Today: Therapeutic Strategies*, **9**, e91-e100.

Chapter 4

Crystal structure of human RNA N^6 -methyladenine demethylase ALKBH5

Most of the work described in this chapter has been reported in Aik *et al.* 'Structure of human RNA N^6 -methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by Oxford University Press (1).

4.1 Introduction

There are more than 100 different post-transcriptional modifications found on RNA, some of which have important functions in RNA stability, tRNA folding, decoding during translation, recognition of tRNA by aminoacyl-tRNA synthetases, and regulation of gene silencing (2-4). However, the functions of a majority of these modifications are still unknown. Perhaps the most studied of these modifications is the mRNA 5' end 7-methylguanosine cap ($m^7G(5')ppp(5')N$), which serves as a regulator for mRNA export, stability and translation (5). N^6 -Methyladenine (m^6A) is the most abundant internal modification on mammalian mRNA and is installed by the methyltransferases (also known as 'writers') METTL3/METTL14 heterodimer complex (6-8). m^6A has recently received considerable attention due to the discovery of its 'eraser' enzymes, i.e. the enzymes that remove the modification from mRNA (9-11). The two identified erasers are the fat mass and obesity associated protein (FTO) and

AlkB homologue 5 (ALKBH5), both of which belong to the AlkB oxygenase subfamily (12-14) (see Chapter 1). RNA sequencing data have also demonstrated that m⁶A is specifically found within the 5'-RRm⁶ACU-3' (where R is purine) consensus motif (15,16). Recently, m⁶A levels were reported to influence mRNA stability through the proteins YTHDF2 and YTHDF3 that bind specifically to m⁶A (15-17). These proteins are also known as 'readers'. Taken together, discoveries of the writers, readers, and erasers of m⁶A indicate that this modification is reversible and is involved in the regulation of gene expression (9).

ALKBH5 is a human homologue of an *E. coli* DNA repair protein, AlkB (13). It is most highly expressed in the lung, followed by testis, pancreas, spleen and ovary (14). ALKBH5 has been shown to be a 2OG oxygenase because it has 2-oxoglutarate (2OG) and Fe(II) dependent activity (13). ALKBH5 is also upregulated in hypoxia and is localised in the nucleus (13). However, at the time of these preliminary characterisations, the substrate of ALKBH5 was unknown. ALKBH5 has recently been shown to demethylate m⁶A in single-stranded RNA (ssRNA) (similar to FTO) (Figure 4 - 1), and also colocalises with splice factors in nuclear speckles, suggesting that it may be involved in mRNA processing, perhaps in splicing (11). Additionally, Zheng *et al.* demonstrated that siRNA knockdown of ALKBH5 in HeLa cells affected mRNA export from the nucleus to the cytoplasm, and they also reported that a global knockout of ALKBH5 in mice impaired spermatogenesis (11).

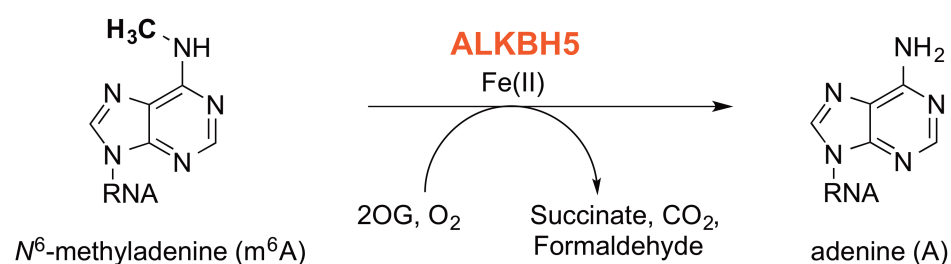


Figure 4 - 1. ALKBH5 catalyses demethylation of *N*⁶-methyladenine (m⁶A).

Due to the physiological importance of ALKBH5 and because ALKBH5 shares a similar substrate to that for FTO, i.e. m⁶A, structural studies on ALKBH5 were undertaken. Moreover, since the catalytic domains of FTO and ALKBH5 belong to the AlkB subfamily, a structure of ALKBH5 would aid development of inhibitors that are selective for ALKBH5, and likewise, for FTO. This chapter describes crystallographic studies on human ALKBH5 and compares a structure of ALKBH5 with structures of other nucleic acid modifying oxygenases (NAOX). The development of an inhibition assay for ALKBH5 is also described in this chapter.

4.2 Results and Discussions

4.2.1 Protein expression and purification

A construct of hexahistidine-tagged human ALKBH5 residues 66-292 (ALKBH5₆₆₋₂₉₂) was expressed in *Escherichia coli* BL21 (DE3). The produced protein was purified using a method adapted from FTO purification, which comprises of nickel affinity chromatography, followed by heparin chromatography and MonoQ anion exchange chromatography. The resulting protein appeared as a

single symmetrically shaped peak on MonoQ chromatography and was >95% pure according to SDS-PAGE analysis (Figure 4 - 2b).

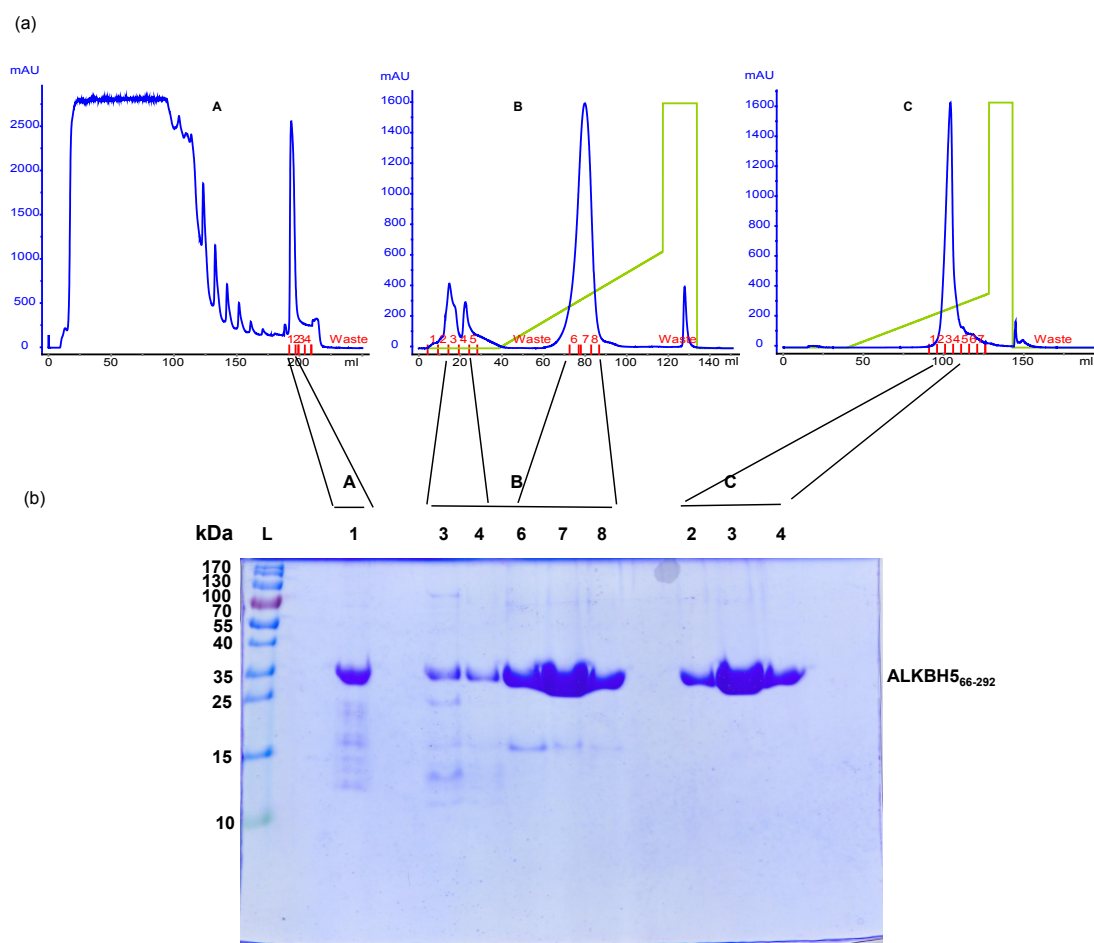


Figure 4 - 2. Purification of ALKBH5₆₆₋₂₉₂. (a) AKTA FPLC UV chromatogram of nickel affinity purification (A), heparin affinity (B) and MonoQ anion exchange (B). Green trace indicates concentration of NaCl; red lines and numbers indicate fractions. (b) SDS-PAGE analysis of fractions from (a). A1 fraction is the elution from nickel affinity chromatography. Protein ladder is labelled L and fraction numbers for each purification step are indicated.

4.2.2 Crystallisation of ALKBH5₆₆₋₂₉₂

Crystallisation trials for ALKBH5₆₆₋₂₉₂ were set up in the presence of the generic 2OG oxygenase inhibitor, *N*-oxalylglycine (NOG) (see Chapter 2), using commercial broad screens (JCSG+ and PACT Premier (Molecular Dimensions)).

Needle-shaped crystals with approximate dimensions of ~30 µm long and diameter < 5µm appeared after 3 weeks in JCSG+ well A12 of the JCSG+ screen (0.2 M potassium nitrate, 20% PEG 3350) (Figure 4 - 3a). However, these crystals did not diffract in-house.

Crystallisation optimisation of ALKBH5₆₆₋₂₉₂ was subsequently set up around the conditions of JCSG+ A12, in the presence of various generic inhibitors of 2OG oxygenases (NOG, 2,4-pyridine dicarboxylate (2,4-PDCA), and 5-carboxy-8-hydroxyquinoline (IOX1)) and the FTO and HIF prolyl-hydroxylase PHD2 inhibitor, (1-chloro-4-hydroxyisoquinoline-3-carbonyl)glycine (IOX3/FG-2216) (18,19). Needle-like crystals that were bigger than the initial hits appeared after a month in crystal drops with 2,4-PDCA added (Figure 4- 3b). After 3 months, a crystal measuring 20 µm in diameter and 500 µm in length appeared in the conditions that contained 2,4-PDCA (Figure 4 -3c). Crystals of a different morphology (~100 µm in the longest dimension) appeared in the conditions that contained IOX3 (Figure 4 - 3d).

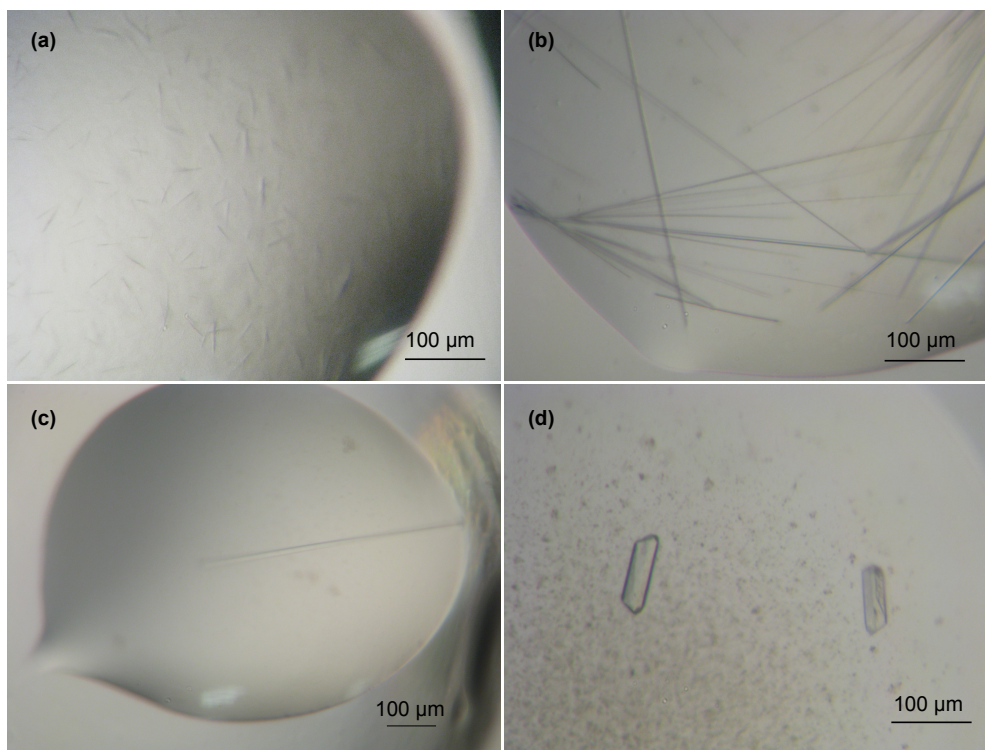


Figure 4 - 3. Images of crystals of ALKBH5₆₆₋₂₉₂. (a) Small needle-shaped crystals from the initial hit from JCSG+ A12, grown in the presence of NOG. (b) Optimised crystals of ALKBH5₆₆₋₂₉₂ grown in the presence of 2,4-PDCA. (c) The longest ALKBH5₆₆₋₂₉₂ crystal found in this study was grown in the presence of 2,4-PDCA. This crystal diffracted to 1.4 Å at the synchrotron. (d) ALKBH5₆₆₋₂₉₂ crystals grown in the presence of IOX3 that appeared after 3 months. Note the different morphologies between ALKBH5₆₆₋₂₉₂ crystals grown in the presence of IOX3 and 2,4-PDCA.

4.2.3 Data Collection and Processing

ALKBH5-2,4-PDCA

A crystal of ALKBH5 grown in the presence of 2,4-PDCA (Figure 4 - 3c) diffracted to 1.4 Å resolution at the synchrotron. Data were collected at 100 K using a helical strategy, i.e. by rotating the needle-shaped crystal (Figure 4 - 3c) while translating along the length of the crystal, to reduce localised radiation damage. The space group was determined to be $P2_12_12_1$ with unit cell constants $a = 38.92$ Å, $b = 74.53$ Å, $c = 111.37$ Å, $\alpha = \beta = \gamma = 90^\circ$. The calculated Matthews coefficient ($V_m = 2.72$ Å³) with a solvent content of 54.74% estimated 1 molecule per

asymmetric unit (20-22). However, the structure could not be solved by molecular replacement using an ensemble of NAOX structures (ALKBH2, ALKBH3, ALKBH8 DSBH domain and FTO DSBH domain) as a search model.

Using a structure of ALKBH5-IOX3 (described in Section 4.2.4) as a search model for molecular replacement on the ALKBH5-2,4-PDCA dataset also did not produce any correct solutions. Subsequently, the data were reprocessed in lower symmetry space groups (i.e. $P2$ and $P1$) and molecular replacement was reattempted (see Table 4 - 1 for data collection statistics processed in various space groups). However, these strategies did not result in a correct molecular replacement solution. Phenix.Xtriage (23) analysis of the data processed in the $P2$ space group identified a possible pseudo-merohedral twin operator ($h, -k, -l$), with an estimated twin fraction of 0.478 based on a maximum likelihood method. Therefore, potential twinning may explain the difficulty in obtaining a molecular replacement solution despite using a structure of the same protein as a search model. Twinning may also cause the space group to appear to have a higher symmetry (i.e. $P2\ 2\ 2$).

Table 4 - 1. Data collection statistics of an ALKBH5 crystal grown in the presence of 2,4-PDCA

Protein	ALKBH5-2,4-PDCA		
Radiation Source	Diamond Light Source Beamline I24		
Detector	Pilatus3-6M		
X-ray Wavelength (Å)	0.9686		
Temperature (K)	100		
Resolution Range (Å)	50.00 – 1.40 (1.45 – 1.40) [‡]	50.00 – 1.48 (1.53 – 1.48) [‡]	50.00 – 1.42 (1.47 – 1.42) [‡]
Space Group	<i>P</i> 2 2 2	<i>P</i> 2	<i>P</i> 1
Unit Cell Dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	38.92, 74.53, 111.37	38.93, 74.54, 111.35	38.93, 74.55, 111.33
α , β , γ (°)	90.00, 90.00, 90.00	90.00, 90.06, 90.00	89.99, 89.94, 90.04
Total Number of Reflections Observed	684601	653452	758655
Number of Unique Reflections	64323	106207	225280
Multiplicity	10.6 (3.9) [‡]	6.2 (3.7) [‡]	3.4 (2.6) [‡]
Completeness (%)	98.3 (86.1) [‡]	99.7 (97.3) [‡]	96.3 (91.8) [‡]
<i>I</i> / σ (<i>I</i>)	19.4 (2.0) [‡]	14.5 (2.3) [‡]	13.3 (2.0) [‡]
R_{merge}	13.4	12.7	8.9
Wilson B Factor (Å ²)	17.00	16.86	12.86
Possible Twin Operators (Number)	No twinning suspected	Pseudo-merohedral (1)	Pseudo-merohedral (3)
Software for Data Reduction	HKL-3000	HKL-3000	HKL-3000

[‡] outermost shell

$$\S R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle \times 100$$

ALKBH5-IOX3

ALKBH5 crystals grown in the presence of IOX3 diffracted to 2.4 Å using the in-house source. At the synchrotron, a crystal diffracted to 2.0 Å. The space group

was determined to be $P2_1 2_1 2_1$ with the unit cell constants $a = 67.1 \text{ \AA}$, $b = 82.7 \text{ \AA}$, $c = 89.2 \text{ \AA}$; $\alpha = \beta = \gamma = 90^\circ$. The calculated Matthews coefficient ($V_m = 2.2 \text{ \AA}^3$) with a solvent content of 43% indicated two molecules per asymmetric unit (Figure 4 - 4) (20-22).

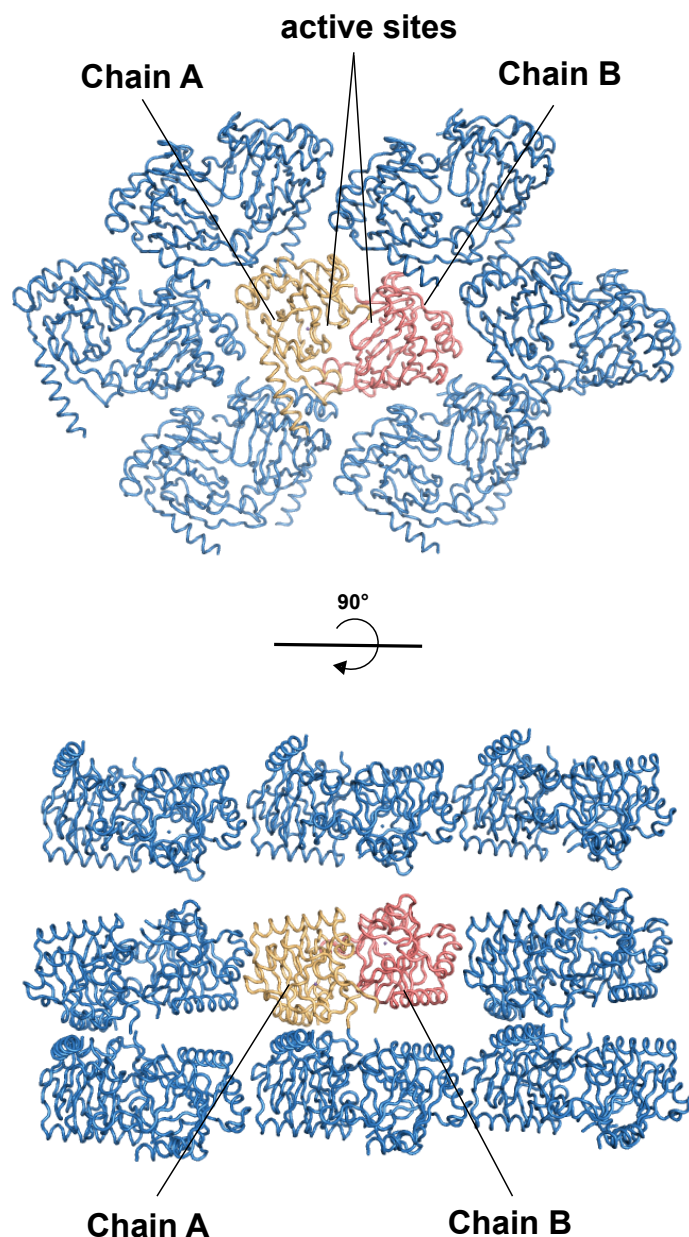


Figure 4 - 4. Crystal packing of ALKBH5₆₆₋₂₉₂ protein molecules in two orientations rotated at 90° relative to each other. There are two ALKBH5 molecules in the asymmetric unit. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

4.2.4 Structure Solution

A structure of ALKBH5 crystallised in the presence of IOX3 was solved by molecular replacement using PHASER (25) using an ensemble (25) of ALKBH2, ALKBH3, ALKBH8 and FTO as a search model. A solution was found and it gave a translation function Z score of 8.8. The initial model output from PHASER (25) was based on FTO. AUTOBUILD (26) was then used to generate a model of ALKBH5 with R and R_{free} values of 0.239 and 0.271. The structure was subsequently refined to R and R_{free} values of 0.165 and 0.218. (See Table 4 - 2 for data collection and refinement statistics.)

Table 4 - 2. Crystallographic data collection and refinement statistics.

Structure	ALKBH5₆₆₋₂₉₂-IOX3
PDB ID	4NJ4
Data Collection	
Radiation Source	Synchrotron Diamond Beamline I24
Detector	Pilatus3-6M
X-ray Wavelength (Å)	0.9789
Temperature (K)	100
Resolution Range (Å)	50.00 - 2.02 (2.09 – 2.02) [‡]
Space Group	<i>P</i> 2 ₁ 2 ₁
Unit Cell Dimensions	
a, b, c (Å)	67.1, 82.7, 89.2
α, β, γ (°)	90, 90, 90
Total Number of Reflections Observed	199163
Number of Unique Reflections	33274
Multiplicity	6.0 (5.6) [‡]
Completeness (%)	100.0 (99.9) [‡]
I/σ(I)	5.9 (2.0) [‡]
[§]R_{merge}	0.142
Wilson B Factor (Å²)	30.22
Refinement	
* R_{work} (%)	16.5
† R_{free} (%)	21.8
RMS Deviation from Ideality	
bonds	0.010 Å
angles	1.190°
Ramachandran Outliers (%)	0
Average B Factor (Å²)	40.53
Number of Water Molecules	224

$$^{\S} R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle \times 100$$

$$^* R_{\text{work}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / |F_{\text{obs}}| \times 100$$

† R_{free}, based on 6% of the total reflections

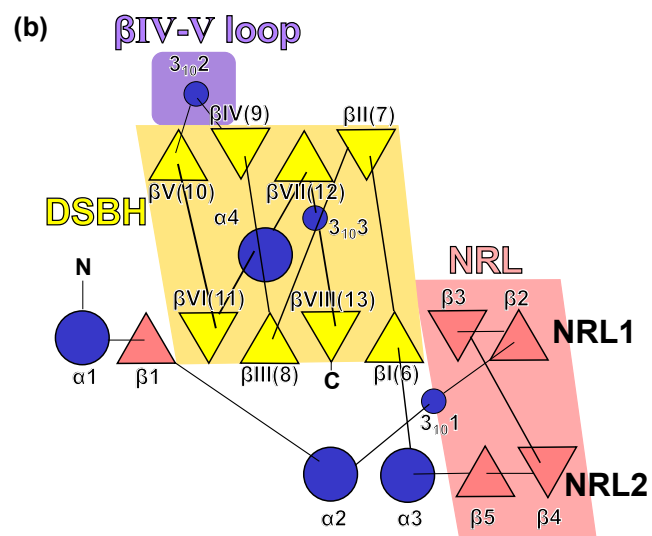
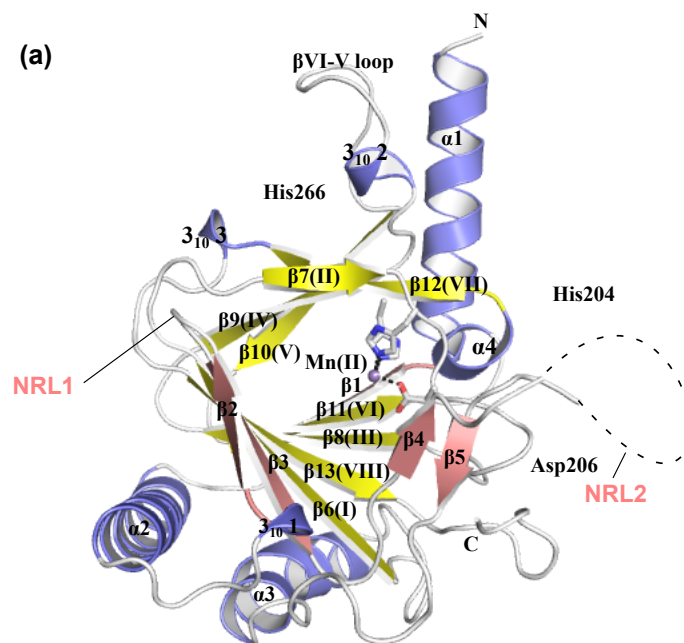
‡ outermost shell

Table is reprinted (adapted) from Aik *et al.* 'Structure of human RNA N⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press.

4.2.5 Overall ALKBH5 Structure

The refined ALKBH5₆₆₋₂₉₂ structure has electron density observed for 219 residues of Chain A (residues 66-68 and 145-149 were not observed) and 206 residues of Chain B (residues 66-77 and 143-151 were not observed). The structures of Chain A and B are similar with a calculated root-mean-square deviation of 0.31 Å for the 206 C_α atoms.

As observed for other 2OG oxygenases (27,28), ALKBH5 contains a conserved double-stranded beta helix (DSBH) core fold, which is formed by 8 anti-parallel beta-strands βI-VIII (β6-13) arranged into two β-sheets. The major β-sheet consists of strands β6, 8, 11 and 13 and the minor β-sheet consists of strands β7, 9, 10, 12. Three extra β-strands – β1, β2 and β3 – extend the major β-sheet. The DSBH core is partially wrapped around by 3 α-helices α1, α2 and α3 and they pack against the major β-sheet (Figure 4 - 5a, b). The core of the DSBH contains the conserved Fe(II)-coordinating residues His204, Asp206 and His266, also commonly known as the HXD/E...H facial triad. Mn(II) was observed bound in the Fe(II) binding site (substituting for Fe(II)), coordinated by the facial triad (Figure 4 - 5c). The 2OG binding pocket is positioned in a cavity between the two β-sheets of the DSBH with the more open end of the cavity apparently providing substrate access to the active site. IOX3, an inhibitor added to the crystallisation drop, was not observed in the active site. The active site openings of both Chains A and B are buried at the interface between the two chains in the asymmetric unit (Figure 4 - 7b).



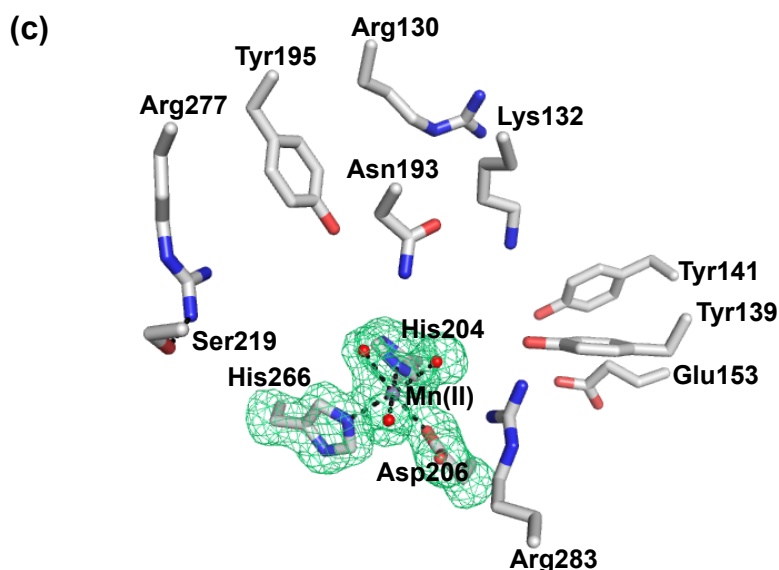


Figure 4 - 5. Views from a crystal structure of ALKBH5 (resolution: 2.02 Å). (a) Ribbons representation of the ALKBH5₆₆₋₂₉₂ structure showing the active site metal Mn(II) (purple sphere) substituting for Fe(II); residues His204, Asp206 and His266 (white sticks); nucleotide recognition lid NRL1 and NRL2 (pink fonts); disordered residues (dashed line); the DSBH β -strands I to VIII (yellow); other β -strands (salmon); and α - and 3_{10} -helices (blue). (b) Topology of the structure of ALKBH5₆₆₋₂₉₂. β -strands are shown as triangles, α -helices as big circles, 3_{10} -helices as small circles. The β IV-V loop is highlighted in purple, the DSBH in yellow, and the NRLs in salmon. (c) Active site residues of ALKBH5₆₆₋₂₉₂ with representative electron density (3.0σ mF_o-DF_c OMIT; green mesh) for side chains of His204, Asp 206, His266 and water molecules (red spheres) all coordinated (black dashed lines) to the active site Mn(II) ion (purple sphere). Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

4.2.6 N-Terminal Region of ALKBH5

The N-terminal region of the ALKBH5₆₆₋₂₉₂ structure begins with a long α -helix. The first 65 residues of the protein are missing in this construct. Secondary structure prediction (29) of the missing N-terminal region resulted in two α -helices. These helices may form a helical bundle with the α 1-helix. The α 1-helix is followed by the β 1 strand that is positioned at the 'back' of the DSBH relative to the active site opening, and is anti-parallel to β VI strand that is part of the

DSBH core. A couple of α -helices and loops were observed following β 1 before the N-terminal extension reaches two β hairpin-like loops that form the nucleic acid recognition lid (NRL).

4.2.7 C-Terminal Region of ALKBH5

The C-terminal region of the ALKBH5 structure consists of a short loop following the end of DSBH β VIII. The protein construct is missing 102 residues that were predicted (29) to be largely disordered. This region contains numerous arginine- and serine-residues which might play a role in RNA binding, as Arg-Ser rich regions of SR protein is known to be involved in RNA binding (30).

4.2.8 Beta IV-V loop

A loop linking strands β IV and β V (residues 229-242) of ALKBH5, which contains a 3_{10} helix ($3_{10}2$), extends from the DSBH and forms a 'finger' close to the opening of the active site (Figure 4 - 6). A β IV-V loop is commonly involved in substrate binding in some other 2OG oxygenases (28,31). This loop appears to be constrained by the interaction with the α 1 helix and a disulphide bond between Cys230 of β IV-V loop and Cys267 of DSBH strand β VII. Adjacent to Cys267, a cysteine-residue (Cys227) presents a surface exposed thiol, which raises the possibility of a thiol/disulphide 'redox shuffle', forming a Cys227-Cys267 disulphide and in turn relaxing the β IV-V loop conformation (Figure 4 - 6).

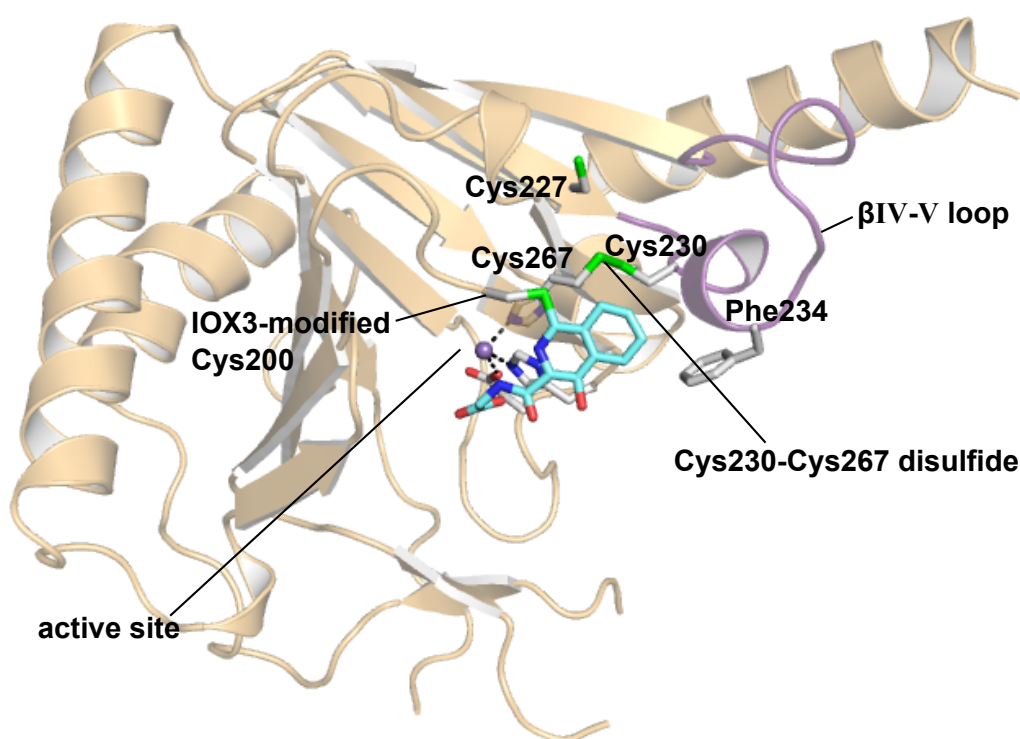


Figure 4 - 6. Surface cysteine-residues of ALKBH5. Cys230 and Cys267 form a disulphide bond. Cys200 is modified by IOX3. Reduction of the Cys230-Cys267 disulphide bond might result in a more flexible β IV-V loop for induced-fit substrate binding. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

4.2.9 Cys200-IOX3 modification

During structure refinement of the ALKBH5₆₆₋₂₉₂ structure, some electron density that did not correspond to the protein side chain or main chain was observed outside the active sites of both Chains A and B, at the interface of the two molecules in the asymmetric unit. The shape of this electron density suggested that it was derived from the ALKBH5 inhibitor IOX3, which was present in the crystallisation condition. Subsequent refinement led to the assignment of IOX3 that is covalently linked to Cys200 side chain thiol with the loss of chlorine (Figure 4 - 7a). Interestingly, two molecules of the covalently

linked IOX3, each attached to a different ALKBH5 molecule at Cys200, are stacked against each other through their aromatic bicyclic rings, presumably aiding crystal packing and formation (Figure 4 - 7b). This observation would explain the different crystal form and unit cell constants of the ALKBH5-IOX3 crystals in contrast to the needle-like crystals that formed in the presence of 2,4-PDCA or NOG.

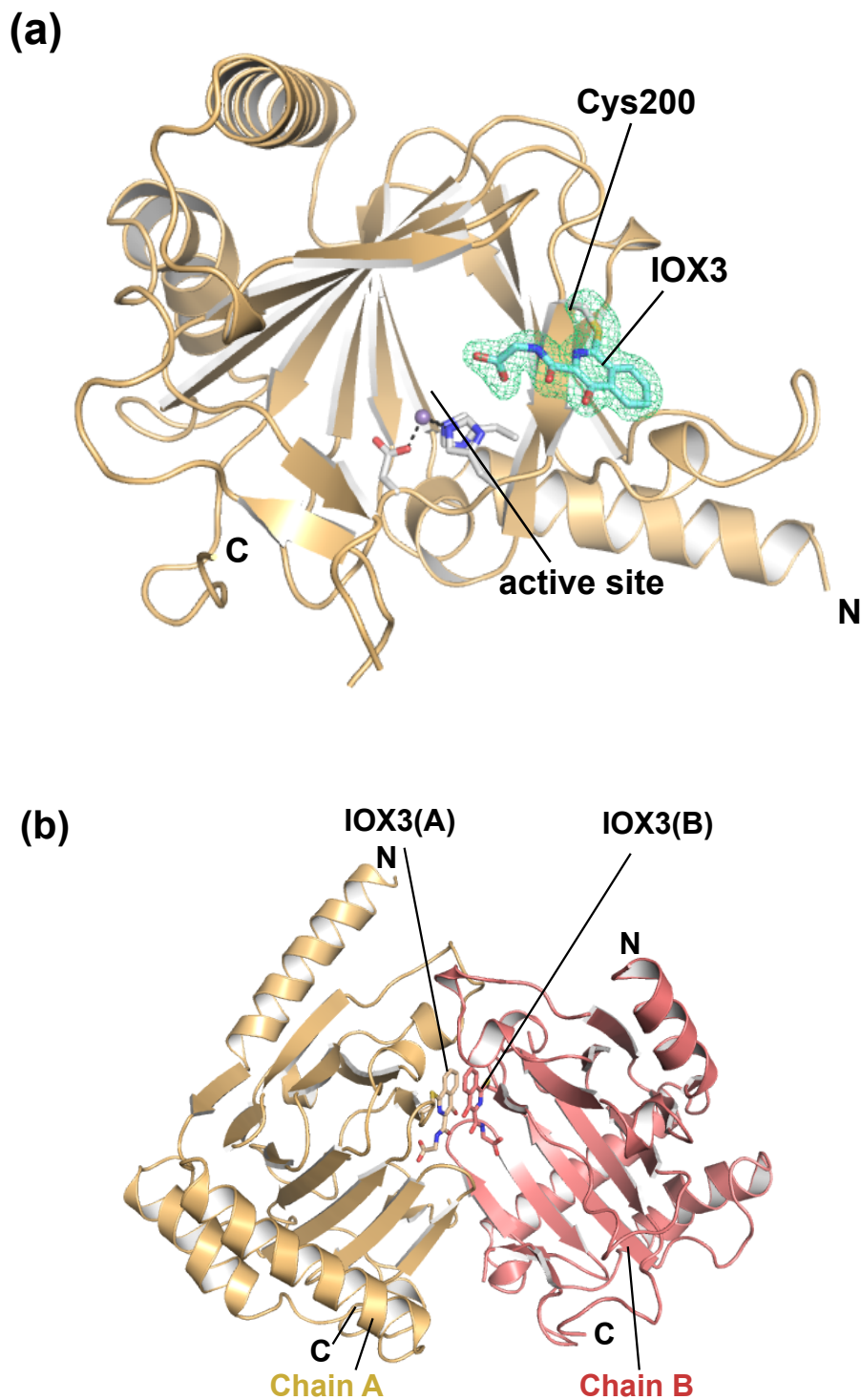


Figure 4 - 7. Binding of 1-chloro-4-hydroxyisoquinoline-3-carbonyl)glycine (IOX3) to ALKBH5 involves covalent attachment. (a) The small molecule IOX3 (cyan sticks) reacts and forms a covalent bond with the side chain of Cys200 (white sticks); the electron density map (3.0σ mF_o-DF_c OMIT; green mesh) is shown. (b) Two protein molecules in an asymmetric unit of an ALKBH5₆₆₋₂₉₂ crystal. Covalently attached IOX3 molecules from each protein molecule stack against each other via π - π interactions. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

To further confirm the crystallographically observed covalent modification on Cys200, fragmentation mass spectrometry (MS/MS) was carried out on a solubilised and Glu-C digested ALKBH5 crystal. MS/MS analysis of a peptide containing Cys200 (Tyr195-Asp206) revealed various y and b ions corresponding to the masses of the fragmented peptide (Figure 4 - 8). One particular peak at $m/z = 850.30$ corresponds to the mass of a peptide encompassing Tyr195-(IOX3-modified Cys200) (Figure 4 - 8), supporting the crystallographically observed modification on Cys200. (Glu-C digestion and MS/MS were performed by John Scotti and Dr Lingzhi Gong.)

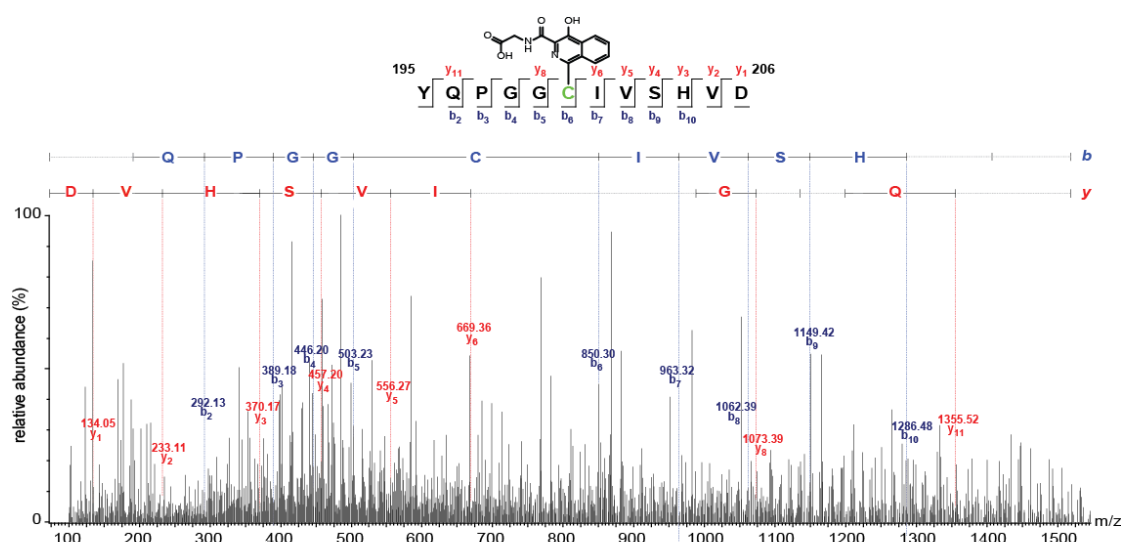


Figure 4 - 8. MS/MS spectrum of ALKBH5 crystal from a crystallisation drop that was incubated for 16 weeks. A peak at $m/z=850.30$ corresponding to a b ion of YQPGGC-IOX3 is observed. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press.

Notably, the position of the Cys200-attached IOX3 is too far from the active site metal for it to chelate the metal in a bidentate manner in the case of ALKBH5. IOX3 has been shown to occupy the 2OG binding sites in PHD2 and FTO and chelates the metal in a bidentate manner (18,32) (Figure 4 - 9).

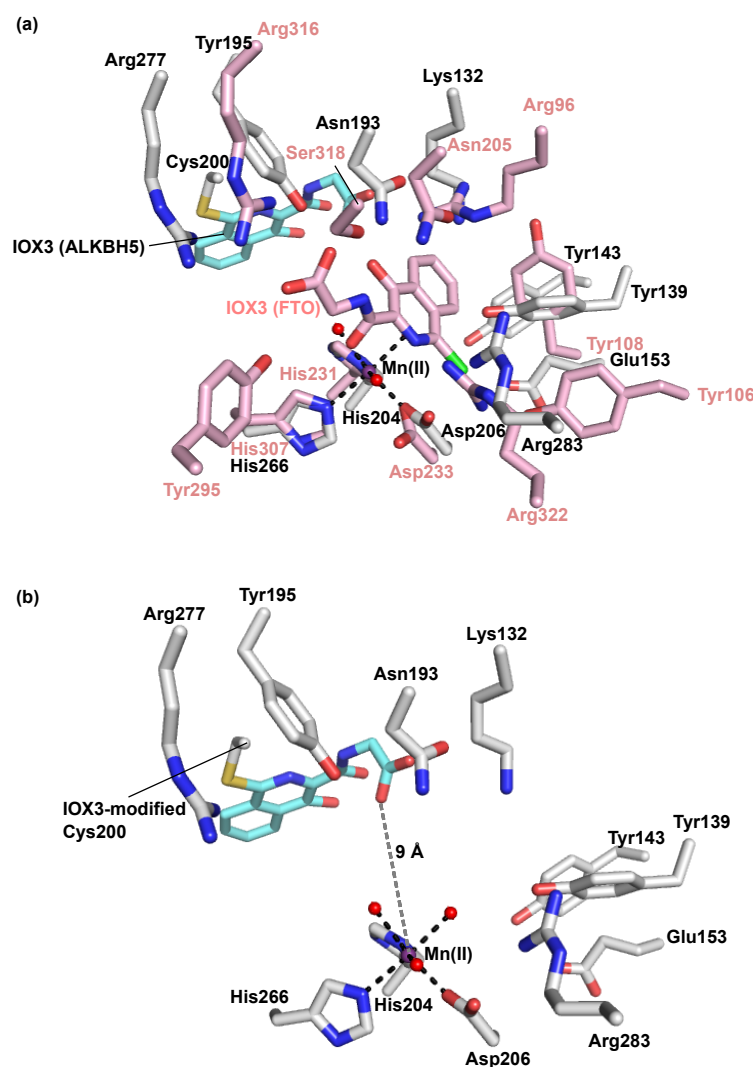


Figure 4 - 9. Position of IOX3-modified Cys200 of ALKBH5. (a) Superimposition of the FTO-IOX3 complex structure (residues and IOX3 coloured pink) (PDB ID 4IE6) with the ALKBH5 structure (residues coloured white, IOX3 coloured cyan). (b) Active site residues of ALKBH5 showing the position of the IOX3-modified Cys200. Note: in the crystallographically observed mode of IOX3, it is not possible for it to coordinate to the active site metal as observed for FTO (18). Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

The mechanisms of formation of the Cys200-IOX3 covalent adduct are proposed to proceed via aromatic nucleophilic substitution (S_NAr) (Figure 4 - 10), which occurs over the prolonged crystallisation process (up to 12 weeks). S_NAr of amino acids and residues occurs during derivitisation of the Sanger reagent, and has been proposed for the irreversible inhibition of thyroid hormone receptor TR β by methylsulfonylnitrobenzoates (33,34). Although the covalent modification of ALKBH5 by IOX3 is unlikely to be important in the timescale of inhibition work carried out for ALKBH5, it does raise the question as to whether IOX3, or related compounds, may undergo S_NAr type substitution in prolonged clinical application.

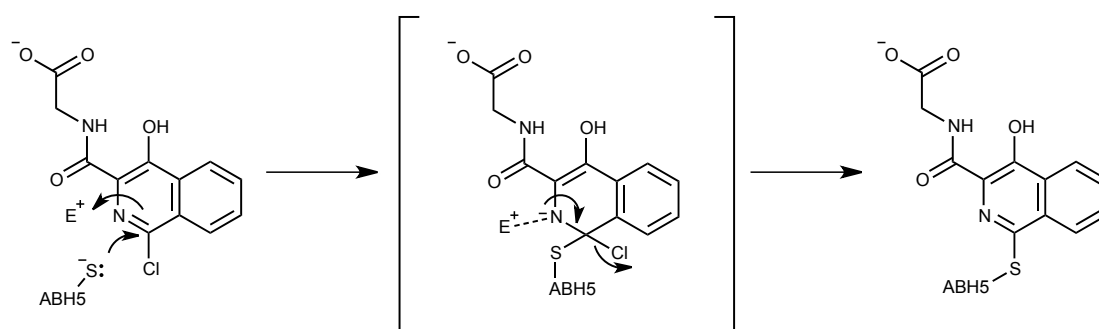


Figure 4 - 10. Proposed nucleophilic aromatic substitution mechanism for the crystallographically observed covalent modification of Cys200 by IOX3. Figure and legend are published in Aik et al. 'Structure of human RNA N^6 -methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by Oxford University Press.

4.2.10 Conserved catalytic triad and overall active site architecture

The active site of ALKBH5 (Figure 4 - 11a) contains an active site metal, which is octahedrally chelated by the conserved HXD...H motif (His204, Asp206, and His266) and three water molecules. Mn(II) was used as an Fe(II) substitute to produce a catalytically inactive complex. The active site of ALKBH5 also has the

conserved 2OG interacting residues: Arg277 (DSBH β VIII), Tyr195 (DSBH β I), Asn193 (DSBH β I) and Arg283 (DSBH β VIII) (Figure 4 – 11a). The 2OG binding pocket is lined by various hydrophobic residues (Ile281, Ile268, Leu226, Ile201, and Val191). The region where the prime substrate nucleotide is predicted to bind contains a mixture of aromatic, acidic and basic residues such as Tyr139, Tyr141, Glu153, Arg130 and Lys132 (Figure 4 – 11a).

4.2.11 2OG binding site

Studies with other 2OG oxygenases show that differences in the 2OG binding pocket can be exploited for the development of selective inhibitors (35-37). For example, hypoxia inducible factor (HIF) prolyl hydroxylase 2 (PHD2) is preferentially inhibited by *N*-oxalyl-*L*-alanine derivatives while HIF asparaginyl hydroxylase (FIH) and the histone demethylase Jumonji domain-containing protein 2A (JMJD2A) is inhibited by the *D*-enantiomers (35,37,38). These differences can be rationalised by structural differences in the 2OG binding pockets (35-37,39). The 2OG binding pocket of ALKBH5 includes residues Arg277 and Tyr195 (Figure 4 - 11a), which probably interact with the C5-carboxylate of 2OG as deduced by comparison of the ALKBH5 structure with those of NAOX-2OG complexes (Figure 4 - 11) (40-44). Interestingly, while FTO possesses the conserved arginine involved in 2OG C5-carboxylate binding as well as a tyrosine residue, the 2OG binding tyrosine (Tyr295) in FTO comes from DSBH β VI (Figure 4 - 11c), differentiating it from ALKBH5 and the other AlkB homologues where a 2OG binding tyrosine is derived from DSBH β I.

All of the AlkB-like NAOXs, including ALKBH5, also have a serine residue that interacts with the 2OG C5-carboxylate binding arginine (Arg277 for ALKBH5) (Figure 4 - 11). This serine is not present in the wybutosine tRNA hydroxylase TYW5, the 2OG binding pocket architecture of which is considerably different from those in the AlkB subfamily (40-45). Rather than belonging to the AlkB-like NAOX subfamily of 2OG oxygenases, TYW5 belongs to the JmjC oxygenase subfamily (45-47), most of which act as histone N^ε-methyllysine demethylases and lysyl- or arginyl-hydroxylases (48). Most of the JmjC oxygenases, including the NAOX TYW5, use a lysine (derived from DSBH βIV) and a tyrosine (derived from a β-strand involved in extending the major β-sheet positioned anti-parallel to DSBH βI) for 2OG C5-carboxylate binding (Figure 4 - 11g).

A conserved asparagine-residue (Asn193) that interacts with the C1-carboxylate oxygen of 2OG is present in the active site of ALKBH5 (Figure 4 - 11a). Arg283 is adjacent to the Fe(II) binding site in ALKBH5. An arginine at this position is conserved in all other structurally characterised NAOXs (Figure 4 - 11) (40-44). This conserved arginine is proposed to be involved in oxygen activation and its substitution has been shown to abrogate activity in the case of ALKBH3 (42).

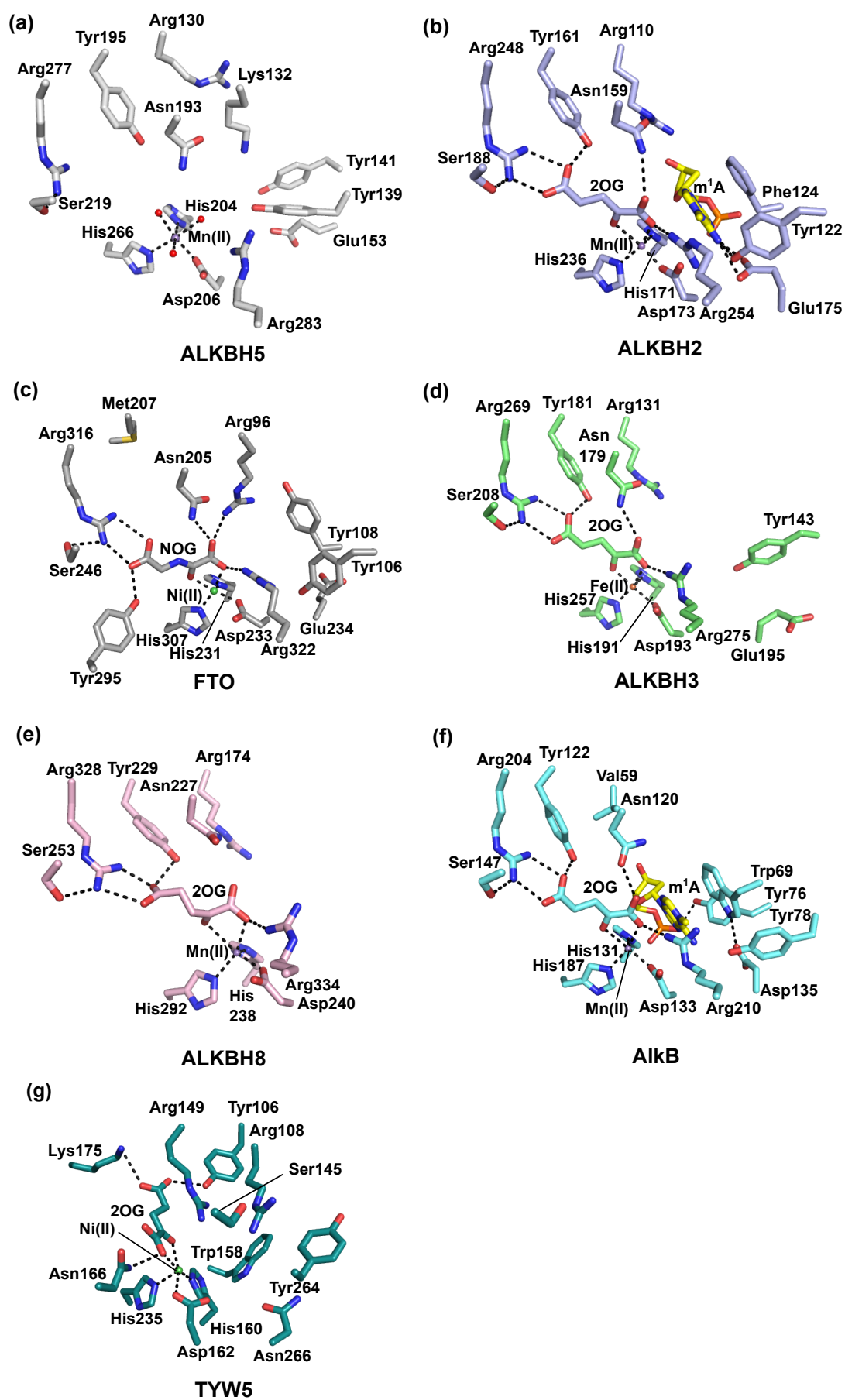


Figure 4 - 11. Comparison of the active site of ALKBH5 with those of other nucleic acid oxygenases. Active site residues of **(a)** ALKBH5 (white sticks) (PDB ID 4NJ4; resolution: 2.02 Å), **(b)** ALKBH2 (light blue sticks) (PDB ID 3BUC; resolution: 2.59 Å), **(c)** FTO (grey sticks) (PDB ID 4IDZ; resolution: 2.46 Å), **(d)** ALKBH3 (green sticks) (PDB ID 2IUW; resolution: 1.50 Å), **(e)** ALKBH8 (pink sticks) (PDB ID 3THT; 3.01 Å), **(f)** AlkB (cyan sticks) (PDB ID 3BIE; resolution: 1.68 Å) and **(g)** TYW5 (teal sticks) (PDB ID 3AL6; resolution: 2.80 Å). Oxygen (red), Nitrogen (blue), Phosphorous (orange), Sulphur (yellow), m⁶A base Carbon (yellow), Mn(II) (purple sphere), Fe(II) (orange sphere), Ni(II) (green sphere) and water molecule (red sphere) and electrostatic interaction (black dashed line) are indicated. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA N⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

4.2.12 Residues involved in substrate binding

Both identified m⁶A demethylases, ALKBH5 and FTO, have a basic residue (Lys132 and Arg96, respectively) in their prime substrate binding sites, which might be involved in substrate recognition through interactions with the nucleic acid substrate backbone phosphates. This basic residue could also play a role in product release following demethylation. Proteomic studies have shown that Lys132 of ALKBH5 can be acetylated following application of deacetylase inhibitors in MV4-11 cells, a human acute myeloid leukaemia cell line (49). Acetylation of Lys132 may alter substrate binding in ALKBH5 thus serving as a form of regulation. In FTO, the ALKBH5 Lys132 equivalent Arg96 has been shown to be important for its enzymatic activity. Arg96Met and Arg96Trp mutations almost completely abolish FTO enzymatic activity (40), thus supporting a crucial role for the conserved basic residue at this position in the m⁶A demethylases.

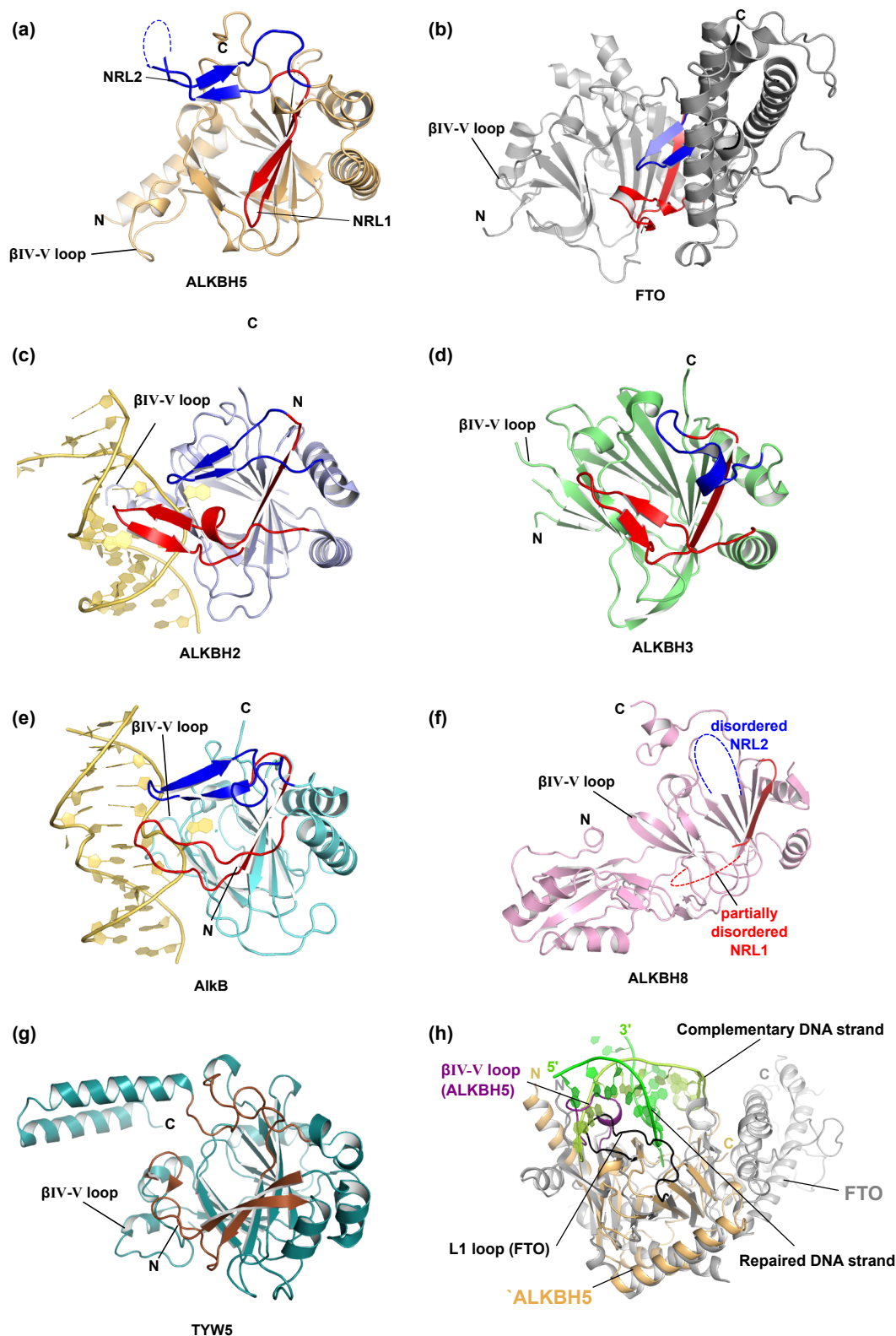
Furthermore, in ALKBH5 there is an acidic residue (Glu153) present, as is the case in most, possibly all, NAOXs (Glu175 in ALKBH2, Glu195 in ALKBH3, Glu234 in FTO, and Asp135 in AlkB) (Figure 4 - 11). In the case of FTO, Glu234 has been shown to be important in substrate recognition because a Glu234Pro substitution abolishes catalytic activity (40). Likewise, the AlkB Asp135Ala

mutant loses activity towards m¹A ssDNA; however, its activity towards m¹G increases (50).

Overall, the active site of ALKBH5 appears to be more open than that of FTO. This may be due to a combination of the longer nucleotide recognition lid (NRL) in FTO combined with its C-terminal domain, both acting to enclose the active site. Whether this apparent difference is reflective of the differences in m⁶A RNA substrate that are accepted by either FTO or ALKBH5 is unknown. It may be that ALKBH5 accepts RNA substrates that have a bulkier secondary structure. The difference in natural substrates of FTO and ALKBH5 will require further investigation.

4.2.13 Substrate Recognition Elements

All NAOX in the AlkB subfamily possess a nucleotide recognition lid (NRL) as part of the N-terminal extension of the DSBH (Figure 4 - 12a-f). The NRL is comprised of mixed beta and loop secondary structures that partially enclose the active site opening. Crystal structures of ALKBH2 and AlkB in complex with DNA substrates have revealed that the NRL plays an important role in substrate recognition (43,44).



stranded DNA (green). The β IV-V loop (purple) of ALKBH5 and L1 loop (black) of FTO overlap with the 'unrepaired' DNA strand (light green), potentially conferring single-strand selectivity for ALKBH5 and FTO. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

The majority of the NRL sequence (residues 159 – 173 and 182 – 191) of ALKBH8 was not observed, i.e. is likely to be disordered, in the ALKBH8 crystal structure, potentially indicating a requirement for flexibility in substrate binding (41). Similarly to the other AlkB subfamily NAOXs, ALKBH5 also possesses an NRL that can be divided into two β -hairpin-like structures: NRL1 (residues 124-137, strands β 2-3) and NRL2 (residues 138-161, strands β 4-5). NRL1 of ALKBH5 forms a β hairpin and a short type I β turn (residues 127-130). Structural alignment (Figure 4 - 13) of this NRL1 region with other AlkB-like NAOXs revealed that NRL1 of ALKBH5 is distinctly shorter than those in AlkB, ALKBH2, ALKBH3, and FTO by 9 to 15 residues. In the AlkB-dsDNA and ALKBH2-dsDNA complex structures, the NRL was shown to play a role in flipping the *N*-methylated DNA base into the active site for demethylation. The conserved Arg130 in NRL1 of ALKBH5 may interact with the substrate phosphate backbone as was observed for the equivalent residue in the ALKBH2-dsDNA complex (Arg110) (43). The ALKBH5 Arg130 equivalent Arg131Ala mutant in ALKBH3 has been shown to be inactive (42).

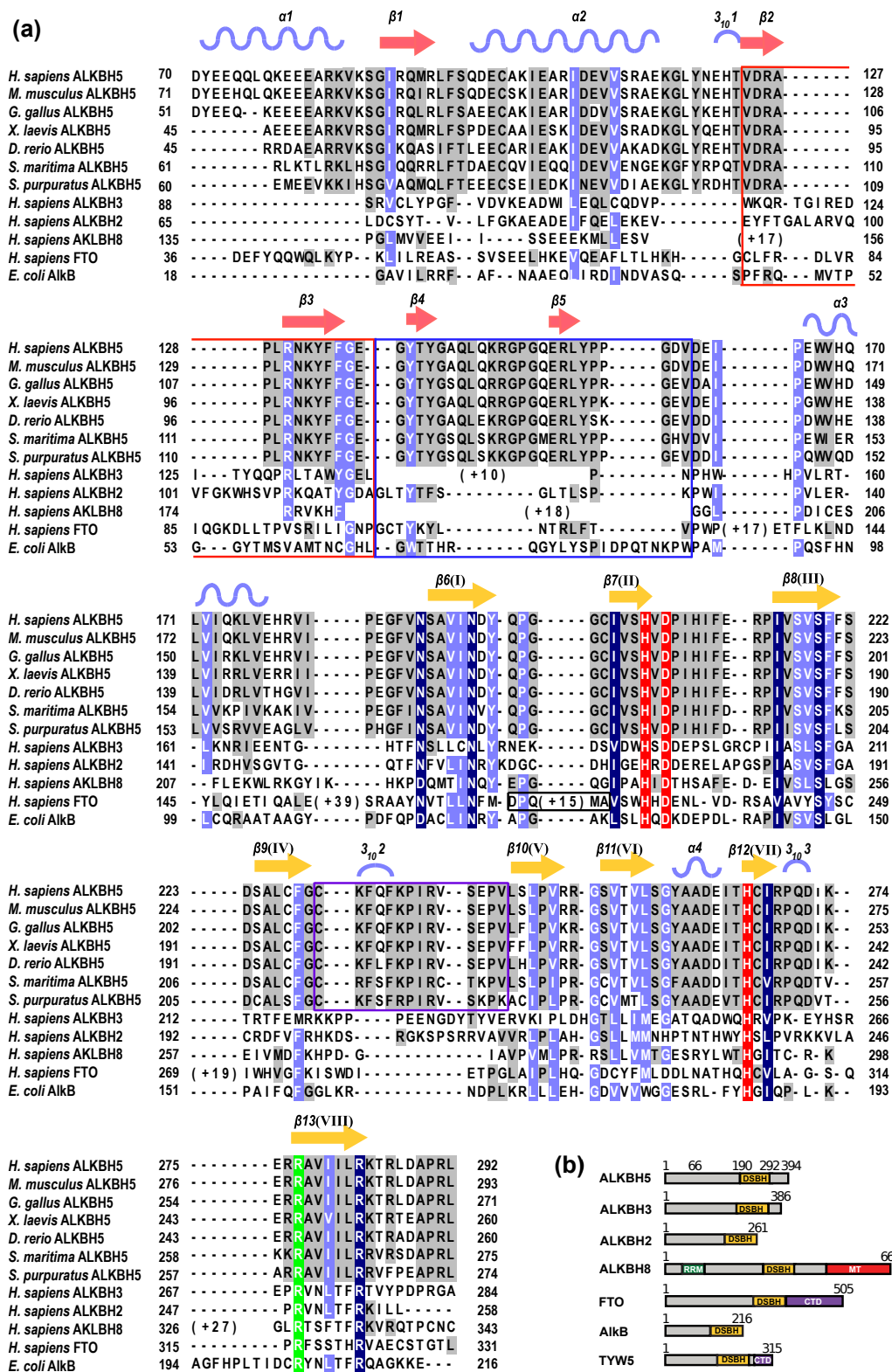


Figure 4 - 13. (a) A ClustalW2 (81) sequence alignment of ALKBH5 homologues from various organisms (from top: *Homo sapiens*: human (sequence identity 100%, PDB ID 4NJ4, Uniprot ID Q6P6C2); *Mus musculus*: mouse (sequence identity 97%, Uniprot ID Q3TSG4); *Gallus gallus*: chicken (sequence identity 86%, F1NIA5); *Xenopus laevis*: frog (sequence identity 78%, Uniprot ID Q6GPB5); *Danio rerio*: zebrafish (sequence identity 72%, Uniprot ID Q08BA6); *Strigamia maritima*: centipede (sequence identity 56%, Uniprot ID T1JJ71); and *Strongylocentrotus purpuratus*: purple sea urchin (sequence identity 52%, Uniprot ID H314D7)); combined with structure-based sequence alignment (82) with ALKBH3 (PDB ID

2IUW; resolution: 1.50 Å); ALKBH2 (PDB ID 3BUC; resolution: 2.59 Å); ALKBH8 (PDB ID 3THT; resolution: 3.01 Å); FTO (PDB ID 3LFM; resolution: 2.50 Å); and AlkB (PDB ID 3I3Q; resolution: 1.40 Å). A few selected manual adjustments were made to the alignment to correct for likely automated errors. Note: Variant residues from the reported ALKBH2 (PDB ID 3BUC; resolution: 2.59 Å) structure sequence were changed to reflect the wild type ALKBH2 sequence (Uniprot Q6NS38). Residues highlighted as: conserved (dark blue); semi-conserved (light blue); weakly conserved (grey); conserved 2OG oxygenase catalytic triad HxD...H, red; conserved 2OG binding arginine (green). Boxed residues indicate those forming: NRL1 (red); NRL2 (blue); the β IV-V loop of ALKBH5 (purple); the L1 loop of FTO (black). Secondary structural elements of *H. sapiens* ALKBH5 are represented as light blue sinusoidal waves (α -helices), red arrows (β -strands excluding the DSBH core), yellow arrows (β -strands of the DSBH core), and single light blue arcs (3_{10} -helix). **(b)** Schematic domain representations of human NAOXs for which structures have been reported. DSBH, double-stranded beta-helix core domain; RRM, RNA recognition motif; MT, methyltransferase domain; CTD, C-terminal domain. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press.

The NRL2 of ALKBH5 is notably longer by 12 residues than those of ALKBH2 and ALKBH3 (42,43). NRL2 of ALKBH5 forms a β -hairpin and the loop apex of the hairpin is partially disordered. The sequence of the apparently disordered apex of NRL2 contains two basic residues Lys147 and Arg148, both of which are conserved across ALKBH5 homologues in various organisms (Figure 4 - 13). These residues may be important for substrate recognition through interactions with the RNA substrate phosphate backbone. ALKBH5 shares a YXY/F motif positioned on the first strand of NRL2 (β 4) with ALKBH2, ALKBH3 and FTO. In AlkB, this motif (YLY) is positioned in its equivalent of the strand β 5 (Figure 4 - 13a). In the ALKBH2-dsDNA structure, the hydroxyl of the proximal Tyr122 interacts with the *N*⁶-atom of the m¹A substrate, while the distal-residue (i.e. the Y/F of the YXY/F motif) Phe124 of ALKBH2 π -stacks with the base (43). The distal Tyr108 of FTO is also observed to π -stack with an *N*³-methylthymidine nucleoside (3meT) base, and substitution of this residue abolishes its activity (40). This 'base-stacking' role is partially replaced by Trp69 in AlkB, aided by the proximal-residue Tyr76, which takes the place of the distal Y/F due to its position on the anti-parallel strand β 5. The equivalent Tyr139 and Tyr141 residues in ALKBH5 might play similar roles in contacting the nucleotide

base substrate. The position of the NRL2 varies in different NAOXs, possibly conferring specificity towards modified-base type and/or sequence context.

β IV-V loop

The β IV-V loop of ALKBH5 contains three basic residues (Lys231, Lys235 and Arg238) with the two lysines situated within a palindromic sequence KFQFK. These basic residues potentially play a role in contacting RNA substrate phosphate backbone considering the position of the β IV-V loop that is near the active site. The β IV-V loop and the neighbouring α 1-helix appear to form a positively charged groove, and this could be important for binding the negatively charged ssRNA phosphate backbone (Figure 4 - 14a, c). Superimposition of ALKBH5 with ALKBH2-dsDNA structures reveals that the β IV-V loop of ALKBH5 overlaps with the complementary 'unrepaired' strand of dsDNA, potentially conferring ssRNA selectivity. A similar ssRNA selectivity role has been proposed for a loop in FTO, the L1 loop, which is located between DSBH strands β I and β II (40). When ALKBH5 and FTO are superimposed, the L1 loop of FTO is situated in the same relative position near the active site as the β IV-V loop of ALKBH5 (Figure 4 - 12h). Almost the whole of the palindromic sequence KFQFK of ALKBH5 forms a 3_{10} helix that is absent in other ALKBHs. The 3_{10} helix contains a solvent exposed phenylalanine, Phe234, which points towards the active site. Observation of this orientation led to the speculation that Phe234 might act as a phenylalanine-finger to flip the m⁶A base into the active site by inserting itself in between flanking bases.

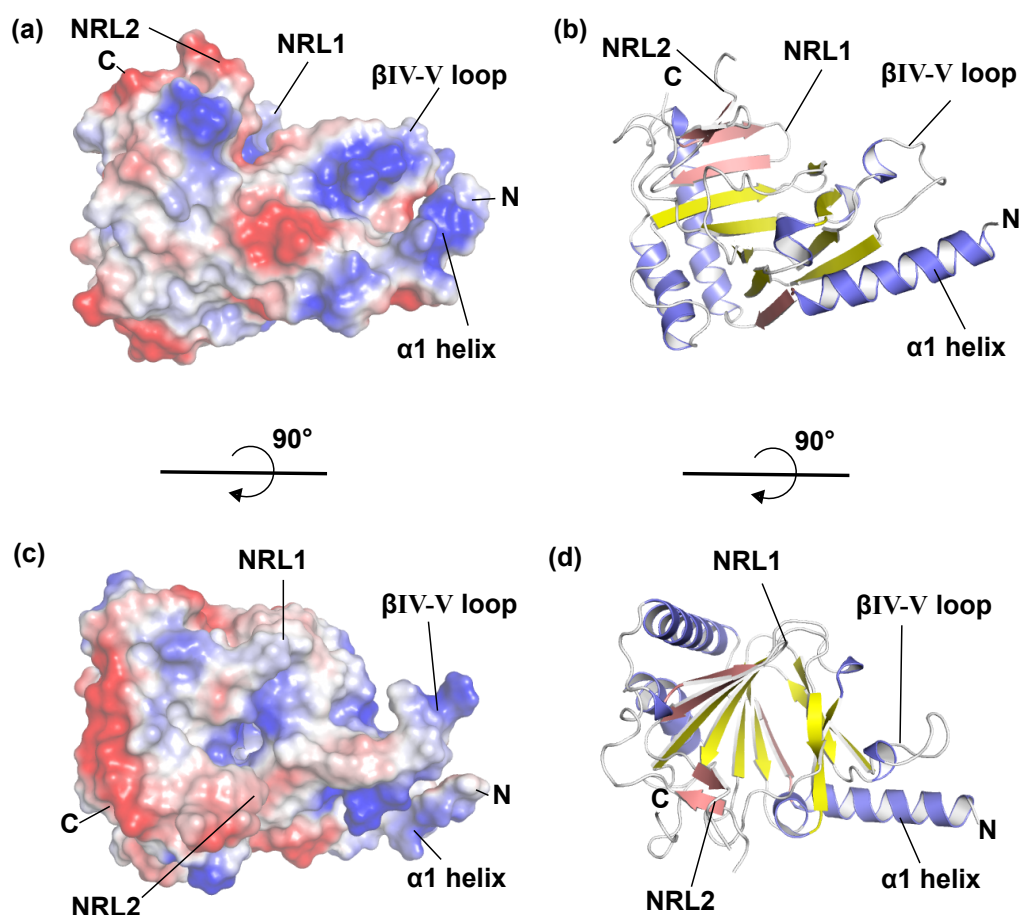


Figure 4 - 14. (a and c) ALKBH5 electrostatic surface representation (basic in blue; acidic in red) with 90° rotation along the x-axis and (b and d) corresponding ribbons representation. The substrate-binding groove around the active site is largely basic for the binding of the negatively charged ssRNA phosphate backbone. The basic region between the βIV-V loop and α1 helix forms a potential substrate binding groove. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press.

4.2.14 Modelling of substrate RNA binding

To investigate the proposed base-flipping mechanism in ALKBH5 using a phenylalanine finger, modelling of RNA substrate binding into ALKBH5 was performed by Dr Hwanho Choi. This mechanism was adopted by ALKBH2 where it uses a phenylalanine finger (Phe102) from its NRL1 to 'nudge' the targeted

base into the active site. Based on the ALKBH2-dsDNA structure, a 5-mer ssRNA substrate with the sequence 5'-GGAm⁶ACU-3' was docked into the active site of ALKBH5. The 5' – 3' direction of the strand through the active site was kept consistent with that observed in the ALKBH2-dsDNA complex (43). The docking was further restrained by preserving the predicted interaction between Arg130 and the phosphate backbone as well as Tyr141 with the phosphate backbone (43,51). The m⁶A substrate's N⁶-methylgroup-metal distance was set to ~4.1 Å, which is the average 'substrate carbon to enzyme metal distance' observed for most 2OG oxygenase-substrate structures (31). Subsequently, an energy minimisation was performed to investigate geometry and steric interactions. A similar docking study was also performed (by Dr Hwanho Choi) based on the AlkB-dsDNA complex structure whereby the base flipping mode involves flanking bases directly stacking against each other (51). The preliminary results suggest that the phenylalanine finger base-flipping mode involving Phe234 from the βIV-V loop to approach the bases flanking m⁶A (Figure 4 - 15) is more likely than the direct base stacking mode, which appeared strain by visual inspection. Although Phe234 appears relatively distant from the substrate binding groove, its position could change upon substrate binding and an induced-fit mechanism aided by reduction of the Cys230-Cys267 disulphide.

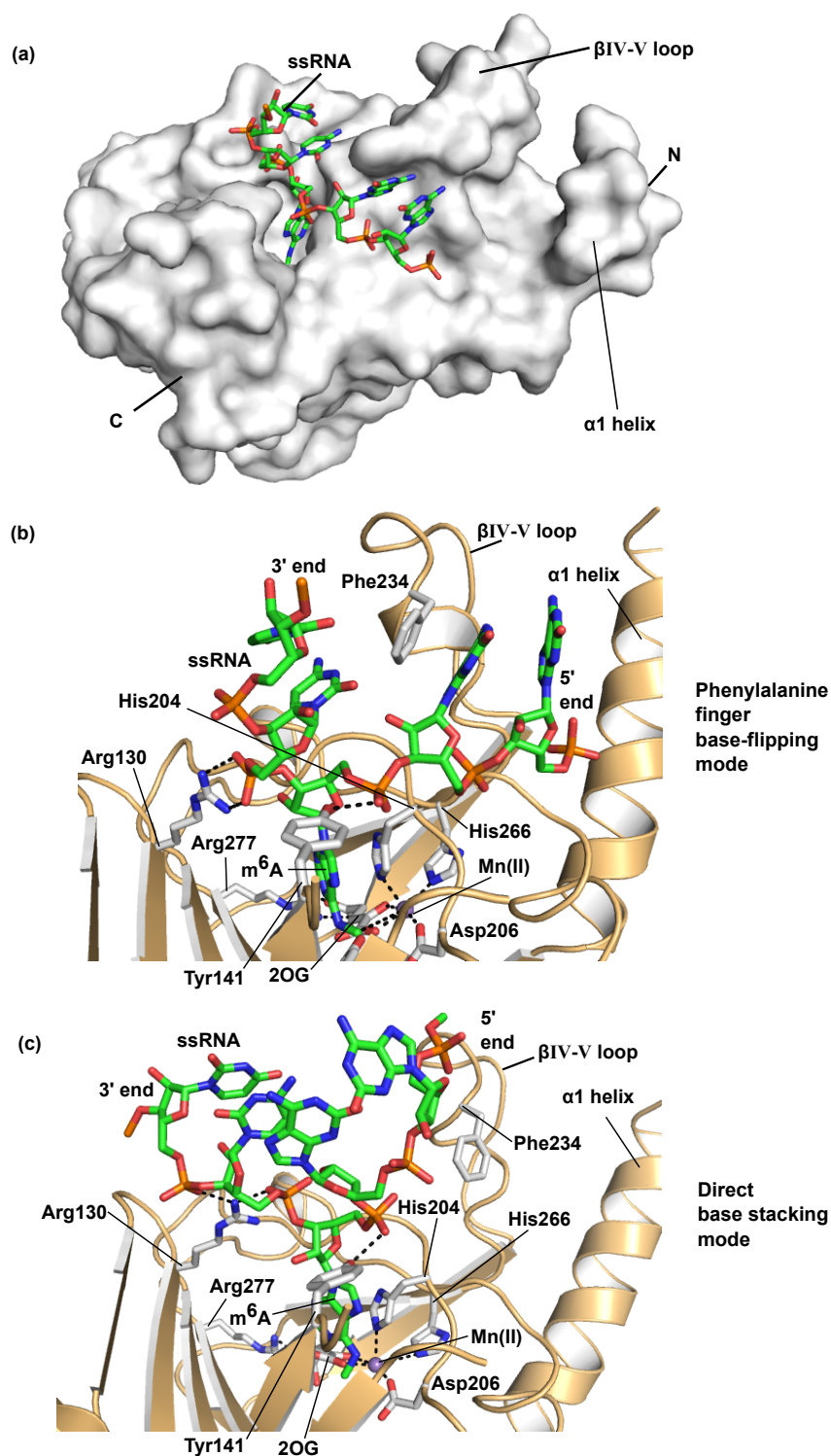


Figure 4 - 15. View of a model of substrate binding to the active site of ALKBH5. (a) surface representation and (b) cartoon representation of a structure of ALKBH5₆₆₋₂₉₂ with modelled 5-mer ssRNA 5'-GGm⁶ACU-3'. The structure of ALKBH2-dsDNA (PDB ID 3BUC; resolution: 2.59 Å) and its phenylalanine base-flipping mode were used as a template (43) for ALKBH5 substrate modelling. (c) cartoon representation of a structure of ALKBH5₆₆₋₂₉₂ with modelled 5-mer ssRNA 5'-GGm⁶ACU-3' in a direct base stacking mode. The structure of AlkB-trinucleotide (PDB ID 3I20; resolution: 1.70 Å) was used as a template (51) for ALKBH5 substrate modeling. The substrate was manually docked onto the ALKBH5 structure followed by energy minimisation. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA N⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press. Protein structure models were produced using MacPyMOL v1.3 (24).

4.2.15 *In vitro* activity of ALKBH5₆₆₋₂₉₂

The crystallised ALKBH5₆₆₋₂₉₂ construct has been shown to be catalytically active for uncoupled 2OG turnover (13). Following the report of a longer construct of ALKBH5 (residues 66-394) catalysing demethylation of m⁶A in a 15-mer ssRNA containing a 5'-GGm⁶ACU-3' motif (11), a 5-mer ssRNA with the sequence 5'-GGm⁶ACU-3' was used as a substrate in this study to test the activity of ALKBH5₆₆₋₂₉₂. The demethylation reaction was carried out with 4 µM ALKBH5₆₆₋₂₉₂ in the presence of 2OG, Fe(II), ascorbate, and 5-mer m⁶A modified ssRNA. The reaction mixture and a control without ALKBH5₆₆₋₂₉₂ were incubated at room temperature for 20 min. In the sample containing ALKBH5₆₆₋₂₉₂, MALDI analysis showed a peak at m/z = 1570, corresponding to a -14 Da shift of the mass of the m⁶A-modified ssRNA (mass = 1584). In the control sample without enzyme, the peak remained at 1584 Da (Figure 4 - 16). These results confirmed that the ALKBH5₆₆₋₂₉₂ construct is active.

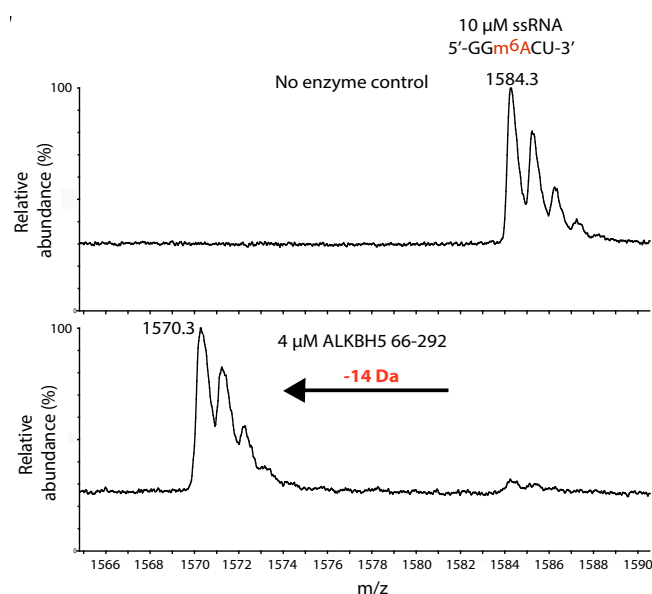


Figure 4 - 16. Activity assays of ALKBH5₆₆₋₂₉₂. MALDI spectra showing demethylation (-14 Da) of 10 µM 5-mer ssRNA (5'-GGm⁶ACU-3') (1584 Da) in the presence of 4 µM ALKBH5₆₆₋₂₉₂ after a 20 min reaction. Figure and figure legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA N⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press.

4.2.16 End-point inhibition assays

End-point inhibition assays were then carried out to investigate the inhibition of ALKBH5 by the generic 2OG oxygenase inhibitors NOG and 2,4-PDCA (52), and IOX3 that was originally developed as a prolyl-hydroxylase inhibitor (36). Relative to no inhibitor control, all three inhibitors inhibited the activity of ALKBH5. 2,4-PDCA was the most potent of the three inhibitors tested (9% residual activity), followed by IOX3 (40% residual activity) and NOG (44% residual activity) (Figure 4 - 17). In contrast, IOX3 was shown to be a more potent inhibitor of FTO than both NOG and 2,4-PDCA (see Chapter 3, Section 3.2.4) (18), thus suggesting that differences in the active site between FTO and ALKBH5 could potentially be further exploited for selectivity. Although IOX3 was observed to be covalently attached to Cys200 in the ALKBH5 crystal structure, this binding mode is most likely not to be relevant to the inhibition studies with purified ALKBH5 at least in a timescale of 5 min (see Section 4.2.19). Nevertheless, the observed reaction does raise the question as to whether *S*-arylation via nucleophilic aromatic substitution may occur during clinical application of C-1 chlorinated isoquinoline derivatives; at least one of such is in clinical trials for anaemia (36).

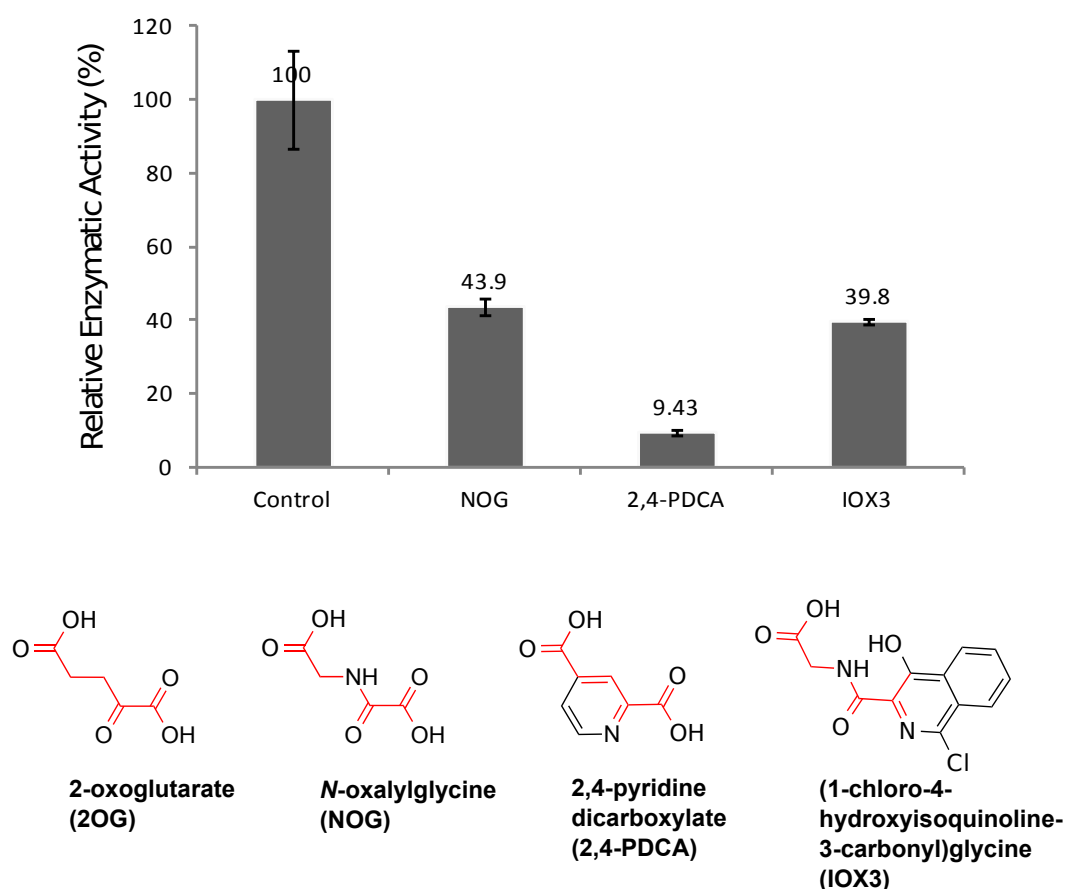


Figure 4 - 17. Inhibition assays of ALKBH5₆₆₋₂₉₂. Percentage residual activity of ALKBH5₆₆₋₂₉₂ relative to no inhibitor control after a 5 min reaction in the presence of 150 μ M N-oxalylglycine (NOG), 150 μ M 2,4-pyridine dicarboxylic acid (2,4-PDCA), and 150 μ M IOX3. Demethylation assays were conducted in triplicate and errors are standard deviations from the mean (n=3). Chemical structures of cosubstrate 2OG and the mimetic inhibitors used are shown. Portions of inhibitors isosteric to cosubstrate 2OG are highlighted as red bonds. Figure and legend are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press.

4.2.17 Development of ALKBH5₆₆₋₂₉₂ inhibition assays

Inhibition studies on ALKBH5 will aid in selective inhibitor development for the closely related FTO protein. Detailed inhibition studies of ALKBH5 can only be carried out with a sensitive inhibition assay, thus an assay for IC₅₀ determination

was optimised from the assay described above. (IC₅₀ assay development was carried out in collaboration with Dr Marina Demetriades.)

The limit of detection using MALDI/ToF mass spectrometry was first tested with 0.5 μ M, 0.7 μ M and 1 μ M ALKBH5₆₆₋₂₉₂ and 5 μ M, 7 μ M and 10 μ M m⁶A-modified RNA substrate. A signal suitable for quantification was found when 0.7 μ M enzyme and 5 μ M substrate were employed.

Kinetic studies of ALKBH5 were carried out by varying 2OG, ascorbate, Fe(II) and RNA substrate concentrations. In each case, the initial reaction rate was plotted against the concentration of each varying component (Figure 4 - 18). It was observed that the ALKBH5 reaction was not fully dependent on ascorbate, i.e. activity was observed without ascorbate (Figure 4 - 18c). This is in contrast with the related FTO and the human collagen prolyl-4-hydroxylase (CPH) tetramer, which showed no detectable *in vitro* enzymatic activity without ascorbate (see Chapter 3, Section 3.2.3 and Chapter 6, Section 6.2.2). Notably, the HIF hydroxylases, PHD2 and FIH, remain active *in vitro* in the absence of ascorbate albeit less active than when ascorbate is present (53). Therefore, the observation that ALKBH5 is not fully dependent on ascorbate for activity is not unprecedented.

For ALKBH5, the K_m values of 2OG and 5-mer RNA containing m⁶A were determined to be 1.29 $\pm$ 0.14 μ M and 0.89 $\pm$ 0.16 μ M, respectively. The concentration of Fe(II) cofactor for optimal activity was 100 μ M. For inhibition studies, 3 μ M 2OG, 3 μ M RNA, 100 μ M Fe(II), 1 mM ascorbate, and 0.7 μ M were used.

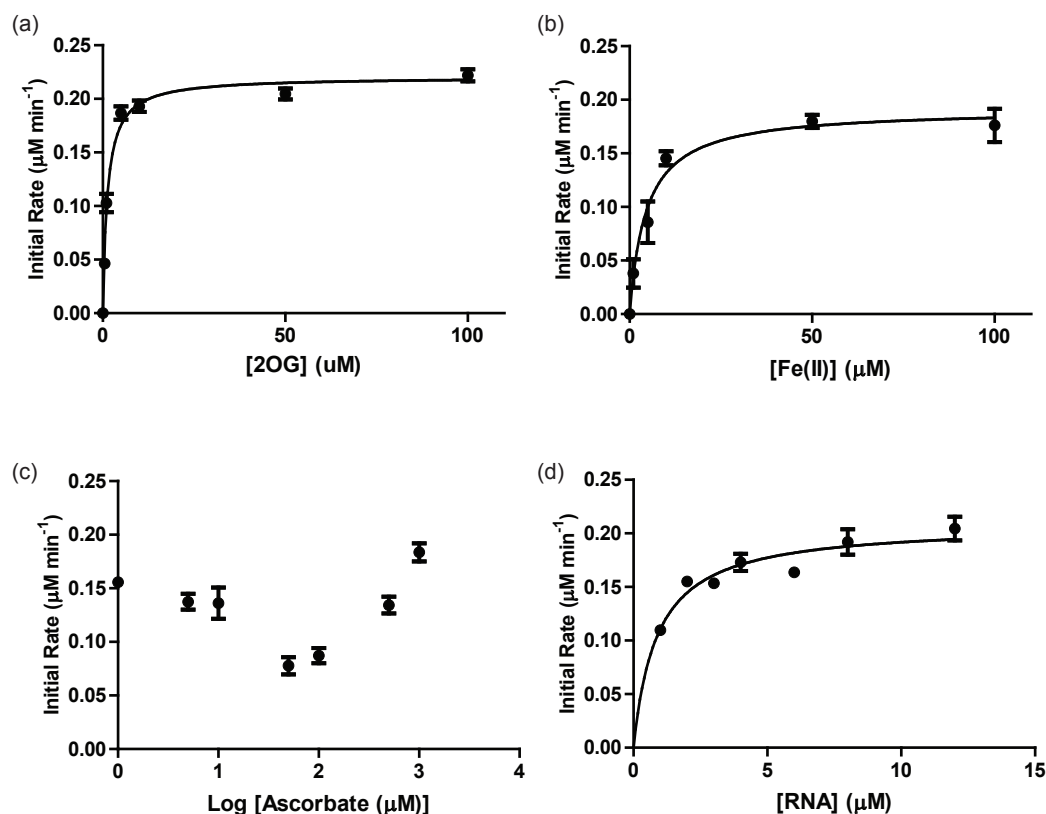


Figure 4 - 18. ALKBH5 kinetic studies. (a) Michaelis-Menten curve of initial rate (μM/min) versus concentration of 2-oxoglutarate (2OG) cosubstrate (μM). K_m of 2OG is $1.29 \pm 0.14 \mu\text{M}$. (b) Graph of initial rate (μM/s) versus concentration of Fe(II) (μM). (c) Data points of initial rate (μM/min) versus Log (concentration of ascorbate (μM)). The activity of ALKBH5 is not dependent on ascorbate concentration; ALKBH5 is still active without ascorbate. (d) Michaelis-Menten curve of initial rate (μM/min) versus concentration of 5-mer RNA (5'-GGm⁶ACU-3') prime substrate (μM). K_m of RNA substrate is $0.89 \pm 0.16 \mu\text{M}$.

Using the above assay, the IC_{50} values of NOG, 2,4-PDCA and IOX3 were determined to be 12.5 μM, 0.389 μM, and 17.8 μM respectively (Figure 4 - 19a-c), confirming results from earlier end-point inhibition assays showing that 2,4-PDCA was the most potent inhibitor of these three compounds (see Section 4.2.16). In contrast with ALKBH5, IOX3 inhibits PHD2 and FTO the most compared with NOG and 2,4-PDCA (Table 4 - 3) (18,54). A crystal structure of ALKBH5₇₄₋₂₉₄ obtained by members of the Min laboratory, a collaborator laboratory from the Structural Genomics Consortium in Toronto, contains a

citrate molecule in the active site (55). Using this optimised assay, citrate was found to be a very weak inhibitor of ALKBH5 with an IC₅₀ value of 488 µM (Figure 4 - 19d) (55), which is comparable with the IC₅₀ value of citrate (300 µM) when tested against FTO as an inhibitor (18). Citrate is also a very weak inhibitor of PHD2, having an IC₅₀ value of 1.8 mM (56).

Table 4 - 3. IC₅₀ values of NOG, 24-PDCA, IOX3 and citrate for ALKBH5, FTO, PHD2 and FIH.

	IC ₅₀ of NOG	IC ₅₀ of 2,4-PDCA	IC ₅₀ of IOX3	IC ₅₀ of citrate
ALKBH5	12.5 µM	0.389 µM	17.8 µM	488 µM
FTO	44 µM (18)	8.3 µM (18)	2.8 µM (18)	300 µM (18)
PHD2	0.8 µM (57)	6 µM (48)	0.3 µM (54)	1800 µM (56)

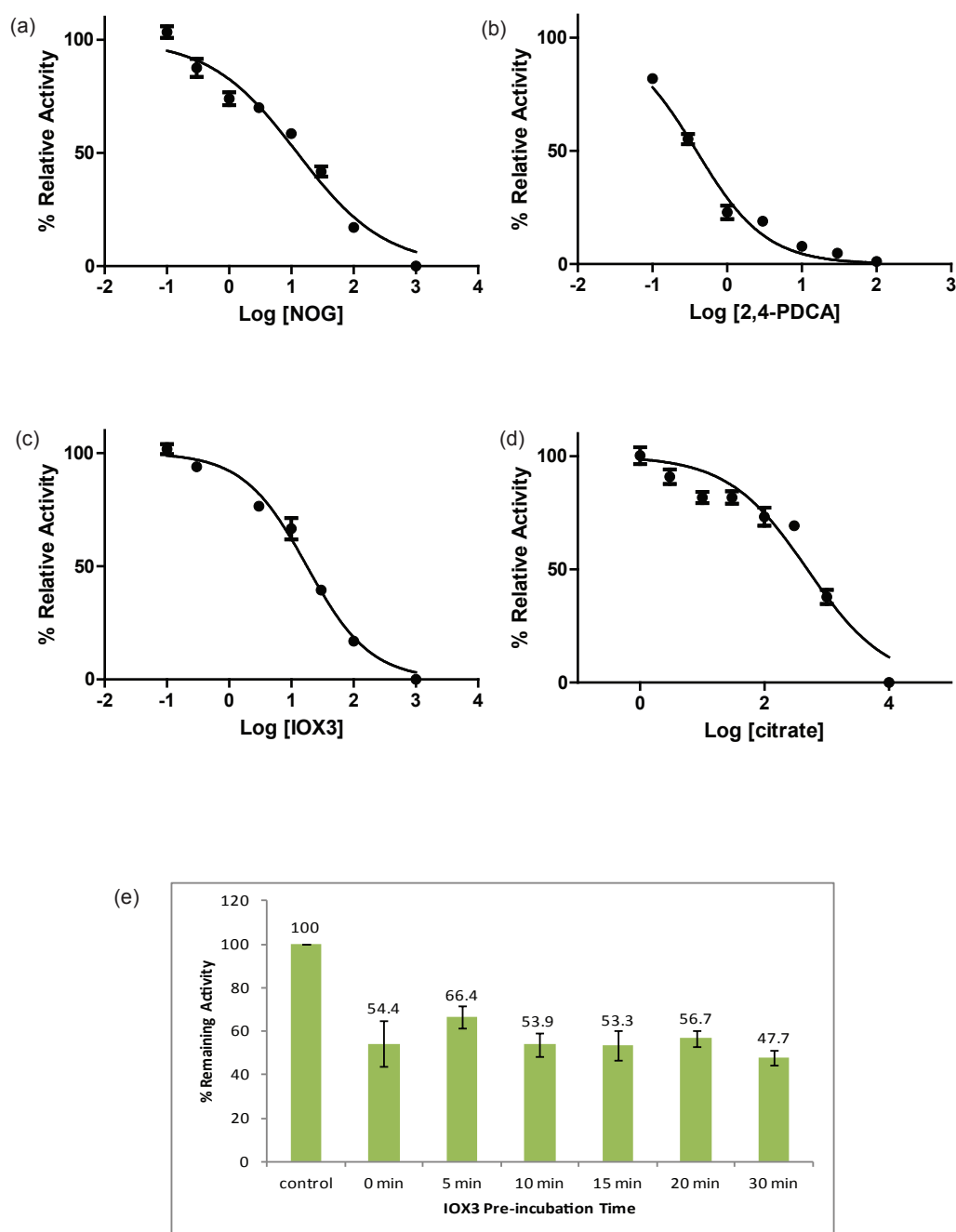


Figure 4 - 19. Inhibition studies on ALKBH5. Dose-response curves of the activity of ALKBH5 in the presence of **(a)** NOG ($IC_{50} = 12.5 \mu M$), **(b)** 2,4-PDCA ($IC_{50} = 0.389 \mu M$), **(c)** IOX3 ($IC_{50} = 17.8 \mu M$), and **(d)** citrate ($IC_{50} = 488 \mu M$). **(e)** Percentage relative activity of ALKBH5₆₆₋₂₉₂ with the enzyme pre-incubated with IOX3 for various durations (0 – 30 min). The final concentration of IOX3 during the reaction is 16 μM . Activity of ALKBH5₆₆₋₂₉₂ obtained is relative to no inhibitor control (control). Standard deviation is indicated by error bar ($n = 3$). Figure 4 -19d: This research was originally published in the Journal of Biological Chemistry. Xu, C., Liu, K., Tempel, W., Demetriades, M., Aik, W., Schofield, C.J. and Min, J. Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single stranded m6A RNA demethylation. *J. Biol. Chem.* 2014 (In press) © the American Society for Biochemistry and Molecular Biology.

4.2.18 IOX3 pre-incubation with ALKBH5

The crystallographic observation of the IOX3-modified Cys200 was presumed to have occurred over the prolonged crystallisation duration (see Section 4.2.9). To investigate whether IOX3-mediated inhibition of ALKBH5 involved modification in the timescale of our inhibition assay, an inhibition assay was conducted with IOX3 pre-incubated with ALKBH5 for different time durations (0 min, 5 min, 10 min, 15 min, 20 min, and 30 min). A control reaction without inhibitor was also set up. The results from these assays show that the relative activities of samples containing IOX3 are $50 \pm 10\%$ (with the exception of 5 min pre-incubation, 66%), and there was no clear pre-incubation time-dependent reduction in activity of ALKBH5 (Figure 4 - 19e). At least within 30 minutes, there was no sign of IOX3 modifying ALKBH5, assuming that any irreversible modification on Cys200 situated near the opening to the active site would affect RNA binding. Therefore, IOX3 most probably inhibits ALKBH5 by competing with 2OG, and possibly also with substrate but without exclusive irreversible covalent modification.

4.3 Conclusions

The determination of a structure of human ALKBH5 has provided insights into its recognition mechanisms of single-stranded nucleic acid. The structure also revealed the similarities and differences between ALKBH5 and its functionally similar cousin, FTO. FTO and ALKBH5 have similar Fe(II) chelating and 2OG-interacting residues apart from the positions of the 2OG-interacting tyrosine (Tyr295 for FTO; Tyr195 for ALKBH5). Further, there are also differences in the

positions of the YXY motif on the NRL2 of FTO and ALKBH5. These differences might allow for exploration of inhibitors that would selectively target each individual enzyme. In summary, the structural and biochemical work described in this chapter lays a foundation for a structure-based inhibition study of ALKBH5, which will be important for the development of selective inhibitors against potential clinical targets within the NAOX subfamily.

4.4 Experimental Procedures

Experimental procedures (except for Section 4.4.5 and 4.4.6) are reprinted (adapted) from Aik *et al.* 'Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press.

4.4.1 Protein expression and purification

A plasmid with the vector backbone pNIC28-Bsa4 encoding a hexahistidine-tagged ALKBH5₆₆₋₂₉₂ construct was transformed into *Escherichia coli* BL21 (DE3) cells (13). The transformed cells were grown at 37 °C until an OD₆₀₀ of 0.6-0.8 was reached. ALKBH5 expression was then induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cell growth was continued for 20 h at 18 °C. The cells were then harvested by centrifugation (Beckman Avanti J-25, rotor JA10, 7000 *g*, 8 min), and the resultant cell pellets were stored at -80 °C. The frozen cell pellets were thawed and resuspended in lysis buffer containing 50 mM tris(hydroxymethyl)aminomethane-hydrochloride (Tris-HCl) pH 7.5, 500 mM

sodium chloride (NaCl), 10 mM imidazole, 1 mg DNaseI, 10 mM MgCl₂, and a Roche cOmplete protease inhibitor cocktail tablet, and subsequently lysed by sonication on ice. The lysates were then centrifuged (Beckman Avanti J-25, rotor JA25.5, 43,400 *g*, 20 min) and the supernatant was loaded onto a 5 mL HisTrap HP column (GE Healthcare) and purified using an AKTA FPLC system. The column was washed with 50 mM Tris-HCl pH 7.5, 500 mM sodium chloride, and 40 mM imidazole, and the protein was eluted with 50 mM Tris-HCl pH 7.5, 500 mM NaCl, 375 mM imidazole. Ethylenediaminetetraacetic acid (EDTA) was added to the protein solution to a final EDTA concentration of 100 mM and then incubated on ice for 30 minutes before being buffer exchanged into 25 mM Tris-HCl pH 7.5, 100 mM NaCl. The protein was further purified using a (5 mL) HiTrap Heparin column (GE Healthcare) followed by a (20 mL) MonoQ column (Pharmacia); in both cases, the protein was eluted with a gradient from 25 mM Tris-HCl pH 7.5 to 500 mM NaCl, 25 mM Tris-HCl pH 7.5. The purified protein was then buffer exchanged into 25 mM Tris-HCl pH 7.5, and concentrated to 11 mg/mL for storage at -80 °C.

4.4.2 Crystallisation

Crystallisation trials of ALKBH5₆₆₋₂₉₂ were carried out in the presence of Mn(II) (as a non-reactive Fe(II) substitute) and various known 2OG oxygenase inhibitors (52). ALKBH5₆₆₋₂₉₂ (MW 28.7 kDa) was crystallised in sitting drops at 293 K by the vapour diffusion method in the presence of Mn(II) and (1-chloro-4-hydroxyisoquinoline-3-carbonyl)glycine (IOX3) or 2,4-pyridine dicarboxylic acid (2,4-PDCA). Crystallisation drops contained 0.2 µL of a protein solution containing a final concentration of 10 mg/mL hexahistidine-tagged

ALKBH5₆₆₋₂₉₂, 0.5 mM manganese(II) chloride and 2 mM IOX3 (41,53) or 2 mM 2,4-PDCA mixed with 0.1 μ L well solution containing 125 mM potassium nitrate and 15% (w/v) polyethylene glycol 3350. Crystals were harvested using nylon loops and cryoprotected by soaking them in well solution diluted with 25% (v/v) glycerol for approximately 30 s and flash cooled in liquid nitrogen.

4.4.3 Data collection and structure determination

Data (0.2° oscillation/image, 180° total rotation) were collected at 100 K on a single crystal at Diamond synchrotron beamline I24 using a wavelength of 0.9686 Å (ALKBH5-2,4-PDCA) or 0.9789 Å (ALKBH5-FG2216) with a Pilatus3 6M detector. All data were indexed, integrated, and scaled using HKL3000 (58). The PHENIX (23) subroutine ENSEMBLER (25) was used to generate an ensemble using the structures of human homologues of

AlkB: FTO (PDB ID 4IE5; resolution: 1.96 Å), ALKBH2 (PDB ID 3S57; resolution: 1.60 Å), ALKBH3 (PDB ID 2IUW; resolution: 1.50 Å), and ALKBH8 (PDB ID 3THT; resolution: 3.01 Å). The ensemble was used as a search model for phasing by molecular replacement. The structure ALKBH5-IOX3 was then solved by molecular replacement using PHASER (25). An initial model of ALKBH5-IOX3 was generated by PHENIX AUTOBUILD (26), which included 411 residues (~97% of the final number of observed residues in the asymmetric unit). Iterative cycles of model building and refinement were performed using COOT (59) and PHENIX (23) until decreasing R and R_{free} no longer converged. (See Table 4 - 2 for detailed data collection and refinement statistics.)

4.4.4 Activity Assays

In vitro demethylation assays (11) were performed in triplicate in a 50 μ L reaction mixture containing 4 μ M ALKBH5₆₆₋₂₉₂, varied concentrations of 5-mer single-stranded RNA (ssRNA) (10 μ M, 20 μ M, 50 μ M, 100 μ M and 200 μ M) with the sequence 5'-GGm⁶ACU-3' (ELLA Biotech, Munich, Germany), 300 μ M 2OG, 2 mM L-ascorbate, 150 μ M diammonium Fe(II) sulphate complex, and 25 mM Tris-HCl pH 7.5. The reaction mixtures were incubated at room temperature and 2 μ L of sample from each reaction was quenched with 2 μ L 20% (v/v) formic acid at specific time points. 1 μ L of each quenched sample was then mixed with 1 μ L of matrix-assisted laser desorption ionisation (MALDI) matrix made up of 2 parts 0.5 M 2,4,6-trihydroxyacetophenone in ethanol and 1 part 0.1 M ammonium citrate dibasic in water. The relative quantities of product and substrate were analysed using MALDI-ToF mass spectrometry. The Michaelis-Menten curve was fitted using non-linear regression and the K_m of the substrate was estimated using GraphPad Prism.

End-point inhibition assays were performed in triplicate for each inhibitor in a 25 μ L reaction mixture (final volume) containing 4 μ M ALKBH5₆₆₋₂₉₂, 80 μ M 5-mer ssRNA with the sequence 5'-GGm⁶ACU-3', 150 μ M 2OG, 2 mM L-ascorbate, 150 μ M diammonium Fe(II) sulphate complex, 150 μ M inhibitor (*N*-oxalylglycine (NOG), 2,4-pyridinedicarboxylate (2,4-PDCA) or IOX3) and 25 mM Tris-HCl pH 7.5. Controls, without inhibitors, in triplicate were also set up. Reactions were incubated at room temperature and quenched after 5 min with an equivolume of 20% (v/v) formic acid and analysed using MALDI-ToF mass spectrometry.

4.4.5 Inhibition assays for IC₅₀ determination

Inhibition assays were performed in triplicate in 25 μ L reaction mixtures (final volume), each containing 700 nM ALKBH5₆₆₋₂₉₂, 3 μ M 5-mer ssRNA with the sequence 5'-GGm⁶ACU-3', 3 μ M 2OG, 1 mM *L*-ascorbate, 100 μ M diammonium Fe(II) sulfate complex, different inhibitor (NOG, 2,4-PDCA, IOX3 and sodium citrate) concentrations ranging from 1 μ M to 1 mM, and 25 mM Tris-HCl pH 7.5. Controls in triplicate without inhibitors were also set up. Reactions were incubated at room temperature and quenched after 6 min with an equivolume of 20% (v/v) formic acid and analysed using MALDI-ToF mass spectrometry.

This research (experimental procedures for IC₅₀ determination of citrate) was originally published in the Journal of Biological Chemistry. Xu, C., Liu, K., Tempel, W., Demetriades, M., Aik, W., Schofield, C.J. and Min, J. Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single stranded m6A RNA demethylation. *J. Biol. Chem.* 2014 (In press) © the American Society for Biochemistry and Molecular Biology.

4.4.6 Assays involving pre-incubation of IOX3 with ALKBH5

Assays were carried out in 50 μ L final volume using similar conditions to those stated in Section 4.4.5 with the exception of 16 μ M IOX3 as the final inhibitor concentration. For each reaction, a mixture of 5 μ L of 160 μ M IOX3, 5 μ L 7 μ M ALKBH5₆₆₋₂₉₂, and 25 μ L of 25 mM Tris-HCl pH 7.5 was preincubated for different durations of time (0, 5, 10, 15, 20, and 30 min). The enzyme-IOX3 mixtures were then added to the other components (to obtain the final concentrations as stated in section 4.4.5) to start the reactions. The reactions

were quenched after 6 min with an equivolume of 20% (v/v) formic acid and analysed using MALDI-ToF mass spectrometry.

Experimental details on ALKBH5 Cys200-IOX3 mass spectrometry and theoretical modeling of substrate can be found in the original publication (Aik *et al.* (2014) 'Structure of human RNA N6-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation.' *Nucleic Acids Res*, **42**, 4741-4754.) attached in the Appendix.

4.5 References

1. Sengoku, T. and Yokoyama, S. (2011) Structural basis for histone H3 Lys 27 demethylation by UTX/KDM6A. *Genes Dev*, **25**, 2266-2277.
2. Carell, T., Brandmayr, C., Hienzsche, A., Muller, M., Pearson, D., Reiter, V., Thoma, I., Thumbs, P. and Wagner, M. (2012) Structure and function of noncanonical nucleobases. *Angew Chem Int Ed Engl*, **51**, 7110-7131.
3. Cantara, W.A., Crain, P.F., Rozenski, J., McCloskey, J.A., Harris, K.A., Zhang, X., Vendeix, F.A., Fabris, D. and Agris, P.F. (2011) The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res*, **39**, D195-201.
4. Machnicka, M.A., Milanowska, K., Osman Oglou, O., Purta, E., Kurkowska, M., Olchowik, A., Januszewski, W., Kalinowski, S., Dunin-Horkawicz, S., Rother, K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res*, **41**, D262-267.
5. Furuichi, Y. and Shatkin, A.J. (2000) Viral and cellular mRNA capping: past and prospects. *Adv Virus Res*, **55**, 135-184.
6. Bokar, J.A. (2005) In Grosjean, H. (ed.), *Fine-Tuning of RNA Functions by Modification and Editing*. Springer, Vol. 12, pp. 141-177.
7. Bokar, J.A., Rath-Shambaugh, M.E., Ludwiczak, R., Narayan, P. and Rottman, F. (1994) Characterization and partial purification of mRNA N6-adenosine methyltransferase from HeLa cell nuclei. Internal mRNA methylation requires a multisubunit complex. *J Biol Chem*, **269**, 17697-17704.
8. Chang, B., Chen, Y., Zhao, Y. and Bruick, R.K. (2007) JMJD6 is a histone arginine demethylase. *Science*, **318**, 444-447.

9. Jia, G., Fu, Y. and He, C. (2013) Reversible RNA adenosine methylation in biological regulation. *Trends Genet*, **29**, 108-115.
10. Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G. *et al.* (2011) N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat Chem Biol*, **7**, 885-887.
11. Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vagbo, C.B., Shi, Y., Wang, W.L., Song, S.H. *et al.* (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell*, **49**, 18-29.
12. Gerken, T., Girard, C.A., Tung, Y.C., Webby, C.J., Saudek, V., Hewitson, K.S., Yeo, G.S., McDonough, M.A., Cunliffe, S., McNeill, L.A. *et al.* (2007) The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science*, **318**, 1469-1472.
13. Thalhammer, A., Bencokova, Z., Poole, R., Loenarz, C., Adam, J., O'Flaherty, L., Schodel, J., Mole, D., Giaslakitotis, K., Schofield, C.J. *et al.* (2011) Human AlkB homologue 5 is a nuclear 2-oxoglutarate dependent oxygenase and a direct target of hypoxia-inducible factor 1alpha (HIF-1alpha). *PLoS One*, **6**, e16210.
14. Tsujikawa, K., Koike, K., Kitae, K., Shinkawa, A., Arima, H., Suzuki, T., Tsuchiya, M., Makino, Y., Furukawa, T., Konishi, N. *et al.* (2007) Expression and sub-cellular localization of human ABH family molecules. *J Cell Mol Med*, **11**, 1105-1116.
15. Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M. *et al.* (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature*, **485**, 201-206.
16. Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E. and Jaffrey, S.R. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell*, **149**, 1635-1646.
17. Schofield, C.J. and Zhang, Z. (1999) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol*, **9**, 722-731.
18. Aik, W., Demetriades, M., Hamdan, M.K., Bagg, E.A., Yeoh, K.K., Lejeune, C., Zhang, Z., McDonough, M.A. and Schofield, C.J. (2013) Structural basis for inhibition of the fat mass and obesity associated protein (FTO). *J Med Chem*, **56**, 3680-3688.
19. Yan, L., Colandrea, V.J. and Hale, J.J. (2010) Prolyl hydroxylase domain-containing protein inhibitors as stabilizers of hypoxia-inducible factor: small molecule-based therapeutics for anemia. *Expert Opin Ther Pat*, **20**, 1219-1245.
20. Blasiak, L.C., Vaillancourt, F.H., Walsh, C.T. and Drennan, C.L. (2006) Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature*, **440**, 368-371.
21. Hewitson, K.S., Holmes, S.L., Ehrismann, D., Hardy, A.P., Chowdhury, R., Schofield, C.J. and McDonough, M.A. (2008) 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.

22. Elkins, J.M., Hewitson, K.S., McNeill, L.A., Seibel, J.F., Schlemminger, I., Pugh, C.W., Ratcliffe, P.J. and Schofield, C.J. (2003) Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 α . *J Biol Chem*, **278**, 1802-1806.
23. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*, **66**, 213-221.
24. Lando, D., Peet, D.J., Gorman, J.J., Whelan, D.A., Whitelaw, M.L. and Bruick, R.K. (2002) FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev*, **16**, 1466-1471.
25. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
26. Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.W., Read, R.J. and Adams, P.D. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr*, **64**, 61-69.
27. Clifton, I.J., McDonough, M.A., Ehrismann, D., Kershaw, N.J., Granatino, N. and Schofield, C.J. (2006) Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem*, **100**, 644-669.
28. McDonough, M.A., Loenarz, C., Chowdhury, R., Clifton, I.J. and Schofield, C.J. (2010) Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struct Biol*, **20**, 659-672.
29. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol*, **292**, 195-202.
30. Shepard, P.J. and Hertel, K.J. (2009) The SR protein family. *Genome Biol*, **10**, 242.
31. Aik, W., McDonough, M.A., Thalhammer, A., Chowdhury, R. and Schofield, C.J. (2012) Role of the jelly-roll fold in substrate binding by 2-oxoglutarate oxygenases. *Curr Opin Struct Biol*, **22**, 691-700.
32. Chowdhury, R., Candela-Lena, J.I., Chan, M.C., Greenald, D.J., Yeoh, K.K., Tian, Y.M., McDonough, M.A., Tumber, A., Rose, N.R., Conejo-Garcia, A. *et al.* (2013) Selective Small Molecule Probes for the Hypoxia Inducible Factor (HIF) Prolyl Hydroxylases. *ACS Chem Biol*.
33. Arnold, L.A., Estebanez-Perpina, E., Togashi, M., Jouravel, N., Shelat, A., McReynolds, A.C., Mar, E., Nguyen, P., Baxter, J.D., Fletterick, R.J. *et al.* (2005) Discovery of small molecule inhibitors of the interaction of the thyroid hormone receptor with transcriptional coregulators. *J Biol Chem*, **280**, 43048-43055.
34. Hwang, J.Y., Huang, W., Arnold, L.A., Huang, R., Attia, R.R., Connelly, M., Wichterman, J., Zhu, F., Augustinaite, I., Austin, C.P. *et al.* (2011) Methylsulfonylnitrobenzoates, a new class of irreversible inhibitors of the interaction of the thyroid hormone receptor and its

- obligate coactivators that functionally antagonizes thyroid hormone. *J Biol Chem*, **286**, 11895-11908.
35. McDonough, M.A., McNeill, L.A., Tilliet, M., Papamicael, C.A., Chen, Q.Y., Banerji, B., Hewitson, K.S. and Schofield, C.J. (2005) Selective inhibition of factor inhibiting hypoxia-inducible factor. *Journal of the American Chemical Society*, **127**, 7680-7681.
 36. Rose, N.R., McDonough, M.A., King, O.N.F., Kawamura, A. and Schofield, C.J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev*, **40**, 4364-4397.
 37. Rose, N.R., Woon, E.C., Kingham, G.L., King, O.N., Mecinovic, J., Clifton, I.J., Ng, S.S., Talib-Hardy, J., Oppermann, U., McDonough, M.A. *et al.* (2010) Selective inhibitors of the JMJD2 histone demethylases: combined nondenaturing mass spectrometric screening and crystallographic approaches. *Journal of Medicinal Chemistry*, **53**, 1810-1818.
 38. Cunliffe, C.J., Franklin, T.J., Hales, N.J. and Hill, G.B. (1992) Novel Inhibitors of Prolyl 4-Hydroxylase .3. Inhibition by the Substrate-Analog N-Oxaloglycine and Its Derivatives. *J Med Chem*, **35**, 2652-2658.
 39. McDonough, M.A., Li, V., Flashman, E., Chowdhury, R., Mohr, C., Lienard, B.M., Zondlo, J., Oldham, N.J., Clifton, I.J., Lewis, J. *et al.* (2006) Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci U S A*, **103**, 9814-9819.
 40. Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y. and Chai, J. (2010) Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature*, **464**, 1205-1209.
 41. Pastore, C., Topalidou, I., Forouhar, F., Yan, A.C., Levy, M. and Hunt, J.F. (2012) Crystal structure and RNA binding properties of the RNA recognition motif (RRM) and AlkB domains in human AlkB homolog 8 (ABH8), an enzyme catalyzing tRNA hypermodification. *J Biol Chem*, **287**, 2130-2143.
 42. Sundheim, O., Vagbo, C.B., Bjoras, M., Sousa, M.M., Talstad, V., Aas, P.A., Drablos, F., Krokan, H.E., Tainer, J.A. and Slupphaug, G. (2006) Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *Embo J*, **25**, 3389-3397.
 43. Yang, C.G., Yi, C., Duguid, E.M., Sullivan, C.T., Jian, X., Rice, P.A. and He, C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature*, **452**, 961-965.
 44. Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R. and Hunt, J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879-884.
 45. Kato, M., Arais, Y., Noma, A., Nagao, A., Suzuki, T., Ishitani, R. and Nureki, O. (2011) Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res*, **39**, 1576-1585.

46. Iyer, L.M., Abhiman, S., de Souza, R.F. and Aravind, L. (2010) Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. *Nucleic Acids Res*, **38**, 5261-5279.
47. Noma, A., Ishitani, R., Kato, M., Nagao, A., Nureki, O. and Suzuki, T. (2010) Expanding role of the jumonji C domain as an RNA hydroxylase. *J Biol Chem*, **285**, 34503-34507.
48. van Staalduinen, L.M., Novakowski, S.K. and Jia, Z. (2014) Structure and Functional Analysis of YcfD, a Novel 2-Oxoglutarate/Fe-Dependent Oxygenase Involved in Translational Regulation in Escherichiacoli. *J Mol Biol*.
49. Choudhary, C., Kumar, C., Gnäd, F., Nielsen, M.L., Rehman, M., Walther, T.C., Olsen, J.V. and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, **325**, 834-840.
50. Holland, P.J. and Hollis, T. (2010) Structural and Mutational Analysis of Escherichia coli AlkB Provides Insight into Substrate Specificity and DNA Damage Searching. *PLoS One*, **5**.
51. Yu, B. and Hunt, J.F. (2009) Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc Natl Acad Sci U S A*, **106**, 14315-14320.
52. Hopkinson, R.J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K.H., Li, X.S., Kristensen, J.B.L., King, O.N.F., Chan, M.C. *et al.* (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci*, **4**, 3110-3117.
53. Mantri, M., Krojer, T., Bagg, E.A., Webby, C.A., Butler, D.S., Kochan, G., Kavanagh, K.L., Oppermann, U., McDonough, M.A. and Schofield, C.J. (2010) Crystal Structure of the 2-Oxoglutarate- and Fe(II)-Dependent Lysyl Hydroxylase JMJD6. *J Mol Biol*.
54. Chowdhury, R., Candela-Lena, J.I., Chan, M.C., Greenald, D.J., Yeoh, K.K., Tian, Y.M., McDonough, M.A., Tumber, A., Rose, N.R., Conejo-Garcia, A. *et al.* (2013) Selective small molecule probes for the hypoxia inducible factor (HIF) prolyl hydroxylases. *ACS Chem Biol*, **8**, 1488-1496.
55. Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y. and Cheng, X. (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol*, **17**, 38-43.
56. Webby, C.J., Wolf, A., Gromak, N., Dreger, M., Kramer, H., Kessler, B., Nielsen, M.L., Schmitz, C., Butler, D.S., Yates, J.R., 3rd *et al.* (2009) Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science*, **325**, 90-93.
57. Hausinger, R.P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol*, **39**, 21-68.
58. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol*, **276**, 307-326.
59. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D*, **66**, 486-501.

Chapter 5

Preliminary Characterisation of OGFOD2

5.1 Introduction

Collagen is the most abundant protein in humans. Its various roles include serving as a structural protein in connective tissues, thus the process of collagen biosynthesis is of paramount importance (1). Several 2OG oxygenases such as the collagen prolyl-4-hydroxylase (CPH) and pro-collagen lysyl-hydroxylases (PLODs) have been found to play essential roles in the hydroxylation of collagen, aiding in the formation of the collagen triple helix (2-5). To date, crystal structures of the oxygenase domains of CPH and PLODs have not been determined, hampering efforts in studying the mechanisms of action of this important group of enzymes.

Protein sequence alignment performed by Aravind and Koonin categorised a novel protein, 2-oxoglutarate and iron-dependent oxygenase domain-containing protein 2 (OGFOD2), into the lysyl hydroxylase subfamily of 2OG oxygenases (Figure 5 - 1) (6). Its closest relatives within the family are the pro-collagen lysyl-hydroxylases (LH1, LH2, and LH3) (6). However, the cellular and biochemical functions of OGFOD2 are unknown. Preliminary data from yeast two hybrid experiments identified elongation factor 1-gamma and histone acetyltransferase KAT5 as potential OGFOD2 interactors (7). This chapter describes preliminary functional, biochemical and structural characterisations of OGFOD2.

OGFOD2	MATVGA ^W PRHF ^W CR ^W CAC ^W CTD ^W NL ^W YVARY ^W GL ^W -HVR ^W FRGEQQLRRDY ^W GPI ^W LRSR ^W - - - - - G	50
PLOD1	WGAL ^W SADGY ^W ARSEDY ^W -VDI ^W VQGR ^W RVGVW ^W NPY ^W ISNI ^W YLI ^W K ^W -GSAL ^W RGE ^W LQSSDL ^W F	472
PLOD2	WGAL ^W SPDGY ^W ARSEDY ^W -VDI ^W VQGN ^W RVGVW ^W NPY ^W MANV ^W YLI ^W K ^W -GKTL ^W RSEM ^W NERN ^W YF	482
PLOD3	WGAL ^W SPDEY ^W ARSEDY ^W -VEL ^W VQRK ^W RVGVW ^W NPY ^W ISQAY ^W VI ^W R ^W -GDTL ^W RMEL ^W PQRDV ^W F	482
OGFOD2	CVSAK ^W DFQQL ^W LAELE ^W QEVER ^W RQRL ^W GQESA ^W ARKA ^W LI ^W ASSY ^W HPAR ^W -PEVY ^W DSL ^W QDAAL ^W	105
PLOD1	HHSKL ^W DPDMA ^W FCANI ^W RQQDV ^W FMFL ^W TNRHT ^W LGHL ^W SLDSY ^W RTTH ^W LHND ^W LWEV ^W FSNP ^W ED	529
PLOD2	VRDKL ^W DPDMAL ^W CRNARE ^W MGV ^W MYISNR ^W HEFGRL ^W STANY ^W NTSHY ^W NNDL ^W WQI ^W FENP ^W VD	539
PLOD3	SGSDT ^W DPDMA ^W FCCKS ^W FRDKGI ^W FLHL ^W SNQHE ^W FGRL ^W LATSR ^W YDTEH ^W LHPD ^W LWQI ^W FDNP ^W VD	539
OGFOD2	AP ^W EFLAVTE ^W YSVSPDADL ^W KGL ^W LQRLET ^W VSEEK ^W RYRV ^W PVFTAPF ^W CQALL ^W E ^W ELEHFE ^W Q	162
PLOD1	WKEKYI ^W HQNYT ^W - - - - - KALAGK ^W -LVET ^W PCPDV ^W YWFPI ^W FTEVAC ^W DELVE ^W MEHFG ^W Q	578
PLOD2	WKEKYI ^W NRDYS ^W - - - - - KIFTEN ^W -IVEQPCPD ^W VWFPI ^W FSEKAC ^W DELVE ^W MEHYG ^W K	588
PLOD3	WKEQYI ^W HENYS ^W - - - - - RALEGEGI ^W VEQPCPD ^W VWFPL ^W SEQMC ^W DELVA ^W MEHYG ^W Q	589
OGFOD2	SDMPKGRPN ^W - - - - - TMN ^W NYGVL ^W HEL ^W GLDEPL ^W MTPLRER ^W FIQPL ^W MALLY ^W PDCGGG ^W	212
PLOD1	WSLGN ^W NDNRI ^W QGGYEN ^W VPTDI ^W HMNQI ^W GFERE ^W WHKFL ^W LEYI ^W APMTE ^W KLYPGYYT ^W -	633
PLOD2	WSGGK ^W HDSRI ^W SGGYEN ^W VPTDDI ^W HMKQV ^W DLENV ^W WLDFIRE ^W FIAPVT ^W LKVFA ^W GYT ^W -	643
PLOD3	WSGGR ^W HDSRLAGGYEN ^W VPTVDI ^W HMKQV ^W GYEDQWL ^W -QLLRT ^W YVGPMT ^W ESL ^W FPGYHT ^W -	644
OGFOD2	RLDSHRA ^W FVVKY ^W APGQDLE ^W LGCHY ^W DNAEL ^W TLNVAL ^W - - GKVF ^W TGGALY ^W FGG ^W LFQAPT ^W	266
PLOD1	RAQFDL ^W AFVVRYP ^W DEQPSL ^W MPH ^W HDASTFTI ^W NIALNRV ^W GVDYEGG ^W GCRFLRYNC ^W - - -	687
PLOD2	KGFALL ^W NFVVKYSP ^W ERQRS ^W LRPH ^W HDASTFTI ^W NIALNNV ^W GEDFQGG ^W GCKFLRYNC ^W - - -	697
PLOD3	KARAVMN ^W FVVRYP ^W DEQPSL ^W LRPH ^W DSSTFT ^W LNVAL ^W NHKG ^W LDYEGG ^W GCRFLRYDC ^W - - -	698
OGFOD2	ALTEP ^W LEVEHV ^W GQGV ^W LHRGGQL ^W HGAR ^W PLGTGE ^W WNLV ^W VWLR ^W A	309
PLOD1	SIRAP ^W RKGWTL ^W MHPG ^W - - - RLTHY ^W HEGLPT ^W TRGT ^W RYI ^W AVSFVDP ^W	727
PLOD2	SIESP ^W RKGWSF ^W MHPG ^W - - - RLTHL ^W HEGLPV ^W KNGT ^W RYI ^W AVSFIDP ^W	737
PLOD3	VISSP ^W RKGWALL ^W HPG ^W - - - RLTHY ^W HEGLPT ^W TWGT ^W RYI ^W MVSFVDP ^W	738

Figure 5 - 1. Sequence alignment of human OGFOD2 (Uniprot ID: Q6N063), human procollagen lysyl hydroxylase 1 (PLOD1, Uniprot ID: Q02809), human procollagen lysyl hydroxylase 2 (PLOD2, Uniprot ID: O00469) and human procollagen lysyl hydroxylase 3 (PLOD3, Uniprot ID: O60568). Residues highlighted: conserved (dark blue); semi-conserved (light blue); weakly conserved (grey); conserved 2OG oxygenase catalytic triad HXD...H (red); conserved 2OG binding arginine (green).

5.2 Results and Discussions

According to Uniprot, there are four splice variants for OGFOD2: Isoform 1 (350 amino acids), Isoform 2 (290 amino acids), Isoform 3 (129 amino acids) and Isoform 4 (186 amino acids). OGFOD2 Isoforms 3 and 4 do not contain the complete 2OG oxygenase domain. The OGFOD2 Isoform 2 construct, obtained from a cDNA library derived from MCF-7 cells, was cloned into the pEGFP-C1 and pEGFP-N1 vectors (Clontech) by Dr Alexander Wolf. As the OGFOD2 Isoform 2 (*O2i2*) gene construct was available in the laboratory, initial studies were carried out using OGFOD2 Isoform 2 protein (*O2i2*). The OGFOD2 Isoform 1 (*O2i1*) gene

encoding for OGFOD2 Isoform 1 protein (O2i1) was later synthesised by GeneArt, Invitrogen.

5.2.1 Cellular Localisation Studies

In order to obtain information about the localisation of OGFOD2 in the cell, immunofluorescence experiments were carried out in HeLa cells. The O2i2 construct was inserted into an N-terminal FLAG-tag vector, p3XFLAG-CMV-10. Plasmids encoding O2i2 were transfected into HeLa cells for over-production of N- and C-terminal green fluorescent protein (GFP)-tagged OGFOD2 and N-terminal FLAG-tagged O2i2. The GFP-OGFOD2 transfected HeLa cells were fixed and stained using 4',6'-diamidino-2-phenylindole (DAPI; a DNA stain) 24 h post-transfection. FLAG-tagged O2i2 producing cells were stained with mouse anti-FLAG primary antibody and anti-mouse-Cy3 conjugated secondary antibody, followed by DAPI. Observation of the cells by fluorescence microscopy showed that transiently over-produced N- and C-terminal GFP-tagged O2i2 proteins were widespread in the cell, localising in both the nucleus and the cytoplasm. There is little difference between images obtained for GFP-tagged O2i2 and GFP control. However, GFP-tagged O2i2 appears absent from certain regions of the nucleus, potentially nucleoli (Figure 5 - 2a-c). Similarly, transiently over-produced N-terminal FLAG-tagged O2i2 was observed to be localised in both the nucleus and cytoplasm (Figure 5 - 2d). FLAG-tagged O2i2 in the cytoplasm appears to form a web-like structure (Figure 5 - 2d).

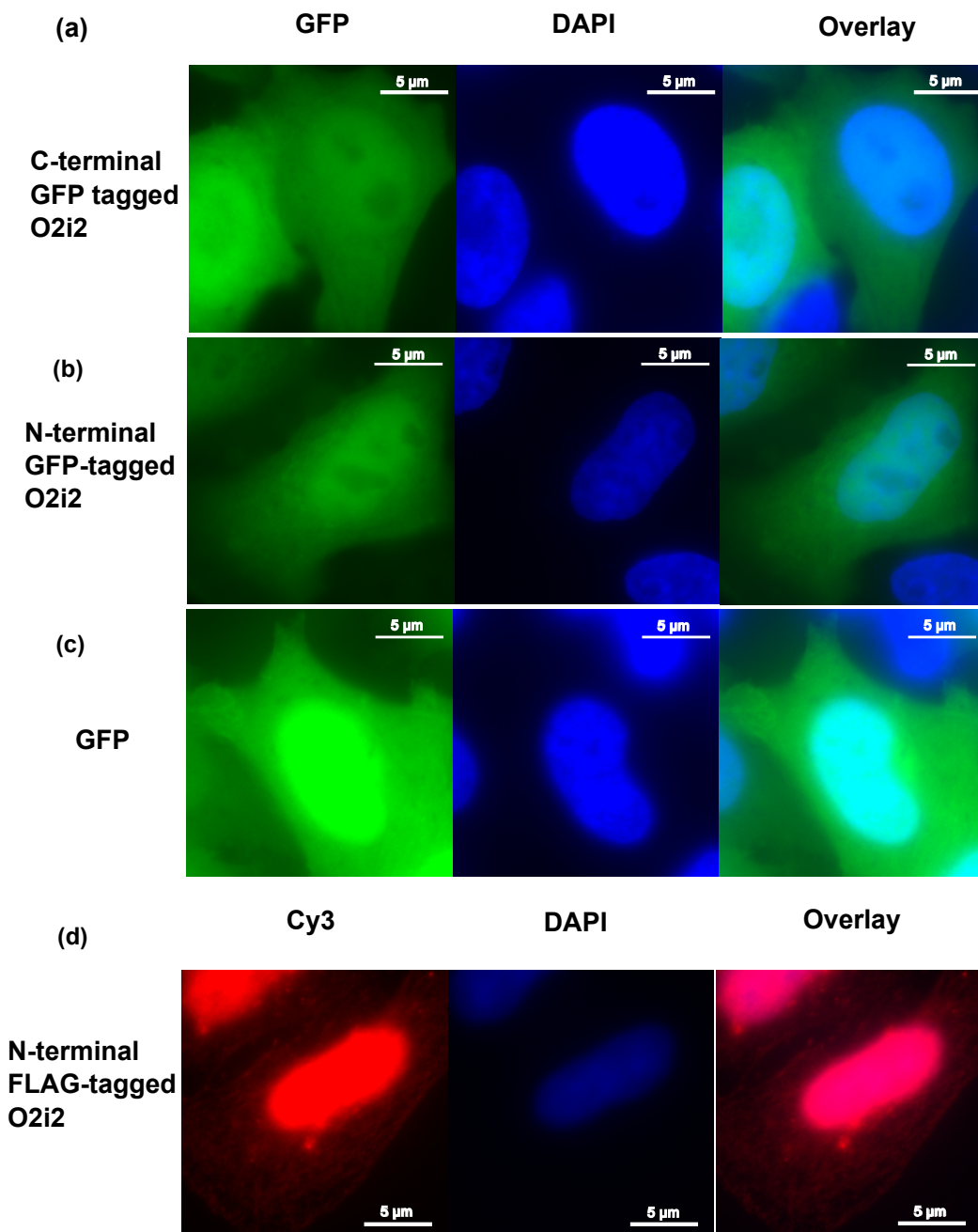


Figure 5 - 2. Immunofluorescence study on O2i2. Fluorescence microscopy images of HeLa cells **(a)** expressing C-terminal GFP tagged OGFOD2 isoform 2 (O2i2), **(b)** N-terminal GFP-tagged O2i2, **(c)** GFP, and **(d)** expressing N-terminal FLAG-tagged O2i2. (a,b,c) Images from left: GFP staining; DAPI staining (DNA); and an overlay image. (d) Images from left: Cy3 signal representing FLAG-tagged O2i2; DAPI staining (DNA); an overlay image.

Protein over-production often causes intense staining and might not reflect the true subcellular localisation of endogenous protein. In the case of

O2i2, it is hard to differentiate between the subcellular localisations of GFP-tagged O2i2 and the GFP control because the nucleus and cytoplasm were strongly fluorescent (Figure 5 - 2). Hence, it was thought that observation of endogenous OGFOD2 might provide a better indication of OGFOD2 localisation. Untransfected HEK293T cells and HeLa cells were stained by an anti-OGFOD2 primary antibody and an anti-mouse-Cy3 secondary antibody. The anti-OGFOD2 had been validated through its ability to stain over-produced OGFOD2 in HeLa cells. A control experiment was performed without using the anti-OGFOD2 primary antibody. Fluorescence microscopy of HEK293T and HeLa cells did not show obvious staining of endogenous OGFOD2 (data not shown), even though speckles of staining in the cytoplasm can be observed at high exposure, suggesting that endogenous OGFOD2 is not produced at a significant level in those cells.

To further investigate the subcellular localisation of OGFOD2, the synthesised *O2i1* gene was cloned into the p3XFLAG_CMV-10 vector and transfected into HeLa cells. The cells were then stained with anti-FLAG primary antibody and anti-mouse-Cy3 conjugated secondary antibody. Immunofluorescence microscopy initially showed similar nuclear and cytoplasmic staining of FLAG-tagged O2i1 as previously observed for FLAG-tagged O2i2. But after adjustment of the focal plane to the nucleus using DAPI staining as a guide, an intense ring of Cy3 staining was observed to localise predominantly around the nucleus (Figure 5 - 3a), suggesting that the FLAG-tagged O2i1 was localised in the endoplasmic reticulum (ER). Therefore, the initial inference that OGFOD2 is nuclear localised was most probably inaccurate; this proposal is likely to be due to the microscope's focus on the nuclear surface.

Notably, the closest relatives of OGFOD2, the procollagen lysyl hydroxylases (PLODs), are also localised in the ER (8-11).

To further investigate whether O2i1 is localised in the ER, a colocalisation study of FLAG-tagged O2i1 and a known ER localised protein, protein disulphide isomerase (PDI) (12), was carried out. A rabbit anti-PDI primary antibody was used to stain PDI and a mouse anti-FLAG primary antibody was used to stain FLAG-O2i1. Staining was then performed with anti-rabbit-cy3 conjugated secondary antibody (red) and anti-mouse-FITC conjugated antibody (green). Fluorescence microscopy showed that the red and green stains largely overlap around the nucleus (Figure 5 - 3b), suggesting colocalisation of FLAG-O2i1 and PDI, thus supporting the proposal that FLAG-O2i1 is ER localised. A similar cellular localisation pattern was also observed in U2OS cells (Figure 5 - 3c). Subsequently, endogenous OGFOD2 was then investigated in U2OS cells because endogenous OGFOD2 in HEK293T and HeLa cells could not be observed by fluorescence microscopy. U2OS cells were fixed and stained with mouse anti-OGFOD2 primary antibody, rabbit anti-PDI primary antibody, anti-mouse-FITC conjugated secondary antibody (green), and anti-rabbit-Cy3 conjugated secondary antibody (red). Endogenous OGFOD2 was observed around the nucleus of U2OS cells and it also colocalised with PDI (Figure 5 - 3d).

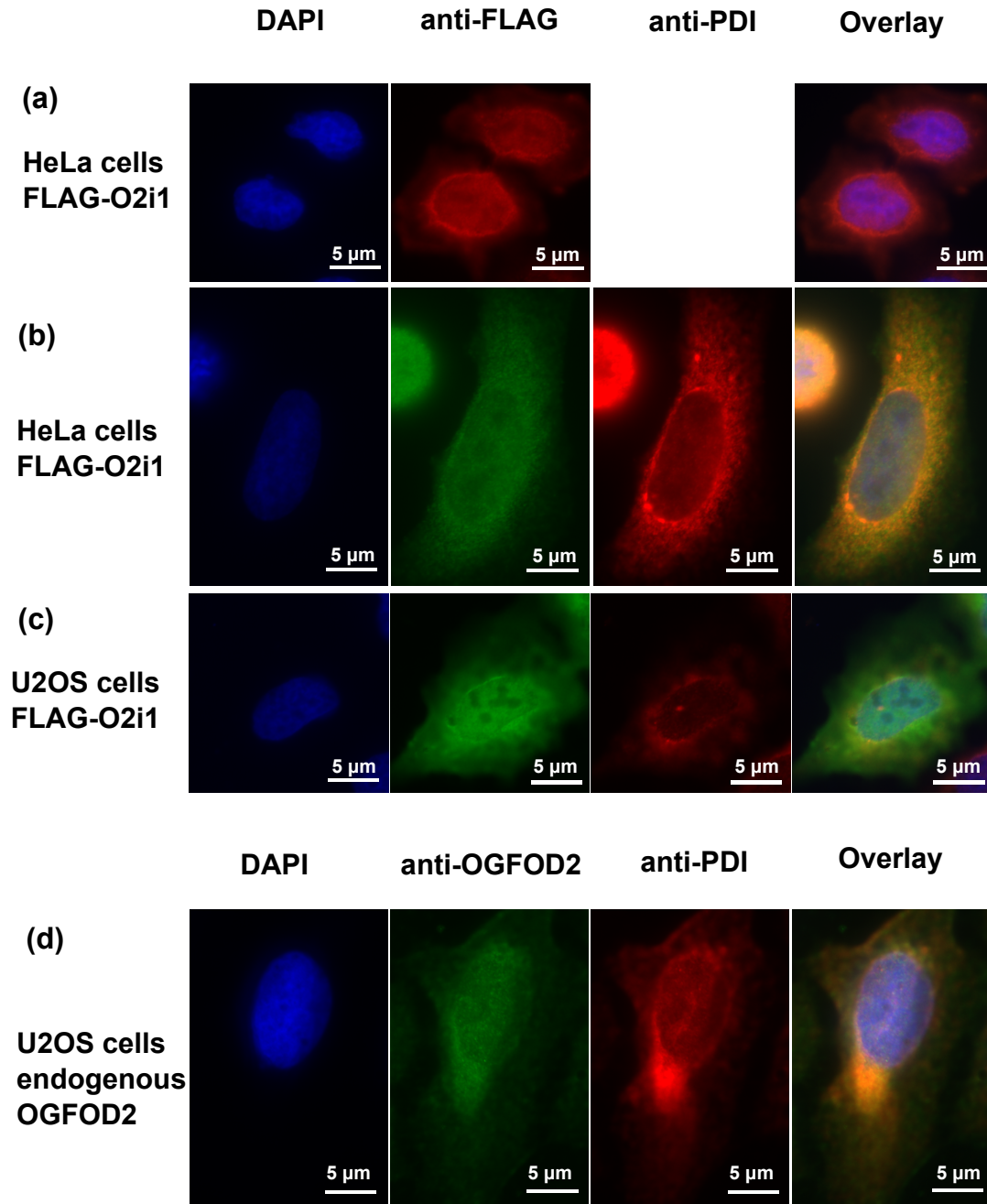


Figure 5 - 3. Immunofluorescence study on O2i1. Fluorescence microscopy images of **(a)** HeLa cells expressing N-terminal FLAG-tagged OGFOD2 isoform 1 (O2i1) stained with anti-FLAG antibody, **(b)** HeLa cells expressing N-terminal FLAG-tagged O2i1 stained with anti-FLAG and anti-PDI antibodies, **(c)** U2OS cells expressing N-terminal FLAG-tagged O2i1 stained with anti-FLAG and anti-PDI antibodies, and **(d)** U2OS cells expressing endogenous OGFOD2 and stained with anti-OGFOD2 and anti-PDI antibodies. (a) Images from left: DAPI staining (DNA); Cy3 signal staining FLAG-tagged O2i1; and an overlay image. (b and c) DAPI staining (DNA); FITC signal staining FLAG-tagged O2i1; Cy3 signal staining PDI (ER marker); an overlay image. (d) Images from left: DAPI staining (DNA); FITC signal staining endogenous OGFOD2; Cy3 signal staining PDI; an overlay image.

The combined fluorescence microscopic data suggest that OGFOD2 is a protein that resides predominantly in the ER.

5.2.2 Attempted substrate identification in cells and *in vitro*

Identification of the natural substrate and binding partners of OGFOD2 should be useful in understanding the functional role of OGFOD2. It was proposed that the identification of OGFOD2 substrate might be achieved by performing co-immunoprecipitation (co-IP) experiments coupled to fragmentation mass spectrometry (MS/MS) analysis. N-Terminal tagged GFP-O2i2, C-terminal tagged GFP-O2i2 and GFP (as a control) were transiently produced in HEK293T cells. Anti-GFP beads (GFP-trap_A; Chromotek, Germany) were used for co-IP of O2i2 with potential binding partners. Liquid chromatography-mass spectrometry (LC-MS/MS) analysis followed by Mascot searches were then conducted to identify these potential binding partners. Table A - 3 and Table A - 4 (Appendix) show lists of potential binding partners (with an exponentially modified Protein Abundance Index (emPAI) score (13) of at least 0.1) for the C- and N-terminal GFP-tagged OGFOD2 samples that were not identified in the GFP control IP. emPAI score is a scale used for estimating absolute protein abundance and is calculated using the following equation (13):

$$\text{emPAI} = 10^{\text{PAI}} - 1, \text{PAI} = \frac{N_{\text{obsd}}}{N_{\text{obsl}}}$$

N_{obsd} = number of observed peptides per protein

N_{obsl} = number of observable peptides per protein

For the N-terminal tagged GFP-O2i2, tubulin beta-4, beta-2A, beta-3, alpha-1B and alpha-8 chains had top abundance and emPAI score; for the C-terminal tagged GFP-O2i2, tubulin beta-2C, beta-3 and alpha-4A chains had top abundance and emPAI scores. However, cytoskeletal proteins are ubiquitous in cells and their presence is a common occurrence in proteomic data (14), thus the tubulin proteins identified from co-IP and MS/MS were probably not substrates.

As OGFOD2 is similar by sequence comparison to the PLODs, a total of 32 peptides from tubulin, collagen, and other collagen-like proteins such as adiponectin, complement C1q, and mannose binding lectin (synthesised by Dr Wei Ge and Dr Monica Mantri) (see Table 5 – 5 in Section 5.4.11 for sequence of peptides), were tested as substrates for recombinant O2i1 and O2i2 (see Section 5.2.3 on purification of recombinant OGFOD2) in the presence of 2OG, Fe(II) and ascorbate for hydroxylase activities, using matrix assisted laser desorption ionisation - time of flight mass spectrometry (MALDI-ToF MS) for mass shift analyses. However, the MALDI-ToF MS analyses did not show any +16 Da mass shifts, suggesting that OGFOD2 did not hydroxylate any of the tested peptides.

5.2.3 Expression, purification and biophysical studies

The production and purification of hexahistidine-tagged O2i1 and O2i2 were carried out in order to obtain recombinant proteins for *in vitro* biochemical studies. O2i1 was largely insoluble resulting in a very low yield after nickel affinity purification. O2i2 was mostly soluble and a substantial amount of protein was produced and purified. O2i2 was purified using nickel affinity purification followed by Superdex 75 (S75) size exclusion chromatography. Size exclusion

chromatography of O2i2 and SDS-PAGE analysis showed two main peaks corresponding to O2i2 (Figure 5 - 4a, b) that eluted after the void volume of the S75 size exclusion column, suggesting that O2i2 can potentially oligomerise in solution.

Native PAGE was used to investigate the oligomerisation states of O2i2. Fractions from the two peaks from size exclusion chromatography were separately collected and concentrated. Samples corresponding to each peak were analysed using native PAGE. Bovine serum albumin (monomer MW=67 kDa) was used as a standard. Native PAGE analysis (Figure 5 - 4c) showed that Peak 1 gave predominantly a band about the same molecular weight as a BSA monomer, suggesting that it corresponds to an O2i2 dimer (MW=64 kDa). Peak 2, which eluted later on size exclusion ran at a lower molecular weight than the BSA monomer on native PAGE gel, indicating that it is predominantly monomer. Results from native PAGE analysis confirmed that O2i2 could form a dimer in solution.

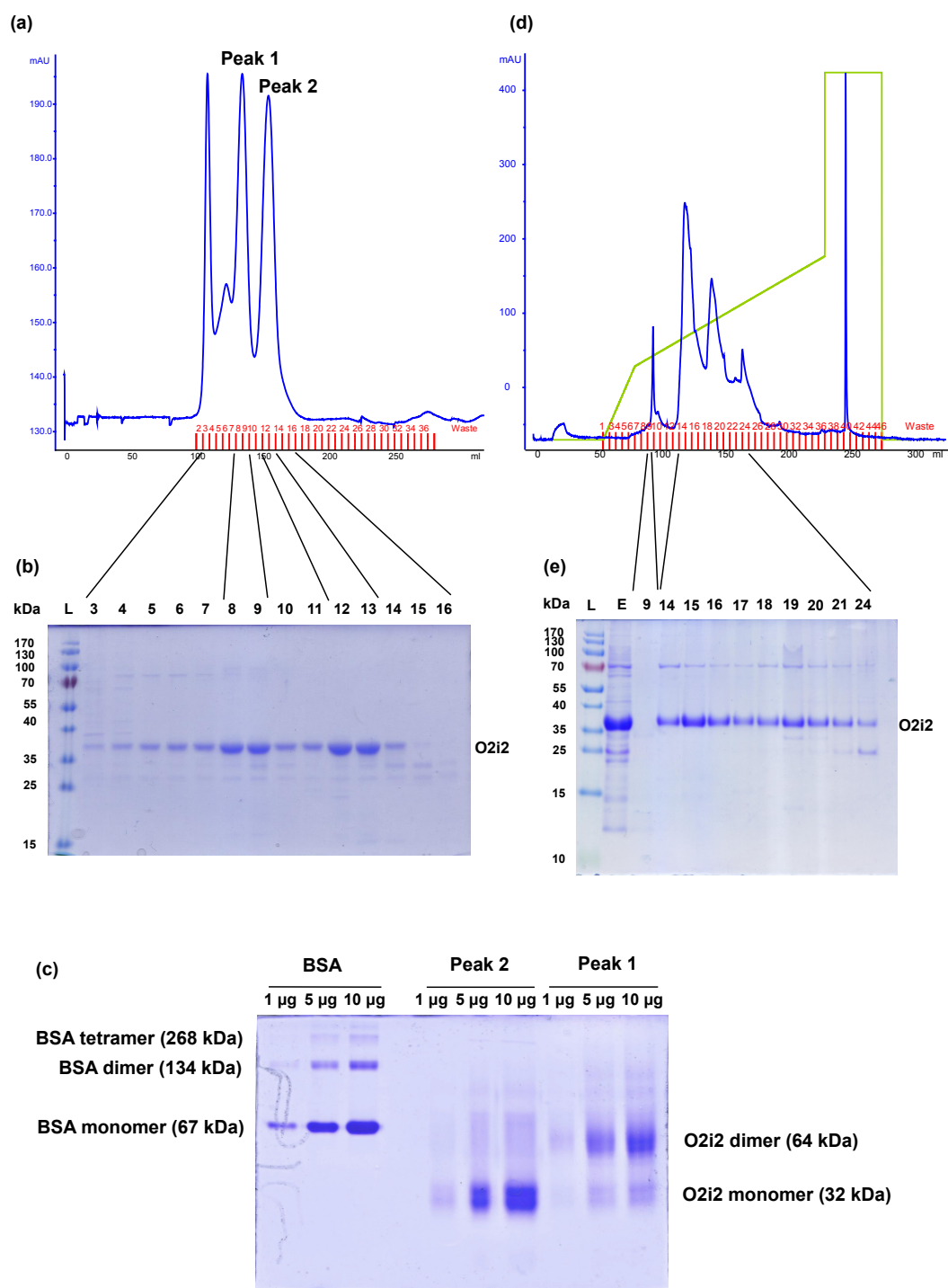


Figure 5 - 4. Chromatograms and PAGE analyses of OGFOD2 isoform 2 (O2i2). (a) S75 size exclusion chromatography of O2i2 after nickel affinity purification. (b) 12% SDS-PAGE gel analysis of fractions from S75 size exclusion chromatography. Two peaks from S75 chromatography correspond to OGFOD2 (Peak 1 and Peak 2). (c) 12.5% Native PAGE analysis of proteins from Peak 1 and Peak 2 showed that Peak 1 is predominantly O2i2 dimer and Peak 2 is predominantly O2i2 monomer. Bovine serum albumin (BSA) was used as a standard. (d) UV trace (blue) of MonoQ anion exchange of O2i2 after nickel affinity purification. Green line represents concentration of buffer B (25 mM Tris-HCl pH7.5, 1 M NaCl). Two major peaks could be observed. (e) 12% SDS-PAGE analysis of fractions from MonoQ anion exchange (b). Numbers above the SDS-PAGE gels in (b) and (e) represent fraction numbers, L represents protein ladder, and the numbers on the left of the gels indicate molecular weight in kDa.

5.2.4 OGFOD2 catalyses substrate uncoupled 2OG turnover

Most 2OG oxygenases are able to oxidise 2OG in the absence of prime substrates, producing carbon dioxide and succinate (2OG uncoupled turnover) (15). A 2OG turnover assay of O2i2 was performed to show that the purified recombinant O2i2 is active and is a potential 2OG oxygenase. Recombinant O2i2 was incubated with 2OG, ammonium iron (II) sulphate and ascorbate for 90 min at 37 °C to investigate whether O2i2 converts 2OG to succinate in the absence of substrate. A control experiment was set up with the same conditions without enzyme. ¹H NMR (performed by Dr Ivanhoe Leung) showed that both samples contained succinate after incubation, but the succinate level of the sample with O2i2 was higher compared to that of the control (Figure 5 - 5), suggesting that OGFOD2 is a potential 2OG oxygenase.

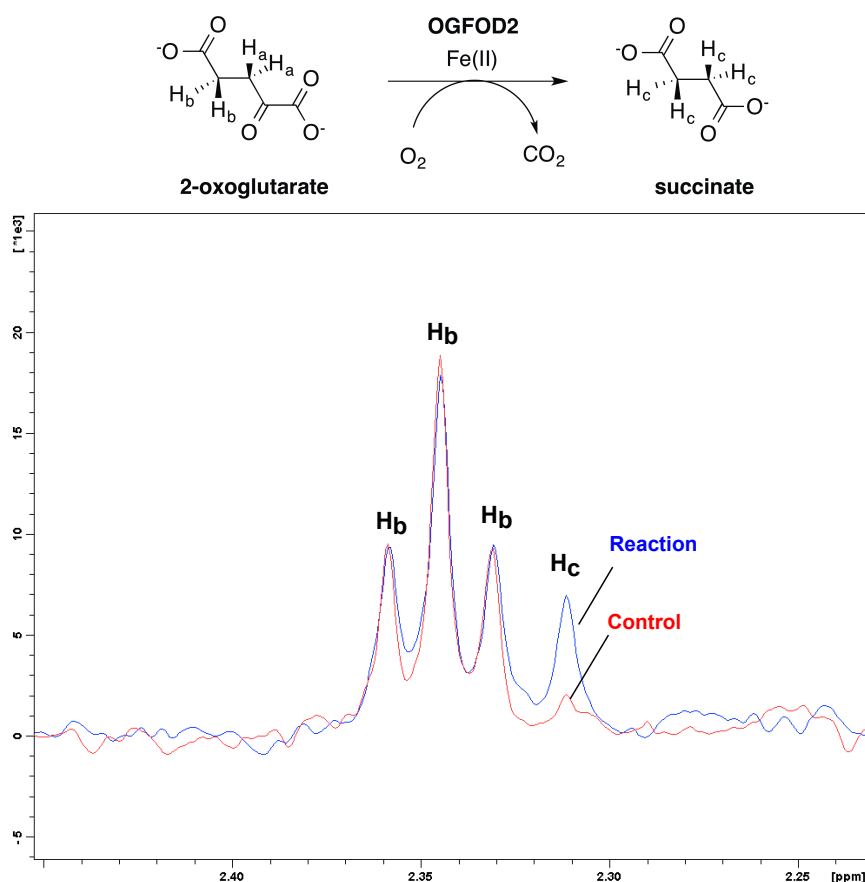


Figure 5 - 5. Proton NMR spectra of OGFOD2 2OG uncoupled turnover assay. Overlay ^1H NMR spectra of a reaction mixture containing O2i2 enzyme, 50 μM Fe(II) and 500 μM 2OG (blue line) and a control reaction without enzyme (red line). Spectrum for reaction sample shows a substantially higher peak corresponding to the $^1\text{H}_c$ protons of succinate ($\delta = 2.30$ ppm) compared to that of the control sample. A triplet peak corresponding to $^1\text{H}_b$ protons of 2OG is observed. Reaction scheme of 2OG uncoupled turnover is shown above the spectra.

5.2.5 Crystallographic studies

Crystallisation trials of OGFOD2 (with and without his-tag cleavage) were carried out using the JCSG+ screen (Hampton Research) in the presence of Mn(II) or Ni(II) and broad spectrum 2OG oxygenase inhibitors *N*-oxalylglycine (NOG) or 2,4-pyridine dicarboxylate (2,4-PDCA) (16,17), but crystal formation was not observed. The purification protocol of OGFOD2 was then optimised by including MonoQ anion exchange (Figure 5 - 4b, d), in place of size exclusion

chromatography. Protein purified using this method crystallised in several conditions:

- (i) 0.1 M trisodium citrate pH 5.5, 20% PEG 3000
- (ii) 1 M diammonium phosphate, 0.1 M sodium acetate pH 4.5
- (iii) 0.8 M succinic acid pH 7.0
- (iv) 1 M ammonium sulfate, 0.1 M bis-tris pH 5.5, 1% PEG 3350

The crystals had a hexagonal bipyramid morphology (Figure 5 - 6a). One of the crystals (size $\sim 100\ \mu\text{m}$) from (iv) diffracted to $4.05\ \text{\AA}$ resolution using a synchrotron radiation source (Diamond Beamline I04-1) (Figure 5 - 6b). Crystal growth optimisation was then carried out around the above conditions. Further crystallisation optimisation steps including the use of additive and detergent screens (Hampton Research), manual hanging drop, and micro-seeding were carried out. However, optimisation efforts did not produce crystals with better diffraction quality.

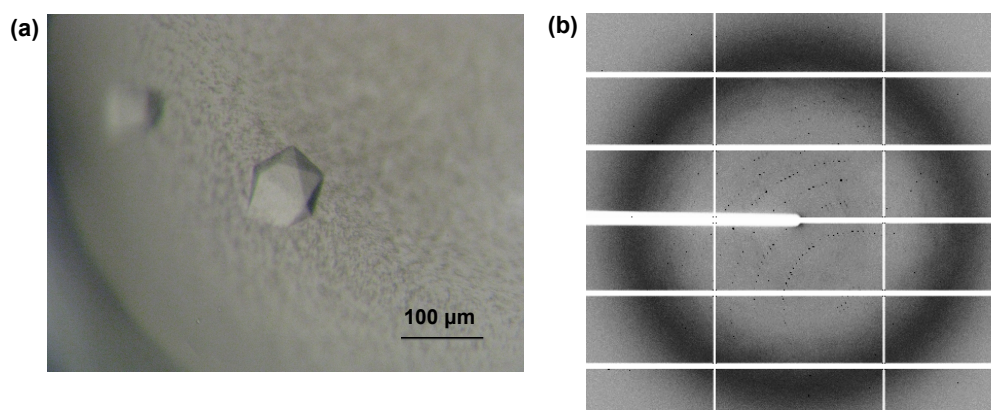


Figure 5 - 6. Crystallographic study on OGFOD2 isoform 2. (a) Crystals of OGFOD2 isoform 2 (O2i2) with the hexagonal bipyramid morphology. **(b)** X-ray diffraction pattern of an O2i2 crystal.

Data were collected using synchrotron radiation (wavelength Å) at 100 K and the space group was processed in $P6_3 2 2$ with the unit cell constants $a = 149.43$ Å, $b = 149.43$ Å, $c = 240.92$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, and $\gamma = 120^\circ$. The calculated Matthews coefficient ($V_m = 2.26$) with a solvent content of 45.67% estimated 5 molecules per asymmetric unit (18-20).

Molecular replacement was performed by searching all possible space groups in the same point group, using an ensemble of prolyl-hydroxylases as a search model. The ensemble consists of *Chlamydomonas reinhardtii* collagen prolyl-hydroxylase (PDB ID 3GZE; resolution: 1.98 Å) (21), human OGFOD1 N-terminal domain (unpublished; resolution: 2.10 Å), *Saccharomyces cerevisiae* TPA1 N-terminal domain (unpublished; resolution: 1.90 Å), human HIF prolyl-hydroxylase 2 (PDB ID 3HQR; resolution: 2.00 Å) (22), and *Pseudomonas putida* prolyl-4-hydroxylase (PDB ID 4IW3; resolution: 2.70 Å). The loops and flexible regions of the ensemble model were deleted to reduce clashes in packing during molecular replacement. However, no correct solutions with all components were found (translation function Z (TFZ) scores <8).

Information from the molecular replacement log files indicated that there were solutions with TFZ scores >8 being rejected because they failed packing tests, suggesting that there could potentially be correct solutions that were not chosen. Subsequent molecular replacement runs were carried out with a lower stringency in packing constraints, i.e. increasing the allowance of percentage atom clashes to 10% and 15% (default setting is 5%). However, this strategy still did not provide any full solution with TFZ>8. Partial solutions (TFZ<8) obtained from this strategy did not appear to provide the correct solution through visual inspection of the models and electron density maps.

The unsuccessful attempts to obtain a molecular replacement solution may be due to a low structural similarity between the search model and the structure of OGFOD2. A structure of a 2OG oxygenase from the lysyl hydroxylase subfamily would be more suitable than those from the prolyl hydroxylases as a search model, but no structures of the lysyl hydroxylases have been reported to date. Moreover, because there were probably 5 molecules in an asymmetric unit, an ensemble of prolyl hydroxylases only account for roughly 10% of the asymmetric unit assuming 50% solvent content, hence the difficulty in obtaining a full solution. Future attempts to obtain a molecular replacement solution could include using different ensembles of structures of 2OG oxygenases as search models. As an alternative phasing strategy, a selenomethionine (SeMet)-derivative OGFOD2 crystal would be desirable for multiwavelength anomalous dispersion (MAD) or potentially for single wavelength anomalous dispersion (SAD) if the diffraction quality of the OGFOD2 crystals could be improved.

Table 5 - 1. Data collection statistics for an OGFOD2 isoform 2 crystal.

Protein	OGFOD2 isoform 2
Radiation Source	Diamond I04-1
Detector	Pilatus 2M
X-ray wavelength (Å)	0.9200
Temperature (K)	100
Resolution Range (Å)	50.00 - 4.05 (4.19 – 4.05) [‡]
Space Group	<i>P</i> 6 ₃ 2 2
Unit Cell Dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	149.43, 149.43, 240.92
α , β , γ (°)	90, 90, 120
Total Number of Reflections Observed	254360
Number of Unique Reflections	13748
Multiplicity	18.5 (13.4) [‡]
Completeness (%)	100.0 (100.0) [‡]
<i>I</i> / σ (<i>I</i>)	21.0 (2.0) [‡]
R_{merge}	14.4
Wilson B Factor (Å ²)	163.78
Software for Data Reduction	HKL-3000
Possible Twin Operators	No twinning suspected

[‡] outermost shell

$$\S R_{\text{merge}} = \frac{\sum |I - \langle I \rangle|}{\sum \langle I \rangle} \times 100$$

5.3 Conclusions

OGFOD2 remains a human 2OG oxygenase of unknown function. Preliminary biochemical and biophysical studies suggest that OGFOD2 has 2OG turnover activity and forms a dimer in solution. Immunofluorescence studies showed that OGFOD2 is localised in the endoplasmic reticulum, similar to its closest relatives, the PLODs, as well as collagen-prolyl-4-hydroxylases (CPHs) (3,9-11).

Preliminary X-ray diffraction data from OGFOD2 crystals showed that it crystallises in a hexagonal space group with an estimated five molecules per asymmetric unit. Further investigations will be needed to unveil the structure and function of OGFOD2.

5.4 Experimental Procedures

5.4.1 General subcloning procedures

Procedure A (polymerase chain reaction amplification)

Specific construct was amplified from the pEGFP-C1_ogfod2 plasmid by polymerase chain reaction (PCR) with forward primer and reverse primer. The PCR reaction mixture contained 1X Pfu buffer (Stratagene®), 0.25 mM dNTP mix, 0.06 U/μL PfuTurbo® DNA polymerase (Stratagene®), 0.50 μM forward primer, 0.50 μM reverse primer, and 17.50 ng/μL pEGFP-C1_ogfod2 as the template.

Procedure B (restriction and ligation)

Insert and target vector were digested with specific restriction enzymes. Restricted insert was ligated with restricted target vector. Ligated plasmid was transformed into *E. coli* XL1 Blue competent cells and selected on appropriate antibiotic-containing LB agar plates. Plasmids containing the insert were verified by DNA sequencing (Source Bioscience).

5.4.2 Subcloning of Ogfod2 constructs into p3XFLAG-CMV-10 and pET28a vectors

The Ogfod2 isoform 2 (O2i2) construct was amplified from pEGFP-C1_ogfod2 by polymerase chain reaction (PCR) following Procedure A (see Section 5.4.1) and

ligated into p3XFLAG-CMV-10 and pET28a vectors following Procedure B (see Section 5.4.1). See Table 5 – 2 for information on primers and restriction enzymes used for each target vector.

The Ogfod2 isoform 1 (O2i1) construct containing KpnI, XbaI, NdeI and HindIII restriction sites was synthesised by GeneArt (Invitrogen). The construct was ligated into p3XFLAG-CMV-10 and pET28a vectors following Procedure B (See section 5.4.1) using the same restriction enzymes used for O2i2.

Table 5 - 2. Primers and restriction enzymes used for subcloning of ogfod2 constructs into vectors.

Target vector	Primers	Restriction enzymes
p3XFLAG-CMV-10	Forward: 5' TGTCGAGGTACCAATGGTTGAGCTTGAGCAGGAGG 3' Reverse: 5' TGTCGAAGATCTTCAGGTGAGCGCACATACATCC 3'	KpnI (NEB) XbaI (NEB)
pET28a	Forward 5' TGTCGACATATGATGGTTGAGCTTGAGCAGGAGG 3' Reverse 5' TGTCGAAAGCTTTAGGTGAGCGCACATACATCC 3'	NdeI (NEB) HindIII (NEB)

5.4.3 Transfection of pEGFP-N1-O2i2, pEGFP-C1_O2i2, pEGFP-C1, and p3XFLAG-CMV-10-O2i1/O2i2 into HeLa cells, HEK293T cells and U2OS cells for immunofluorescence studies

HeLa cells or HEK293T cells were cultured on cover slips in 12-well plates to 50 % confluency in DMEM media (Lonza). For each well that required transfection, 1 µg plasmid in 100 µl DMEM media and 2 µL lipofectamine (Invitrogen) in 100 µL DMEM were incubated at room temperature for 5 min before being mixed together and incubated for a further 15 min at room temperature. The

transfection mixture was then pipetted dropwise into each well containing cells. Replacement of fresh media was performed 3 h post-transfection.

5.4.4 Fixing and antibody staining of HeLa, HEK293T, and U2OS cells for immunofluorescence studies

HeLa or HEK293T cells were fixed with 4% formaldehyde in phosphate saline buffer (PBS) for 15 min. For cells that required antibody staining, they were permeabilised in 1% Triton X-100 in PBS for 10 min, followed by 1 h of blocking in block solution (10% fetal bovine serum (FBS), 0.2% Tween-20 in PBS), 1 h primary antibody incubation, 3 times 10 min washes with wash solution (1% FBS, 0.2% Tween-20 in PBS), 1 h secondary antibody incubation, and 3 times 10 min washes with wash solution. For nuclear staining, cells were incubated in PBS containing 10 ng/mL DAPI for 5 min. Cover slips containing fixed and stained cells were mounted on glass slides and viewed using a fluorescence microscope at 60X objective magnification.

Table 5 - 3. Primary antibodies used for immunofluorescence studies.

Primary antibody	Host	Company	Product No.
anti-FLAG	mouse	Sigma	F1804
anti-OGFOD2	mouse	Novus Biologicala	H00079676-B01
anti-PDI	rabbit	Abcam	ab3672

Table 5 - 4. Secondary antibodies used for immunofluorescence studies.

Secondary antibody	Host	Company	Product No.
anti-mouse-Cy3	sheep	Jackson ImmunoResearch	515-165-062
anti-rabbit-Cy3	goat	GE Healthcare	PA43004
anti-mouse-FITC	goat	Abcam	ab6785

5.4.5 Large-scale transfection of HEK293T cells with pEGFP-N1-O2i2, pEGFP-C1-O2i2 and pEGFP-C1

The transfection method was based on a polyethylenimine (PEI) transfection method described in Ehrhardt *et al.* (23). For each plasmid (pEGFP-N1-O2i2, pEGFP-C1-O2i2, and pEGFP-C1), HEK293T cells were cultured in 8 dishes with a diameter of 150 mm to 60% cell density. A transfection mixture containing 470 µg plasmid, 2.2 mL polyethylenimine (PEI), and 18 mL DMEM media was incubated for 10 min at room temperature before being pipetted dropwise onto HEK293T cells in all dishes. After 3 h, media was replaced with fresh media. Transfected cells were incubated at 37 °C for 48 h before being harvested.

5.4.6 Co-immunoprecipitation with O2i2 and GFP

HEK293T cells (for each plasmid, 8 x 150 mm culture dishes of cells with 100% cell density) transiently transfected with pEGFP-N1-O2i2, pEGFP-C1-O2i2 and pEGFP-C1 were separately lysed in lysis buffer (10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.5% Nonidet P40, 10 mg/mL DNase I, and 1 mM MgCl₂) with an addition of protease inhibitor cocktail (Sigma), phosphatase inhibitor cocktail 1 and 2 (Sigma), for 30 min on ice. Lysate was centrifuged at 24000 rpm for 10 min at 4 °C. The supernatant was added to anti-GFP antibody beads (GFP-TRAP®; Chromotek) pre-equilibrated with lysis buffer and the mixture was incubated on a slow rotating roller for 1 h at 4 °C to ensure complete binding of GFP to anti-GFP antibody. The mixture was subsequently centrifuged at 2000 rpm for 2 min at 4 °C, and the supernatant discarded. Beads were washed three times with wash buffer (10 mM Tris-HCl, pH 7.5 and 300 mM NaCl). Bound proteins were eluted by addition of 100 µl of 2X SDS-loading buffer and the beads boiled for 10 min at 95 °C. SDS-PAGE was carried out for 30 µL of each sample using a NuPAGE (Invitrogen) 4-12% gradient gel. The NuPAGE gel was then fixed with fixing solution (50% v/v methanol and 10% v/v acetic acid in water), stained with Colloidal coomassie stain (Invitrogen), and destained with water.

5.4.7 In-gel tryptic digestion and peptide extraction

Protein bands of interest were excised from the relevant gel and further cut into smaller pieces between 1 mm³ and 2 mm³. For each protein band, the gel pieces were rinsed with wash solution (50% (v/v) methanol, 5% (v/v) acetic acid) for 3 h. Wash solution was removed and gel pieces were rinsed again with fresh wash

solution overnight. Gel pieces were then dehydrated with acetonitrile for 5 min at room temperature. The acetonitrile was removed and 10 mM dithiothreitol (DTT) was added to reduce proteins for 40 min at room temperature. The DTT was removed and proteins were alkylated with 100 mM iodoacetamide at room temperature for 30 min. The iodoacetamide was removed and the gel pieces were dehydrated with acetonitrile, rehydrated with 100 mM ammonium bicarbonate and dehydrated again with acetonitrile as mentioned above. Trypsin solution (20 ng/ μ L trypsin in 50 mM ammonium bicarbonate) was added to rehydrate the gel pieces on ice for 10 min. Excess trypsin solution was removed and a small amount of 5 mM ammonium bicarbonate was added to the gel pieces. Trypsinolysis was carried out overnight at 37 °C. Tryptic peptides were extracted with extraction buffer (50% v/v acetonitrile, 5% v/v formic acid) 3 times. Solvent was then evaporated and peptides reconstituted with 1% acetic acid. Samples were analysed by LC-MS/MS analysis by Dr Mukram Mackeen, Kessler group, based on a previously reported protocol (24).

5.4.8 LC-MS/MS analysis

The tryptic peptides were subjected to nano-ultra performance liquid chromatography tandem MS analysis (nano-UPLC-MS/MS) using a 75 μ m-inner diameter x 25 cm C18 nanoAcquity™ UPLC™ column (1.7- μ m particle size; Waters) with a gradient of 2–45% solvent B in A (solvent A: 99.9% H₂O, 0.1% HCOOH acid; solvent B: 99.9% MeCN, 0.1% HCOOH acid) for 90 min on a Waters nanoAcquity UPLC system (final flow rate, 250 nl/min; 7000 p.s.i.) coupled to a Q-TOF Premier tandem mass spectrometer (Waters) run in positive ion mode.

MS analysis was performed in data-directed analysis (DDA) mode MS to MS/MS switching at precursor ion counts greater than 10 with a return from MS/MS to MS survey after 1s (MS/MS collision energy is dependent on precursor ion mass and charge state). All raw MS data were processed to generate pkl files using the PLGS software (version 2.3) including deisotoping and deconvolution, which converts masses with multiple charge states to $m/z = 1$. Correction of mass accuracy of the raw data was performed by infusing Glu-fibrinopeptide (200 fmol/ μ l; 700 nl/min flow rate; 785.8426 Da $[M + 2H]^{2+}$) into the mass spectrometer as a lock mass during analysis. MS and MS/MS data were calibrated at intervals of 30 s. MS/MS spectra (peak lists) were searched against human entries in the UniProtKB/Swiss-Prot protein database (Version 2010.07.16; 529,056 sequences) using Mascot version 2.3.01 (Matrix Science) with the following parameters: peptide tolerance, 0.2 Da; $^{13}C = 1$; fragment tolerance, 0.1 Da; missed cleavages, 2; instrument type, ESI-Q-TOF-IMM; fixed modification, carbamidomethylation (C); and variable modifications, deamidation (N, Q) and oxidation (M, D, K, N, P, R and M, C, F, H, W, Y separately). The interpretation and presentation of MS/MS data were performed according to published guidelines (25). Assignments of hydroxylation on sites identified by Mascot were verified by manual inspection of MS/MS spectra. emPAI scores (13) from the Mascot results were used to as a guide to assess the abundance of identified proteins. (Protocol carried out and written by Dr Mukram Mackeen (24).)

5.4.9 Recombinant OGFOD2 expression and purification

The pET28a_o2i1 and pET28a_o2i2 plasmids encoding for hexahistidine-tagged O2i1 and O2i2 were transformed into *E. coli* BL21(DE3). Cells were grown in 7.2 L of 2TY medium (1.6% w/v tryptone, 0.5% w/v NaCl, 1% w/v yeast extract) with 30 µg/mL kanamycin at 37 °C until the optical density of 600 nm (OD₆₀₀) was 0.6. Protein production was induced by 1 mM IPTG at 18 °C for 16 h. Cell pellets were resuspended in binding buffer (20 mM Tris-HCl, pH7.5, 500 mM NaCl, 10 mM imidazole), with addition of 100 µL 1 M MgCl₂, a Roche cOmplete EDTA-free protease inhibitor cocktail tablet and 1 mg DNase (bovine pancreas, grade II, Roche). Cells were lysed on ice by sonication, with a Sonics VibraCell VCX-500 and 13 mm probe at 60% intensity, for 4 times 45 s separated by 45 s breaks. Lysate was centrifuged in a Beckman Avanti J25 centrifuge in a JA-25.50 rotor at 23000 rpm for 20 min at 4 °C. The supernatant was filtered with a 0.20 µm filter and loaded onto a 5 mL HisTrap FF column (GE Healthcare) using an AKTA FPLC system for purification. The column was washed with 120 mL of wash buffer (20 mM Tris-HCl, pH7.5, 500 mM NaCl, 40 mM imidazole). The protein of interest was eluted in a gradient of elution buffer (20 mM Tris-HCl, pH7.5, 500 mM NaCl, 500 mM imidazole) and collected as 5 mL fractions.

O2i2 protein was subjected to his-tag cleavage and further purification. Fractions containing the protein of interest (determined by SDS-PAGE) were pooled and the His₆ tag was cleaved by digestion overnight at 4 °C with thrombin (Novagen), using 1 unit of thrombin per 4 mg of protein, in cleavage buffer (20 mM Tris-HCl, 150 mM NaCl, 2.5 mM CaCl₂, pH 8.4). Protein solution was concentrated and loaded onto a 320 mL gel filtration Superdex 75 column (Pharmacia) containing S75 gel filtration resin (Amersham Biosciences) which

had been pre-equilibrated with gel filtration buffer (100 mM Tris-HCl, 100 mM NaCl, pH 7.5). Protein was eluted with 1 column volume of gel filtration buffer, with 5 mL fractions collected. SDS-PAGE was used to determine the fractions containing the protein of interest. The relevant fractions were pooled and concentrated with a 10 kDa MWCO ultrafiltration device (Amicon) to 15 mg/mL and stored at -80 °C.

For crystallisation, protein was first purified using nickel affinity as described above but eluted with 20 mM Tris-HCl pH 7.5, 500 mM NaCl and 350 mM imidazole. The eluted protein solution was then treated with 100 mM EDTA for 1 h and buffer exchanged using PD-10 columns (GE Healthcare) into 25 mM Tris-HCl pH 7.5. The protein was then further purified using a 20 mL MonoQ anion exchange column with a gradient of buffer A (25 mM Tris-HCl pH 7.5) and 50% of buffer B (25 mM Tris-HCl pH 7.5, 1 M NaCl). Fractions corresponding to the two main peaks were separately pooled and buffer exchanged into 25 mM Tris-HCl pH 7.5 using PD-10 columns (GE Healthcare) and concentrated to 10 – 15 mg/mL for storage at -80 °C.

5.4.10 2OG uncoupled turnover assays

A 100 µl solution containing 20 µM of O2i2, 50 µM ammonium Fe(II) sulphate, 500 µM 2OG, 4 mM sodium *L*-ascorbate, and 50 mM Tris-HCl pH 7.5 was prepared. A control sample without OGFOD2 was also set up. Both samples were incubated at 37 °C for 90 min. Proton NMR detection of 2OG to succinate conversion was performed using an Avance II 500 MHz machine with water suppression of proton spectrum (by Dr Ivanhoe Leung).

5.4.11 *In vitro* assays for substrate identification

Reaction mixtures containing a final concentration of 5 μM O2i1/O2i2, 100 μM ascorbate, 100 μM ammonium Fe(II) sulphate, 20 μM peptide (see Table 5 - 5 for peptide sequence), and 100 μM 2OG, and 50 mM Tris-HCl pH 7.5 were set up. Control reactions without enzyme were also set up. The reaction mixtures were incubated at 37 °C for 1 h, quenched with equivolume of methanol, and analysed by MALDI mass spectrometry.

Table 5 - 5. Peptides used for *in vitro* assays with OGFOD2.

No.	Sequence	Molecular Weight (Da)	Protein
P576	PRLHFFAPGFAPLTARGSQQ	2198.7	tubulin
P577	HFFMPGFAPLTARGSQQYRA	2282.8	tubulin
P578	ORLHFFMPGFAPLTARGSQQ	2258.8	tubulin
P579	LPGFAPLTARGSQQYRALTV	2146.7	tubulin
P580	FMPGFAPLTARGSQQYRALT	2212.8	tubulin
P678	PAPKPAPKPAPKPAPKPAPK	1985.5	collagen
P679	SPKPPSPKPPSPKPPSPKPP	2044.4	collagen
P258	PPPPPPPPPPPPPPPPPPPP	1862.9	collagen
P259	SPSPSPSPSPSPSPSPSPSP	1860	collagen
P260	PPGPPGPPGPPGPPGPPGPP	1720	collagen
P300	ACTGWAAGIPGHPGHNGAPG	1828.2	adiponectin
P301	IPGHPGHNGAPGRDGRDGT	1965.3	adiponectin
P302	TPGEKGEKGDPLIGPKGDI	1965.6	adiponectin
P303	EKGDPLIGPKGDIGETGVP	1936.5	adiponectin
P304	ETGVPGAEGPRGFPGUQGRK	2010.5	adiponectin
P305	PGAEGPRGFPGIQGRKGEPG	1964.5	adiponectin
P306	TGCYGIPGAPGLPGAPGKDG	1785.4	complement C1q
P307	GKDG YDGLPGPKGEPGIPAI	1938.6	complement C1q
P308	GPKGEPGIPAIPGIRGPKGQ	1926.6	complement C1q
P309	PKGQKGEPGLPGHPGKNGPA	1923.4	complement C1q
P310	GKNGPAGPPGAPGVPGPAGI	1668.2	complement C1q
P311	PGVPGPAGIPGEPGEEGRYK	1964.4	complement C1q
P312	KKGEAGRPGRGRPLKGEQ	2134.8	complement C1q
P313	GLKGEQGE PGAPGIRTGIQG	1922.5	complement C1q
P314	QGLKGDQGE PGPSGNPGKVG	1879.3	complement C1q
P315	SGNPGKVGYPGPSGPLGARG	1825.4	complement C1q
P316	PSGPLGARGIPGIKGTKGSP	1847.6	complement C1q
	IKGIKGIK	913.2	collagen
	GTKGEKGE	804.9	mannose binding protein (MBP)
	GPKGQKGD	785.9	MBP
	DTKGEKGE	862.9	MBP
	GTKEEKGE	876.9	MBP
	GEKGEKGD	818.8	adiponectin

5.4.12 Crystallisation of OGFOD2

OGFOD2 was buffer exchanged into 25 mM Tris-HCl, pH 7.5 using BioRad desalting spin columns. A protein solution of 8 mg/mL OGFOD2, 2 mM 2,4-pyridine dicarboxylate (2,4-PDCA), and 1 mM NiCl₂ was prepared. The protein solution was mixed with crystallisation drops from the JCSG plus screen (Molecular Dimensions) in 1:2, 1:1, and 2:1 ratios and incubated at 293 K. Crystals from well H2 subwell 2 (conditions: 1 M ammonium sulphate, 0.1 M bis-tris pH 5.5, 1% PEG 3350) were cryoprotected by soaking them in well solution diluted with 25% glycerol or 25% ethylene glycol for approximately 30 s and cooled in liquid nitrogen.

5.4.13 Data Collection and Processing

Data were collected at Diamond Light Source I04-1 and processed using HKL3000 (26). Molecular replacement was attempted using Phaser (27), a subroutine of PHENIX (28). A search model was built using Ensembler (27) with the structures of four prolyl hydroxylases: *Chlamydomonas reinhardtii* collagen prolyl hydroxylase (PDB ID 3GZE; resolution: 1.98 Å) (21), human OGFOD1 N-terminal domain (unpublished; resolution: 2.10 Å), *Saccharomyces cerevisiae* TPA1 N-terminal domain (unpublished; resolution: 1.90 Å), human HIF prolyl hydroxylase 2 (PDB ID 3HQR; resolution: 2.00 Å) (22), and *Pseudomonas putida* prolyl-4-hydroxylase (PDB ID 4IW3; resolution: 2.70 Å).

5.5 References

1. Prockop, D.J. and Kivirikko, K.I. (1995) Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem*, **64**, 403-434.
2. Hautala, T., Byers, M.G., Eddy, R.L., Shows, T.B., Kivirikko, K.I. and Myllyla, R. (1992) Cloning of human lysyl hydroxylase: complete cDNA-derived amino acid sequence and assignment of the gene (PLOD) to chromosome 1p36.3---p36.2. *Genomics*, **13**, 62-69.
3. Myllyharju, J. (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol*, **22**, 15-24.
4. Passoja, K., Rautavuoma, K., Ala-Kokko, L., Kosonen, T. and Kivirikko, K.I. (1998) Cloning and characterization of a third human lysyl hydroxylase isoform. *Proc Natl Acad Sci U S A*, **95**, 10482-10486.
5. Valtavaara, M., Papponen, H., Pirttila, A.M., Hiltunen, K., Helander, H. and Myllyla, R. (1997) Cloning and characterization of a novel human lysyl hydroxylase isoform highly expressed in pancreas and muscle. *J Biol Chem*, **272**, 6831-6834.
6. Aravind, L. and Koonin, E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol*, **2**, RESEARCH0007.
7. Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F.H., Goehler, H., Stroedicke, M., Zenkner, M., Schoenherr, A., Koeppen, S. *et al.* (2005) A human protein-protein interaction network: a resource for annotating the proteome. *Cell*, **122**, 957-968.
8. Kellokumpu, S., Sormunen, R., Heikkinen, J. and Myllyla, R. (1994) Lysyl hydroxylase, a collagen processing enzyme, exemplifies a novel class of lumenally-oriented peripheral membrane proteins in the endoplasmic reticulum. *J Biol Chem*, **269**, 30524-30529.
9. Suokas, M., Lampela, O., Juffer, A.H., Myllyla, R. and Kellokumpu, S. (2003) Retrieval-independent localization of lysyl hydroxylase in the endoplasmic reticulum via a peptide fold in its iron-binding domain. *Biochem J*, **370**, 913-920.
10. Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M. and Myllyla, R. (2000) Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. *J Biol Chem*, **275**, 36158-36163.
11. Myllyla, R., Wang, C., Heikkinen, J., Juffer, A., Lampela, O., Risteli, M., Ruotsalainen, H., Salo, A. and Sipila, L. (2007) Expanding the lysyl hydroxylase toolbox: new

- insights into the localization and activities of lysyl hydroxylase 3 (LH3). *J Cell Physiol*, **212**, 323-329.
12. Noiva, R. (1999) Protein disulfide isomerase: the multifunctional redox chaperone of the endoplasmic reticulum. *Semin Cell Dev Biol*, **10**, 481-493.
 13. Ishihama, Y. (2005) Exponentially Modified Protein Abundance Index (emPAI) for Estimation of Absolute Protein Amount in Proteomics by the Number of Sequenced Peptides per Protein. *Molecular & Cellular Proteomics*, **4**, 1265-1272.
 14. Geiger, T., Wehner, A., Schaab, C., Cox, J. and Mann, M. (2012) Comparative Proteomic Analysis of Eleven Common Cell Lines Reveals Ubiquitous but Varying Expression of Most Proteins. *Molecular & Cellular Proteomics*, **11**.
 15. Hausinger, R.P. (2004) FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol*, **39**, 21-68.
 16. Hopkinson, R.J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K.H., Li, X.S., Kristensen, J.B.L., King, O.N.F., Chan, M.C. *et al.* (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci*, **4**, 3110-3117.
 17. Rose, N.R., McDonough, M.A., King, O.N., Kawamura, A. and Schofield, C.J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev*, **40**, 4364-4397.
 18. Kantardjieff, K.A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci*, **12**, 1865-1871.
 19. Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.
 20. Matthews, B.W. (1976) X-Ray Crystallographic Studies of Proteins. *Annu Rev Phys Chem*, **27**, 493-523.
 21. Koski, M.K., Hieta, R., Hirsila, M., Ronka, A., Myllyharju, J. and Wierenga, R.K. (2009) The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem*, **284**, 25290-25301.
 22. Chowdhury, R., McDonough, M.A., Mecinovic, J., Loenarz, C., Flashman, E., Hewitson, K.S., Domene, C. and Schofield, C.J. (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure*, **17**, 981-989.

23. Ehrhardt, C., Schmolke, M., Matzke, A., Knoblauch, A., Will, C., Wixler, V. and Ludwig, S. (2006) Polyethylenimine, a cost-effective transfection reagent. *Signal Transduction*, **6**, 179-184.
24. Mackeen, M.M., Kramer, H.B., Chang, K.H., Coleman, M.L., Hopkinson, R.J., Schofield, C.J. and Kessler, B.M. (2010) Small-molecule-based inhibition of histone demethylation in cells assessed by quantitative mass spectrometry. *J Proteome Res*, **9**, 4082-4092.
25. Taylor, G.K. and Goodlett, D.R. (2005) Rules governing protein identification by mass spectrometry. *Rapid Communications in Mass Spectrometry*, **19**, 3420-3420.
26. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol*, **276**, 307-326.
27. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
28. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*, **66**, 213-221.

Chapter 6

Structural and biochemical studies on a human collagen prolyl-4-hydroxylase

6.1 Introduction

Collagen, the most abundant protein in mammals, is the predominant structural component in connective tissues (1-4). The formation of the collagen triple helical structure is aided by various post-translational modifications including *trans* C4-hydroxylation, glycosylation and peptide cleavage, all of which occur in the endoplasmic reticulum (ER) (2). Hydroxylation of proline-residues has been shown to stabilise the triple helical structure of collagen (5). Vitamin C (ascorbate), Fe(II) and 2OG were all shown to be essential for the enzyme collagen prolyl-4-hydroxylase (CPH) to catalyse the formation of hydroxyproline at the repeated Gly-X-Y motif in collagen (Y= 4-hydroxyproline) (6-8). Thus CPH was the first identified human Fe(II) and 2OG dependent oxygenase. Recently, CPH was also shown to promote tumour metastasis, suggesting that its inhibition is worth exploring as a potential therapy for cancer (9).

Previous kinetic studies that measured the production of $^{14}\text{CO}_2$ as a consequence of 2OG turnover in the presence of collagen peptides using either partially purified CPH or cell lysate showed that the K_m values for collagen peptides are dependent on peptide length (10-14). To date, direct observation of *in vitro* hydroxylation of collagen peptides by highly purified human CPH has not been reported.

Both the human and mouse CPH α -subunits form an $\alpha_2\beta_2$ heterotetramer with the β -subunit, protein disulphide isomerase (PDI) (15-17). Three isoforms of the α -subunit, CPH1-3, have been found in humans and mice (14). The CPH1 α -subunit consists of three domains: the N-terminal dimerisation domain (N domain), the peptide substrate binding (PSB) domain, and the catalytic (CAT) domain (18,19). Structures of the N domain and the PSB domain of human CPH in complex with a collagen peptide have been determined (19,20) but structures of the human CAT domain or the $\alpha_2\beta_2$ tetramer complex remain elusive. However, crystal structures of related CPH CAT domain-like oxygenases from *Chlamydomonas reinhardtii* and *Bacillus anthracis* have been determined (21,22).

A structure of a CPH $\alpha_2\beta_2$ heterotetrameric complex would provide insightful information on its mechanisms of substrate recognition and hydroxylation. This chapter describes work on crystallographic and *in vitro* biochemical studies of the human CPH1 $\alpha_2\beta_2$ heterotetramer complex.

6.2 Results and Discussion

All experiments and analyses described in this chapter were conducted in collaboration with John Scotti.

6.2.1 Expression and Purification

Successful production of human CPH1 in *E. coli* Origami strains using codon optimised genes has been previously reported (23,24). To obtain codon optimised genes encoding for human collagen prolyl-4-hydroxylase subunit α -1 (referred to as just CPH) (Uniprot ID P13674) and human protein disulphide

isomerase (PDI), the gene constructs encoding CPH and PDI, both omitting the N-terminal ER signal peptides, were designed and synthesized (GeneArt, Invitrogen). These constructs were then sub-cloned into a single pETDuet-1 vector for co-expression, both with N-terminal hexahistidine tags, in *E. coli* Origami 2 cells. Preliminary expression trials resulted in production of soluble protein for both constructs when the *E. coli* cells were grown at 18 °C overnight after induction at OD₆₀₀ = 0.8 with 0.5 mM IPTG. A large-scale (7.2 L culture media) expression was carried out using the conditions stated above.

CPH was initially purified using nickel affinity chromatography (Figure 6 - 1a), followed by Superdex 200 (S200) size exclusion chromatography. Size exclusion chromatography showed a broad elution profile consisting of three unresolved peaks (Figure 6 - 1b), which suggests a heterogeneous mixture of α - and β -subunits.

An alternative purification strategy was therefore needed to separate the CPH $\alpha_2\beta_2$ heterotetramer from both the α - and β -subunits mixture. Drawing on experience from previous successes using heparin affinity chromatography in purifying FTO and ALKBH5 (see Chapters 3 and 4), heparin affinity chromatography was attempted to further purify CPH after nickel affinity chromatography. Because both CPH and PDI have pI < 7 (overall negatively charged at pH 7.5), it was considered unlikely that they would bind to the heparin column due to the negatively charged sulfated glycosaminoglycan. As expected, the majority of the protein eluted from nickel affinity chromatography flowed through the heparin column without binding. However, to our surprise, a fraction of the protein did bind to the column, and was eluted using a gradient 0 – 500 mM NaCl (Figure 6 - 1c).

Protein fractions that eluted from the heparin column were then subjected to additional purification using MonoQ anion exchange chromatography (Figure 6 - 1d) followed by S200 size exclusion chromatography (Figure 6 - 1e). In contrast with the protein obtained from the first purification using nickel affinity chromatography followed by S200 chromatography, the sample derived from S200 chromatography following heparin affinity chromatography and MonoQ anion exchange eluted as a single symmetrical peak (Figure 6 - 1e), indicating that the eluted protein was homogenous. The protein fractions that did not bind to the heparin column were further purified using MonoQ anion exchange chromatography, resulting in two peaks (Figure 6 - 1f).

Native PAGE analysis was conducted on the three protein fractions mentioned above using bovine serum albumin (BSA) as a molecular weight standard. The heparin-bound protein was observed at a higher molecular weight than the heparin flow-through proteins, similar to the BSA tetramer (266 kDa), indicating that it was likely to be a tetramer (~240 kDa) (Figure 6 - 1g). Both fractions from the heparin flow-through, following separation by MonoQ anion exchange, showed two bands close to the BSA monomer (66.5 kDa), suggesting that they consisted of an uncomplexed mixture of CPH and PDI monomers (both ~60 kDa) (Figure 6 - 1g).

Analyses using size exclusion chromatography and native PAGE indicate that the CPH tetramer was highly pure. This CPH tetramer protein was used for subsequent biochemical and crystallographic studies.

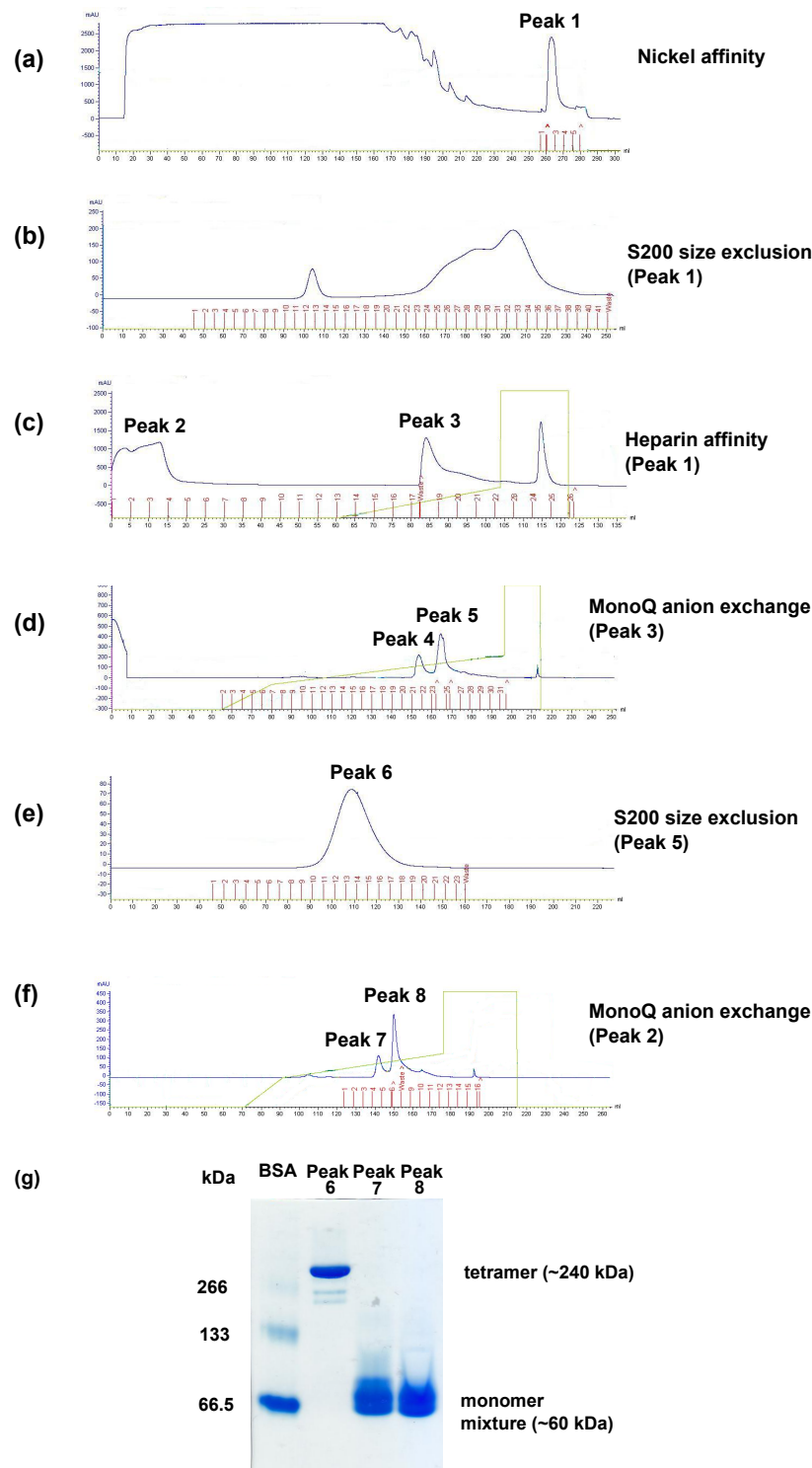


Figure 6 - 1. Purification of CPH tetramer complex. UV chromatogram of **(a)** nickel affinity purification, **(b)** S200 size exclusion chromatography of fractions from Peak 1 in nickel affinity purification, **(c)** heparin affinity of fractions from Peak 1 in nickel affinity purification, **(d)** MonoQ anion exchange of fractions from Peak 3 in heparin affinity, **(e)** S200 size exclusion chromatography of Peak 5 from MonoQ anion exchange, and **(f)** MonoQ anion exchange of fractions from Peak 2 in the heparin affinity flow through. Peak number from where input protein sample was derived is stated in parenthesis. **(g)** 8% native PAGE gel showing that protein from Peak 6 ran as a tetramer complex while protein from Peaks 7 and 8 ran as monomer mixtures of CPH and PDI, which are both ~60 kDa. Bovine serum albumin (BSA) was used for molecular weight standards (kDa).

6.2.2 *In vitro* hydroxylation of collagen peptides

Previously reported kinetic and biochemical studies of CPH were performed by measuring radioactive $^{14}\text{CO}_2$ release from enzymatically coupled radiolabelled 2OG turnover (10-14), which is low throughput and does not directly measure the hydroxylase activity of CPH. Moreover, all reported assays for kinetic studies were conducted with partially purified CPH from insect cell lysates (10-14).

In order to develop a more efficient assay, a hydroxylation assay that detects collagen peptide hydroxylation using fully purified recombinant CPH tetramer was developed. To obtain suitable peptide substrates, we looked for regions of pro-collagen (collagen alpha-1(I)) that contained different numbers of potential hydroxylation sites (X-Pro-Gly), and designed three peptides that have 10 (HCOL1), 6 (HCOL3) and 1 (HCOL5) potential hydroxylation sites respectively (see Figure 6 - 2 for peptide sequences). HCOL1 has 10 repeated proline-proline-glycine (PPG)₁₀ and thus contains 10 potential hydroxylation sites; HCOL3 has 2 groups of 3 hydroxylation sites separated by 6 amino acids in between, which was designed to bind across the PSB domain and the CAT domain of CPH; HCOL5 contains a single hydroxylation site, which would be useful for quantification of enzymatic activity.

The HCOL1, HCOL3 and HCOL5 peptides were incubated with the purified CPH-PDI tetramer complex in the presence of 2OG, ascorbate and Fe(II). Controls without enzyme were also set up in parallel. Matrix assisted laser desorption ionisation-time of flight mass spectrometry (MALDI-ToF MS) was employed to analyse the hydroxylation reactions. MALDI-ToF analysis recorded mass shifts corresponding to either a single +16 Da or multiples of +16 Da for all three peptides, which were not present in the controls without enzyme. These results

indicated that CPH hydroxylated all three peptides. More specifically, mass peaks corresponding to HCOL1 hydroxylated at 1 to 9 sites could be observed (Figure 6 - 2a); for HCOL3, peptide mass peaks corresponding to 2 and 3 hydroxylations were observed (Figure 6 - 2b); and a peptide mass peak corresponding to HCOL5 with a single hydroxylation was observed (Figure 6 - 2c). Purified proteins from the heparin flow-through fractions that were a mixture of CPH and PDI also showed hydroxylase activity against the three peptides, but they appeared to be less active than CPH tetramer (Figure 6 - 2).

Kinetic studies were then carried out on the CPH tetramer by quantifying the rate of reaction when varying the concentrations of the assay components: 2OG, ascorbate, Fe(II), and HCOL5 peptide substrate (Figure 6 - 3). Kinetic analyses showed that the activity of the CPH tetramer was strongly Fe(II) and ascorbate-dependent; in both cases, activity of CPH was not detected in the absence of Fe(II) or ascorbate. Michaelis-Menten analysis of 2OG resulted in a K_m of $38.2 \pm 5.5 \mu\text{M}$. The K_m of the HCOL5 peptide substrate was not determined as it did not reach saturation at a concentration of 1 mM (1:1000 enzyme:substrate ratio).

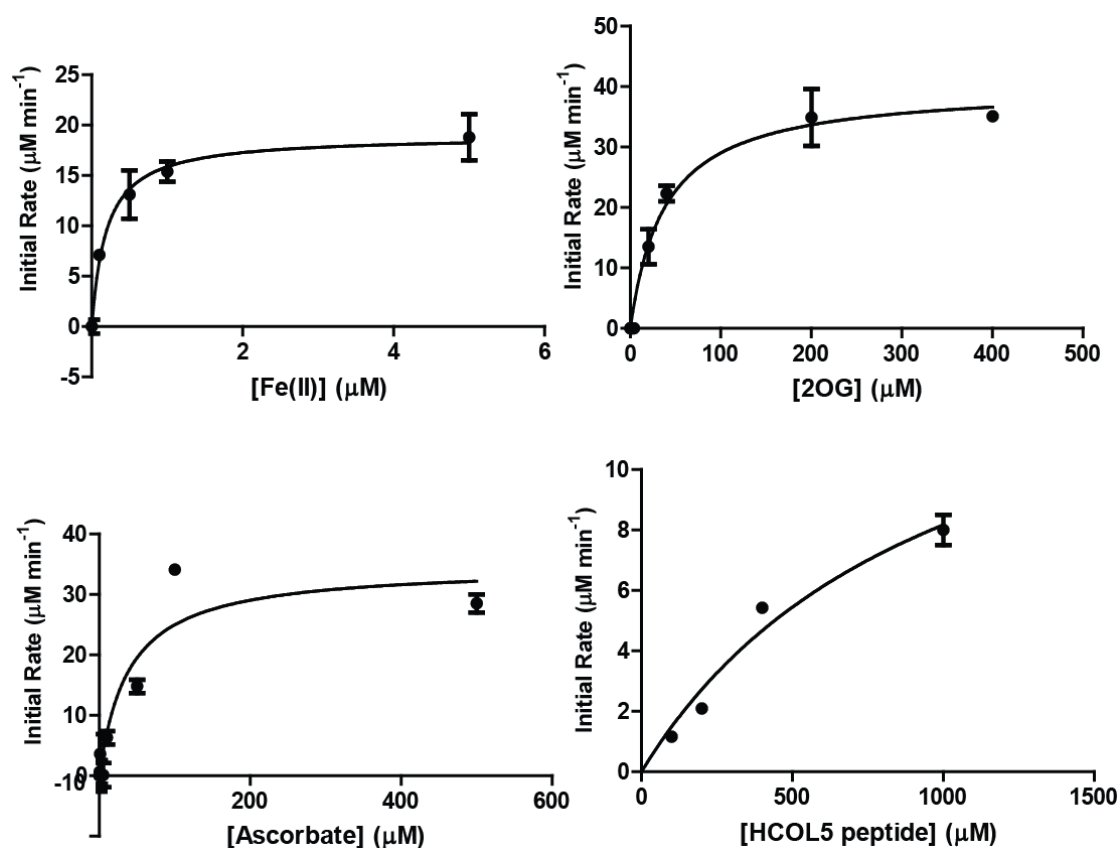


Figure 6 - 3. Kinetic studies of CPH-PDI tetramer. Michaelis-Menten plots of (a) Fe(II), (b) 2-oxoglutarate (2OG), (c) ascorbate, and (d) HCOL5 peptide. K_m of 2OG is $38.2 \pm 5.5 \mu\text{M}$ and K_m of HCOL5 was not determined as saturation did not occur even at 1 mM HCOL5 peptide concentration, possibly due to binding of multiple peptide substrate molecules.

Biochemical and kinetic studies using purified CPH enzyme and detection of hydroxylated collagen peptides are unprecedented because previous reported biochemical studies were indirectly quantified through the 2OG turnover activity of CPH (10-14). These results also provided direct evidence to the function of CPH as a collagen prolyl-hydroxylase. As CPH is involved in various diseases such as breast cancer and fibrosis (2,9), there is significant interest in the scientific and pharmaceutical community to develop inhibitors against CPH. A recent study demonstrated in mouse models and in cells that metastatic breast cancer tumour cells have upregulated levels of CPH via the HIF pathway (9). More importantly, treatment of breast cancer tumour cells with a CPH inhibitor, ethyl 3,4-dihydroxybenzoate, resulted in a significant decrease in collagen content and lung metastasis despite only a modest effect on primary tumour growth reduction (9). This report on the role of CPH in cancer metastasis and the potential therapeutic effect of CPH inhibition has further strengthened the need to develop selective inhibitors for CPH. While ethyl 3,4-dihydroxybenzoate appeared to have a desired effect on retarding metastasis, it is notable that Gilkes *et al.* established a tumour cell line to mitigate confounding effects that ethyl 3,4-dihydroxybenzoate might have on the HIF prolyl-hydroxylases (PHD1-3) (9). Therefore, inhibitors that are selective for CPH are needed if there is more evidence showing that CPH is a viable therapeutic target for cancer treatment. The initial biochemical work on CPH described in this chapter will lead to an *in vitro* inhibition assay for CPH.

6.2.3 Dynamic Scanning Fluorometry

To identify stabilising components to aid in crystallisation of the CPH $\alpha_2\beta_2$ tetramer, a dynamic scanning fluorometry (DSF) assay was performed. CPH $\alpha_2\beta_2$ tetramer with various combinations of Mn(II) (an unreactive Fe(II) substitute), NOG (a broad spectrum 2OG oxygenase inhibitor), and collagen peptides were used for DSF analysis. The results indicated that Mn(II) and NOG increase CPH stability ($\Delta T_m = 1.1$ °C and 5.9 °C, respectively), and that the addition of both Mn(II) and NOG increased stabilisation of the CPH tetramer ($\Delta T_m = 6.5$ °C) more than either component alone (Figure 6 - 4). However, addition of peptides did not stabilise CPH. In conclusion, the results from the DSF assay suggest that addition of Mn(II) and NOG stabilises the CPH tetramer and might aid in crystallisation.

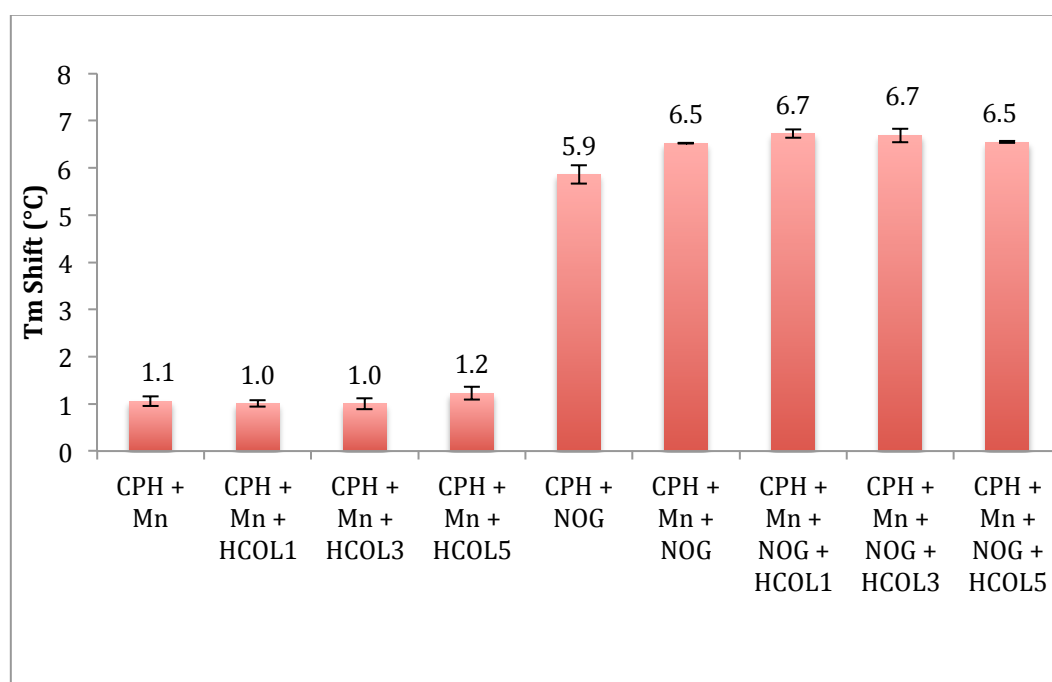


Figure 6 - 4. Dynamic scanning fluorometry analysis of CPH. Bar chart showing thermal stability shift of CPH in the presence of Mn(II), NOG, and collagen peptide substrates (HCOL1, HCOL3 and HCOL5) relative to CPH alone control. Error bars indicate standard deviations from triplicate experiments.

6.2.4 Crystallographic studies

Crystallisation trials of the CPH $\alpha_2\beta_2$ tetramer and monomer mixture of α - and β -subunits were carried out in the presence of Mn(II), NOG, and HCOL5 peptide using conditions from the following commercial broad screens at both 20 °C and 4 °C: JCSG+ (Molecular Dimensions, MD), PACT Premier (MD), Index (Hampton Research), MIDAS (MD), and Proplex (MD). No crystals were observed for the monomer uncomplexed mixture. Hexagonal rod shaped crystals (~75 μ m in the longest dimension) from the tetramer samples appeared in a number of conditions after 2 to 3 weeks (Figure 6 - 6a). Crystal growth optimisation was carried out using Hampton Research additives and Silver Bullet screens. The conditions that yielded crystals were:

- (i) JCSG A3.1/2/3: 0.2 M di-ammonium hydrogen citrate, 20% PEG 3350
- (ii) JCSG A2.3: 0.1 tri-sodium citrate pH 5.5, 20% PEG 3000
- (iii) JCSG B9.2/3: 0.1 M citric acid pH 5.0, 20% PEG 6000
- (iv) JCSG H7.3: 0.2 M ammonium sulfate, 0.1 M bis-tris pH 5.5, 25% PEG 3350
- (v) JCSG B9 + Silver Bullet HT G1: 0.16% w/v 5-sulfosalicylic acid dihydrate, 0.16% w/v dodecanedioic acid, 0.16% w/v hippuric acid, 0.16% w/v mellitic acid, 0.16% w/v oxalacetic acid, 0.16% w/v suberic acid, 0.02 M HEPES sodium pH 6.8
- (vi) JCSG B9 + additive F9: 2% benzamidine
- (vii) ProPlex B4.2: 0.2 M NaCl, 0.1 M MES pH 6.5, 10% PEG 4000
- (viii) ProPlex C4.3: 0.2 M lithium sulfate, 0.1 M MES pH 6.0, 20% PEG 4000

A crystal from condition iii diffracted to ~ 4 Å resolution at the Diamond Light Source synchrotron beamline I02 using X-ray wavelength 0.9786 Å. The crystal was cryoprotected by soaking it for approximately 30 s in well solution containing an additional 25% glycerol. Data were collected at 100 K. The space group was processed in *P6* with unit cell constants $a = 174.84$ Å, $b = 174.84$ Å, $c = 111.24$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, and $\gamma = 120^\circ$. The Matthews coefficient ($V_m = 2.1$ Å³/Da) with a solvent content of 39.81% assumes one tetramer per asymmetric unit (25-27).

Molecular replacement was attempted on all possible space groups in the same point group, using structures of the components of the α -subunit consisting of the CPH α -subunit N double domain and an ensemble of prolyl hydroxylases (for the CPH CAT domain), and the β subunit (i.e. oxidised human PDI), as search models (see experimental procedures Section 6.4.6 for details) (19,21,28-30). The oxidised human PDI was used because the CPH protein was produced in *E. coli* Origami 2 (DE3) strain that has an oxidising cellular environment. However, no solutions for all components were found (translation function Z (TFZ) score <8).

Several strategies were adopted in subsequent molecular replacement runs in an attempt to improve search results (see flow chart, Figure 6 - 5). The following strategies were attempted: (i) manual deletion of loops and flexible regions in the ensemble model, (ii) searching for individual CPH α -subunit N-terminal dimerisation domains and the CPH α -subunit PSB domains to take into account domain movements, (iii) searching for individual domains of the oxidised human PDI (β -subunit) (there are four domains – a, b, b', and a') to take into account domain movements, (iv) varying the search order of the

components, (v) changing the packing constraints from allowing 5% of atom clashes to 10%, and (vi) reprocessing the data in lower symmetry space groups including in $C2\ 2\ 2$, $C2$, $P2$ and $P1$ (Table 6 - 1) before reattempting molecular replacement. However, these strategies did not produce a correct solution for all components (TFZ<8).

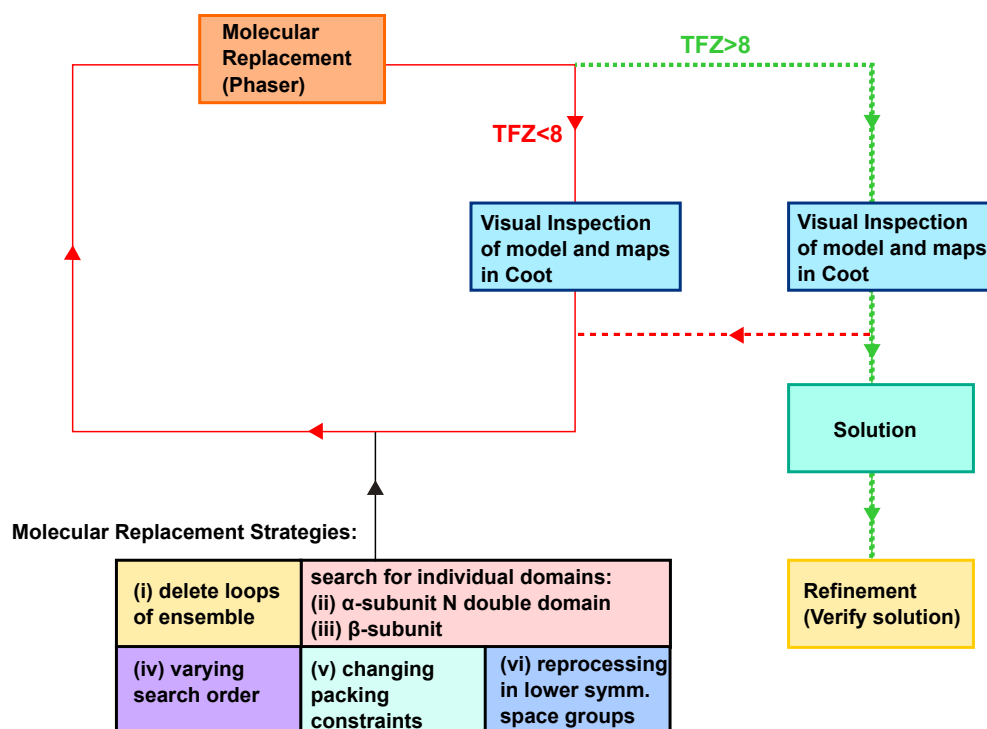


Figure 6 - 5. Flow chart summarising molecular replacement strategies.

A crystal from the crystallisation optimisation attempt (condition vii) diffracted to 3.8 Å resolution (Figure 6 - 6b), the highest resolution diffraction for CPH crystals thus far in this study. The space group processed as $P6_3$ with unit cell constants $a = 177.71$ Å, $b = 177.71$ Å, $c = 111.44$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, and $\gamma = 120^\circ$ (see Table 6 - 1 for data collection statistics). The Matthews coefficient ($V_m = 2.1$ Å³/Da) with a solvent content of 41.86% also assumes one tetramer per asymmetric unit (25-27). Similar molecular replacement strategies (i) - (v) were

employed on this dataset; but as before, they did not yield a correct solution for all components (TFZ<8).

The unsuccessful attempts in obtaining a molecular replacement solution for all components may be due to there being more than 5 components to search for in order to obtain a full solution, with each component consisting of only 10% or less of the asymmetric unit assuming 50% solvent content. Future attempts to obtain a molecular replacement solution could include using different ensembles of structures of 2OG oxygenases as search models for the CPH α -subunit CAT domain. A small-angle X-ray scattering (SAXS) molecular envelope (see Section 6.2.5) may also help in obtaining molecular replacement phases (31-33).

Because molecular replacement was not successful in solving a CPH tetramer structure, alternative phasing methods were attempted. Selenomethionine (SeMet)-derivatised CPH and PDI proteins were produced by co-expression and purified in the same way as native CPH $\alpha_2\beta_2$ tetramer in order to obtain CPH $\alpha_2\beta_2$ tetramer crystals, with selenomethionine replacing methionine, in an attempt to obtain initial phases using single or multiwavelength anomalous dispersion (SAD/MAD) (34-36). However, the CPH $\alpha_2\beta_2$ tetramer was not observed in the SeMet preparation. Another phasing method using halide anomalous signal to obtain phases by MAD was attempted, by soaking native CPH $\alpha_2\beta_2$ tetramer crystals with sodium bromide or 5-amino-2,4,6-triiodoisophthalic acid (I3C) (37,38). However, the soaked crystals either diffracted weakly (maximum resolution 7 Å) or did not diffract X-rays at all.

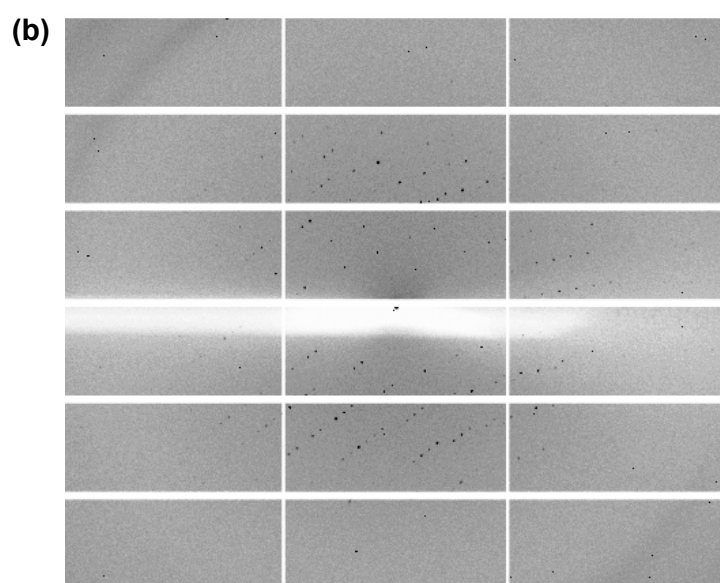
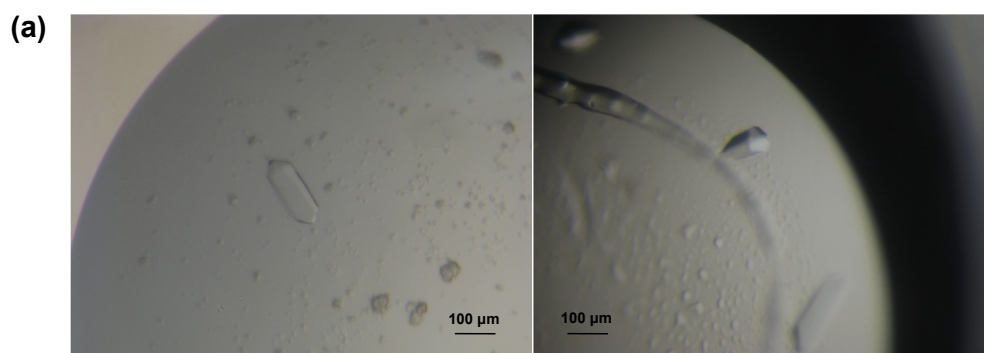


Figure 6 - 6. Crystallographic study on CPH. (a) Hexagonal rod shaped crystals of CPH tetramer. **(b)** X-ray diffraction of a CPH crystal.

Table 6 - 1. Crystallographic data collection statistics for collagen prolyl hydroxylase

Crystallisation Conditions	JCSG+ B9 + Silver Bullet HT G1	JCSG+ B9				
Radiation Source	Diamond I04-1	Diamond I02				
Detector	Pilatus 2M	Pilatus 6M				
X-ray wavelength (Å)	0.9200	0.9786				
Temperature (K)	100	100				
Resolution Range (Å)	50.00 – 3.84 (3.93 – 3.84) [‡]	50.00 – 4.00 (4.14 – 4.00) [‡]	50.00 – 4.20 (4.35 – 4.20) [‡]	50.00 – 4.20 (4.35 – 4.20) [‡]	50.00 – 4.20 (4.35 – 4.20) [‡]	50.00 – 4.20 (4.35 – 4.20) [‡]
Space Group	<i>P</i> 6 ₃	<i>P</i> 6	<i>C</i> 2 2 2	<i>C</i> 2	<i>P</i> 2	<i>P</i> 1
Unit Cell Dimensions						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	177.71, 177.71, 111.44	174.84, 174.84, 111.24	174.99, 304.04, 111.45	302.21, 174.86, 111.17	175.49, 111.51, 175.58	111.30, 175.04, 174.84
α , β , γ (°)	90, 90, 120	90, 90, 120	90, 90, 90	90, 89.81, 90	90, 120.18, 90	59.91, 89.86, 90
Total Number of Reflections Observed	310913	227774	219651	200803	229347	231660
Number of Unique Reflections	19281	16391	22514	42354	43446	81609
Multiplicity	16.1 (15.7) [‡]	13.9 (13.2) [‡]	9.8 (10.2) [‡]	4.7 (4.9) [‡]	5.3 (5.4) [‡]	2.8 (2.9) [‡]
Completeness (%)	100.0 (100.0) [‡]	100.0 (100.0) [‡]	99.9 (99.9) [‡]	98.8 (98.7) [‡]	99.6 (99.6) [‡]	97.2 (96.6) [‡]
<i>I</i> / σ (<i>I</i>)	11.8 (2.0) [‡]	10.6 (1.6) [‡]	5.5 (1.3) [‡]	3.3 (1.1) [‡]	7.0 (1.4) [‡]	4.76 (0.98) [‡]
R_{merge}	21.0	18.8	42.6	39.7	18.1	16.8
Wilson B Factor (Å ²)	124.34	140.49	135.57	133.82	140.17	136.34
Possible twin operators (Number)	Merohedral (1)	Merohedral (1)	Pseudo-merohedral (2)	Pseudo-merohedral (5)	Pseudo-merohedral (5)	Pseudo-merohedral (11)

‡ outermost shell

$$\S R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum \langle I \rangle \times 100$$

6.2.5 Small Angle X-ray Scattering (SAXS) Studies

SAXS data have been previously reported for the CPH N double domain (N-terminal dimerisation domain and PSB domain); they reveal a Z-shaped structure, consistent with crystallographic data of the N double domain (39). Because a structure of the CPH $\alpha_2\beta_2$ tetramer could not be solved by molecular replacement, SAXS was then employed as a means to obtain a molecular envelope to be used to help obtain molecular replacement phases (31-33). (SAXS data collection and preliminary analysis were performed by Dr Edward Lowe, Department of Biochemistry, Oxford).

Based on the autoprocessed SAXS data, an *ab initio* model was built by the DAMMIN software. The model revealed an asymmetrical elongated shape for the CPH $\alpha_2\beta_2$ tetramer with dimensions of $\sim 205 \text{ \AA} \times \sim 100 \text{ \AA} \times \sim 114 \text{ \AA}$ (Figure 6 - 7a,b). The CPH $\alpha_2\beta_2$ complex appears to have two lobes connected by a narrow 'bridge', which is shaped like the letter 'Z'. One of the lobes appears bulkier than the other. A cavity can be observed at the centre of the bulkier lobe (Figure 6 - 7a).

A model of CPH was then built by manually placing the individual components of the α - and β -subunits of CPH, i.e. the N-terminal double domain (dimerisation and substrate binding domains) (PDB ID 4BT8; resolution: 2.20 $\text{\AA}$) (39), the catalytic domain from CrCPH (PDB ID 3GZE; resolution: 1.98 $\text{\AA}$) (21), and human oxidised PDI (PDB ID 4EL1; resolution: 2.88 $\text{\AA}$) (30) into the autoprocessed *ab initio* model (Figure 6 - 7c-f). Unexpectedly, the *ab initio* model

could only accommodate two α -subunits and one β -subunit, forming an $\alpha_2\beta$ heterotrimer, which is not consistent with the well-characterised $\alpha_2\beta_2$ heterotetramer, as deduced from native PAGE analysis reported in the literature (10,12,13,15,16,23) and in Section 6.2.1.

In the SAXS model (Figure 6 - 7), the N-terminal dimerisation domain of CPH α -subunit was modelled onto the central Z-shaped 'bridge', consistent with previously reported crystallographic and SAXS data (39). As such, the CPH α -subunit PSB domain and the CPH α -subunit CAT domain occupy regions adjacent to the bridge at both ends. The bulkier lobe of the *ab initio* model is able to accommodate PDI (β subunit), but the smaller lobe of the *ab initio* model apparently lacked space to accommodate PDI (β -subunit). PDI (β -subunit), the CAT domain and the CPH α -subunit PSB domain were positioned around the cavity in the bulkier lobe, with the active sites of the CPH α -subunit CAT domain and PDI (β -subunit), and the substrate binding site of the CPH α -subunit PSB domain, facing the cavity. Based on this model, together with previous reports that PDI (β -subunit) acts as a procollagen chaperone (40,41) and that the PSB domain binds pro-collagen (19,20), it is proposed that the collagen peptide substrate may be shuttled through the cavity, chaperoned by PDI (β -subunit) and the CPH α -subunit PSB domain into the active site of the CPH α -subunit CAT domain, and subsequently exits through the other end of the cavity.

Whether or not the data obtained from autoprocessed SAXS envelope represent the physiological or crystallised form of CPH remains to be further investigated. It is possible that the $\alpha_2\beta$ trimeric CPH runs on native PAGE gels at a molecular weight equivalent to that of a tetramer (~240 kDa) because of the potentially elongated molecular shape of the CPH trimer, as indicated by the

SAXS model. The other possibility is that the *ab initio* model built by DAMMIN (42) from the autoprocessing software ATSAS (43) did not take into account the symmetry of the $\alpha_2\beta_2$ tetramer, and hence generated an asymmetrical *ab initio* model. Currently, the SAXS data need to be reprocessed with two fold symmetry restraints in place to rule out any possible artefacts.

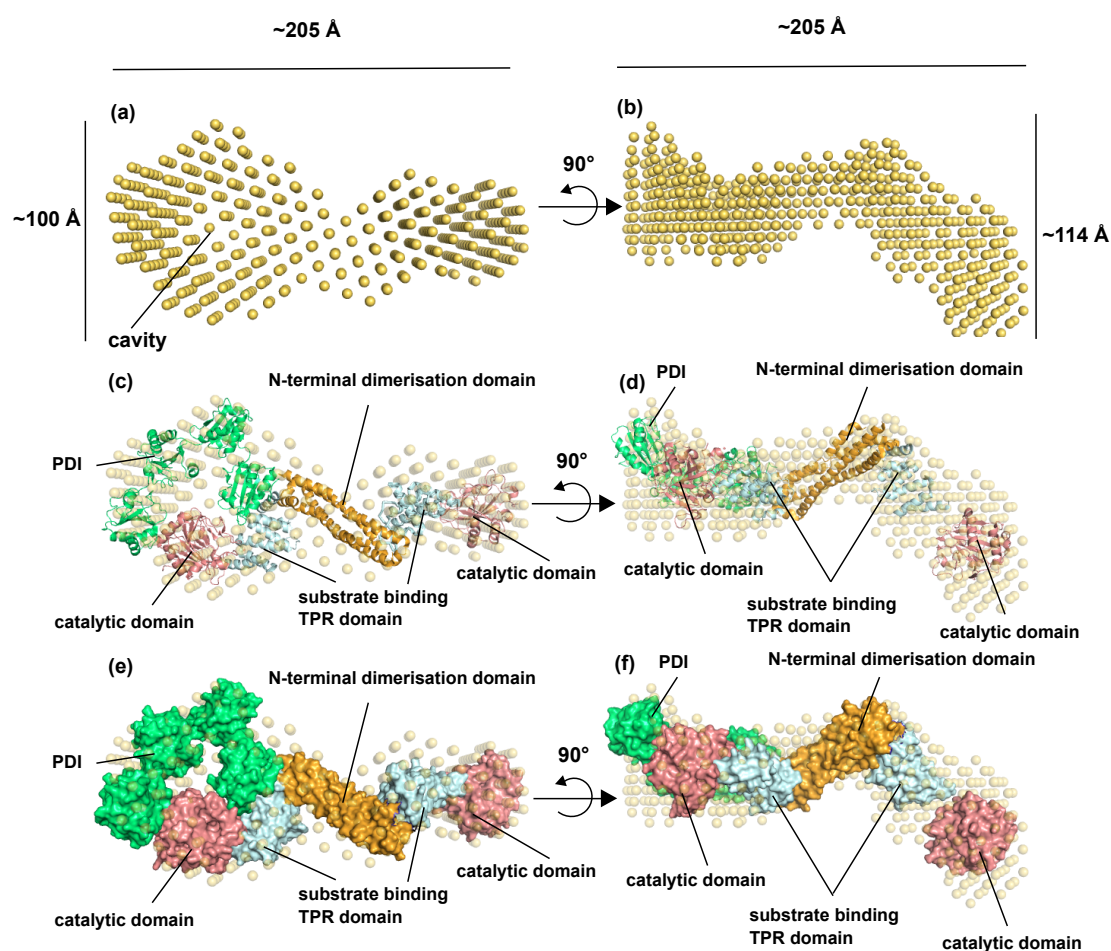


Figure 6 - 7. Small angle X-ray scattering (SAXS) studies on CPH. (a) and (b) *Ab initio* model built by DAMMIN (represented by yellow spheres) based on SAXS data for CPH in two orientations rotated 90° along the X-axis. **(c) and (d)** A cartoon representation model manually placed onto the *ab initio* SAXS model. N-terminal dimerisation domain (PDB ID 4BT8; resolution: 2.20 Å), orange; substrate binding TPR domain (PDB ID 4BT8; resolution: 2.20 Å) (19), light cyan; catalytic domain (PDB ID 3GZE; resolution: 1.98 Å) (21), salmon; human oxidised PDI (PDB ID 4EL1; resolution: 2.88 Å) (30), green. **(e) and (f)** surface representation model manually placed onto the *ab initio* SAXS model. Colour scheme as in (c) and (d). Protein structure models were produced using MacPyMOL v1.3 (44).

6.3 Conclusions

Amidst emerging evidence that CPH is linked to cancer metastasis (9), there is renewed interest in inhibition and functional studies. Although inhibition and biochemical studies on CPH have been ongoing for decades, progress has been somewhat hampered by difficulties in isolating active recombinant enzymes for *in vitro* activity assays. CPH inhibitors identified in early studies such as NOG, 2,5-PDCA, catechols and anthraquinones are 2OG competitors (45). 2,2-bipyridine, another CPH inhibitor, is an iron chelator, and thus also inhibits PHD2 and HIF asparaginyl hydroxylase (FIH) (45-47). Although ethyl 3,4-dihydroxybenzoate was initially identified as a potential selective CPH inhibitor, 3,4-dihydroxybenzoate also inhibits PHD2 (48). All in all, the inhibitors of CPH identified so far have not achieved the desired specificity (45).

This chapter showed that a recombinant form of the CPH $\alpha_2\beta_2$ tetramer can be purified and a biochemical assay to quantify its enzymatic activity was developed. The CPH $\alpha_2\beta_2$ tetramer was crystallised, laying the foundation for obtaining a crystal structure of the $\alpha_2\beta_2$ tetramer, which will provide insights into the mechanisms of collagen prolyl-4-hydroxylation. In an unexpected potential twist, autoprocessed SAXS data of CPH suggests that in solution, CPH may form an $\alpha_2\beta$ trimer instead of an $\alpha_2\beta_2$ tetramer. However, further investigations of these data and the use of other techniques will be needed. In summary, the biochemical and structural work on CPH have brought us a bit closer to structure-based inhibition studies on CPH for potential therapeutic purposes.

6.4 Experimental Procedures

6.4.1 Cloning

Codon-optimised genes encoding for full-length (minus ER signal sequences) human CPH₁₁₈₋₅₃₄ and PDI₁₈₋₅₀₈ (β -subunit) were restricted from pMK-CPH1 and pMK-PDI plasmids (from GeneArt, Invitrogen) using NdeI and KpnI-HF restriction enzymes for CPH and EcoRI-HF and NotI-HF for PDI and subsequently purified using agarose gel electrophoresis. The CPH gene was then ligated into a pETDuet-1 vector (Novagen) using the NdeI and KpnI restriction sites. The PDI gene was inserted into the pETDuet-1-CPH plasmid using the EcoRI and NotI restriction sites. The pETDuet-1-CPH-PDI plasmid was confirmed by DNA sequencing.

6.4.2 Expression and Purification

The pETDuet-1-CPH-PDI plasmid was transformed into *Escherichia coli* Origami 2 (DE3) cells for co-expression. The *E. coli* cells containing the pETDuet-1-CPH-PDI plasmid were grown in 2TY media (or selenomethionine media prepared from premix supplied by Molecular Dimension) at 37 °C until it reached an O.D.₆₀₀ of ~0.8 and were induced by a final concentration of 0.5 mM IPTG. Cell growth continued overnight at 18 °C. The cells were harvested by centrifugation and the pellets stored at -80 °C.

The cell pellets were resuspended in binding buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, and 10 mM imidazole; supplemented by ~1 mg DNaseI and a Roche EDTA-free protease inhibitor cocktail tablet) and lysed by sonication. The lysates were cleared by centrifugation (43,400 g for 20 min), after which the

supernatant was loaded onto a (5 mL) HisTrap HP column (GE Healthcare) and purified using an AKTA FPLC system. The column was washed with wash buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, and 40 mM imidazole). The protein was eluted with 50 mM Tris-HCl pH 7.5, 500 mM NaCl, and 250 mM imidazole. The eluted protein was buffer exchanged into 25 mM Tris-HCl pH 7.5 buffer using PD-10 columns (GE Healthcare) and further purified using a HiTrap Heparin column (GE Healthcare), followed by a (20 mL) MonoQ column. In both cases, the protein was eluted with a gradient from buffer A (25 mM Tris-HCl pH 7.5) to buffer B (25 mM Tris-HCl, 1 M NaCl). The purified protein was then loaded onto an (300 mL) S200 size exclusion column pre-equilibrated with 25 mM Tris-HCl pH7.5 for the final purification step. The purified CPH $\alpha_2\beta_2$ tetrameric complex and monomer mixture proteins were concentrated to 5 – 8 mg/ml and stored at -80 °C for *in vitro* assays or immediately used for crystallisation.

6.4.3 Dynamic Scanning Fluorometry (DSF)

50 μ L of mixtures containing a final concentration of 1.5 μ M CPH $\alpha_2\beta_2$ tetramer protein, 1X Sypro Orange (Invitrogen), 50 mM HEPES pH 7.5, and one or more of the following components (final concentration) were prepared:

- (i) 50 μ M Mn(II)
- (ii) 100 μ M NOG
- (iii) 100 μ M peptide (HCOL1, HCOL3, or HCOL5)

DSF assays were performed using a MiniOpticon™ Real-Time PCR Detection System (Bio-Rad). Fluorescence readings were taken at 25 – 90 °C, increasing the temperature linearly over an 80 min period. Excitation and emission wavelengths

were 470 nm and 570 nm, respectively. Melting curves were analysed using GraphPad Prism[®].

6.4.5 *In vitro* hydroxylation assays

Initial hydroxylation assays were carried out with a final concentration of 5 μ M CPH $\alpha_2\beta_2$ tetramer or monomer (α/β mixture) protein, 50 μ M Fe(II), 1 mM 2OG, 4 mM ascorbate, 100 μ M peptide (HCOL1, HCOL3 or HCOL5), and 25 mM Tris-HCl pH 7.5 at room temperature for 3 h. The reactions were quenched with an equal volume of 20% formic acid and analysed using MALDI-ToF MS. Samples were prepared by mixing 1 μ L of reaction solution with 1 μ L of MALDI matrix saturated solution of α -cyano-4-hydroxycinnamic acid (water : acetonitrile = 1:1)

For kinetic studies, a time course assay was first carried out using 1 μ M CPH tetramer protein, 50 μ M Fe(II), 400 μ M 2OG, 1 mM ascorbate, 100 μ M HCOL5 peptide substrate and 25 mM Tris-HCl pH 7.5. The reactions were quenched with equivolume of 20% formic acid at various time points (2, 5, 10, 20, 30, 45, 60, 90, 120, and 195 min) and analysed by MALDI-ToF mass spectrometry. To investigate the kinetics of each component (with the other components remaining at the same concentration as described in the time course assay), varying concentrations of ascorbate (0 – 500 μ M), 2OG (0 – 400 μ M), and Fe(II) (0 – 100 μ M) were prepared and the reactions were quenched after 20 min before MALDI-ToF MS analyses were carried out. For the kinetic study of HCOL5 peptide substrate, time course assays were conducted using various concentrations of peptide substrates (0 – 1mM) with other conditions remaining as described above. The results were analysed by MALDI-ToF MS and

the initial rates were calculated using linear regions of the time course graphs. Graphs were plotted using GraphPad Prism®.

6.4.6 Crystallographic Studies

Protein samples containing 5 mg/ml CPH $\alpha_2\beta_2$ tetramer or (α/β mixture) monomer, 0.5 mM MnCl_2 , 0.5 mM NOG, and 1 mM HCOL5 were prepared. Samples were then applied to crystallisation drops from broad screens (JCSG+, PACT Premier, MIDAS, and Proplex from Molecular Dimensions and Index from Hampton Research) and optimisation screens (additive and Silver Bullet HT from Hampton Research) in 1:2, 1:1, and 2:1 protein:well solution ratios using a Phenix RE (Art Robbins) crystallisation automation robot, and incubated at 293 K or 277 K. Crystals were cryo-protected by soaking them for approximately 30 s in well solution containing 25% glycerol, then harvested using nylon loops (Hampton Research) or Mylon loops (Mitogen) and flash-cooled in liquid nitrogen. X-ray data were collected at 100 K at Diamond Light Source beamline I02 and I04-1. Data were then processed using HKL-3000 (49). Molecular replacement was carried out using Phaser (29) (subroutine of PHENIX (50)) with structures of the N-terminal CPH dimerisation domain (PDB ID 4BT8; resolution: 2.20 Å), the PSB domain (PDB ID 4BT8; resolution: 2.20 Å) (19), an ensemble (29) of prolyl-hydroxylases and the human oxidised PDI (PDB ID 4EL1; resolution: 2.88 Å) (30) as search models. The ensemble was generated by Ensembler (29) from structures of *Chlamydomonas reinhardtii* collagen prolyl hydroxylases (PDB ID 3GZE; resolution: 1.98 Å) (21), human OGFOD1 N-terminal domain (unpublished; resolution: 2.10 Å; provided by John Scotti and Dr

Shoichiro Horita), *Saccharomyces cerevisiae* TPA1 N-terminal domain (unpublished; resolution: 1.90 Å; provided by John Scotti), human HIF prolyl hydroxylase 2 (PDB ID 3HQR; resolution: 2.00 Å) (28) and *Pseudomonas putida* prolyl-4-hydroxylase (PDB ID 4IW3; resolution: 2.70 Å) (unpublished; provided by John Scotti).

6.4.7 Small Angle X-ray Scattering (SAXS)

A serial 2X dilution of ~3 mg/mL CPH α/β complex was first carried out. SAXS data were collected at the European Synchrotron Radiation Facility BioSAXS beamline BM29 equipped with a Pilatus 1M detector. Data processing, analysis, and molecular modelling were carried out using programmes of the ATSAS package (43) available at the beamline. *Ab initio* models were built using DAMMIN (42). Crystal structures of the components of CPH α - and β - subunits were manually superimposed onto an *ab initio* model generated by DAMMIN to obtain a model for CPH $\alpha_2\beta$ trimer.

6.5 References

1. Di Lullo, G.A. (2001) Mapping the Ligand-binding Sites and Disease-associated Mutations on the Most Abundant Protein in the Human, Type I Collagen. *Journal of Biological Chemistry*, **277**, 4223-4231.
2. Prockop, D.J. and Kivirikko, K.I. (1995) Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem*, **64**, 403-434.
3. Myllyharju, J. (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends in Genetics*, **20**, 33-43.
4. Shoulders, M.D. and Raines, R.T. (2009) Collagen Structure and Stability. *Annu Rev Biochem*, **78**, 929-958.
5. Berg, R.A. and Prockop, D.J. (1973) The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem Biophys Res Commun*, **52**, 115-120.

6. Horton, J.R., Upadhyay, A.K., Qi, H.H., Zhang, X., Shi, Y. and Cheng, X. (2010) Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol*, **17**, 38-43.
7. Peterkofsky, B. and Udenfriend, S. (1965) Enzymatic Hydroxylation of Proline in Microsomal Polypeptide Leading to Formation of Collagen. *Proc Natl Acad Sci U S A*, **53**, 335-342.
8. Prockop, D.J. and Juva, K. (1965) Synthesis of Hydroxyproline in Vitro by the Hydroxylation of Proline in a Precursor of Collagen. *Proc Natl Acad Sci U S A*, **53**, 661-668.
9. Gilkes, D.M., Chaturvedi, P., Bajpai, S., Wong, C.C., Wei, H., Pitcairn, S., Hubbi, M.E., Wirtz, D. and Semenza, G.L. (2013) Collagen prolyl hydroxylases are essential for breast cancer metastasis. *Cancer Res*, **73**, 3285-3296.
10. Kivirikko, K.I. and Myllyharju, J. (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biology*, **16**, 357-368.
11. Kivirikko, K.I. and Myllyla, R. (1982) Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol*, **82 Pt A**, 245-304.
12. Kivirikko, K.I. and Pihlajaniemi, T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv Enzymol Relat Areas Mol Biol*, **72**, 325-398.
13. Kukkola, L., Koivunen, P., Pakkanen, O., Page, A.P. and Myllyharju, J. (2004) Collagen prolyl 4-hydroxylase tetramers and dimers show identical decreases in Km values for peptide substrates with increasing chain length: mutation of one of the two catalytic sites in the tetramer inactivates the enzyme by more than half. *J Biol Chem*, **279**, 18656-18661.
14. Myllyharju, J. (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol*, **22**, 15-24.
15. Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T. and Kivirikko, K.I. (1997) Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer. The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta2 tetramer. *J Biol Chem*, **272**, 17342-17348.
16. Helaakoski, T., Annunen, P., Vuori, K., MacNeil, I.A., Pihlajaniemi, T. and Kivirikko, K.I. (1995) Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase alpha-subunit isoform: formation of an alpha 2 beta 2 tetramer with the protein disulfide-isomerase/beta subunit. *Proc Natl Acad Sci U S A*, **92**, 4427-4431.
17. Kukkola, L., Hieta, R., Kivirikko, K.I. and Myllyharju, J. (2003) Identification and characterization of a third human, rat, and mouse collagen prolyl 4-hydroxylase isoenzyme. *J Biol Chem*, **278**, 47685-47693.
18. Myllyharju, J. and Kivirikko, K.I. (1999) Identification of a novel proline-rich peptide-binding domain in prolyl 4-hydroxylase. *Embo J*, **18**, 306-312.

19. Anantharajan, J., Koski, M.K., Kursula, P., Hieta, R., Bergmann, U., Myllyharju, J. and Wierenga, Rik K. (2013) The Structural Motifs for Substrate Binding and Dimerization of the α Subunit of Collagen Prolyl 4-Hydroxylase. *Structure*, **21**, 2107-2118.
20. Pekkala, M., Hieta, R., Bergmann, U., Kivirikko, K.I., Wierenga, R.K. and Myllyharju, J. (2004) The peptide-substrate-binding domain of collagen prolyl 4-hydroxylases is a tetratricopeptide repeat domain with functional aromatic residues. *J Biol Chem*, **279**, 52255-52261.
21. Koski, M.K., Hieta, R., Hirsila, M., Ronka, A., Myllyharju, J. and Wierenga, R.K. (2009) The crystal structure of an algal prolyl 4-hydroxylase complexed with a proline-rich peptide reveals a novel buried tripeptide binding motif. *J Biol Chem*, **284**, 25290-25301.
22. Culpepper, M.A., Scott, E.E. and Limburg, J. (2010) Crystal structure of prolyl 4-hydroxylase from *Bacillus anthracis*. *Biochemistry*, **49**, 124-133.
23. Neubauer, A., Soini, J., Bollok, M., Zenker, M., Sandqvist, J., Myllyharju, J. and Neubauer, P. (2007) Fermentation process for tetrameric human collagen prolyl 4-hydroxylase in *Escherichia coli*: improvement by gene optimisation of the PDI/beta subunit and repeated addition of the inducer anhydrotetracycline. *J Biotechnol*, **128**, 308-321.
24. Neubauer, A., Neubauer, P. and Myllyharju, J. (2005) High-level production of human collagen prolyl 4-hydroxylase in *Escherichia coli*. *Matrix Biol*, **24**, 59-68.
25. Kantardjieff, K.A. and Rupp, B. (2003) Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci*, **12**, 1865-1871.
26. Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.
27. Matthews, B.W. (1976) X-Ray Crystallographic Studies of Proteins. *Annu Rev Phys Chem*, **27**, 493-523.
28. Chowdhury, R., McDonough, M.A., Mecinovic, J., Loenarz, C., Flashman, E., Hewitson, K.S., Domene, C. and Schofield, C.J. (2009) Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure*, **17**, 981-989.
29. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr*, **40**, 658-674.
30. Wang, C., Li, W., Ren, J., Fang, J., Ke, H., Gong, W., Feng, W. and Wang, C.C. (2013) Structural insights into the redox-regulated dynamic conformations of human protein disulfide isomerase. *Antioxid Redox Signal*, **19**, 36-45.
31. Hao, Q. (2001) Phasing from an envelope. *Acta Crystallogr D Biol Crystallogr*, **57**, 1410-1414.
32. Hao, Q. (2006) Macromolecular envelope determination and envelope-based phasing. *Acta Crystallogr D Biol Crystallogr*, **62**, 909-914.
33. Hao, Q., Dodd, F.E., Grossmann, J.G. and Hasnain, S.S. (1999) Ab initio phasing using molecular envelope from solution X-ray scattering. *Acta Crystallogr D Biol Crystallogr*, **55**, 243-246.

34. Hendrickson, W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51-58.
35. Hendrickson, W.A., Horton, J.R. and LeMaster, D.M. (1990) Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *Embo J*, **9**, 1665-1672.
36. Wang, B.C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol*, **115**, 90-112.
37. Beck, T., Krasauskas, A., Gruene, T. and Sheldrick, G.M. (2008) A magic triangle for experimental phasing of macromolecules. *Acta Crystallogr D Biol Crystallogr*, **64**, 1179-1182.
38. Dauter, Z. and Dauter, N. (1999) Anomalous signal of solvent bromides used for phasing of lysozyme. *J Mol Biol*, **289**, 93-101.
39. Anantharajan, J., Koski, M.K., Kursula, P., Hieta, R., Bergmann, U., Myllyharju, J. and Wierenga, R.K. (2013) The structural motifs for substrate binding and dimerization of the alpha subunit of collagen prolyl 4-hydroxylase. *Structure*, **21**, 2107-2118.
40. Wilson, R., Lees, J.F. and Bulleid, N.J. (1998) Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. *Journal of Biological Chemistry*, **273**, 9637-9643.
41. Kellokumpu, S., Suokas, M., Risteli, L. and Myllyla, R. (1997) Protein disulfide isomerase and newly synthesized procollagen chains form higher-order structures in the lumen of the endoplasmic reticulum. *J Biol Chem*, **272**, 2770-2777.
42. Svergun, D.I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J*, **76**, 2879-2886.
43. Konarev, P.V., Petoukhov, M.V., Volkov, V.V. and Svergun, D.I. (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J Appl Crystallogr*, **39**, 277-286.
44. Schrodinger, LLC. (2010).
45. Rose, N.R., McDonough, M.A., King, O.N., Kawamura, A. and Schofield, C.J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev*, **40**, 4364-4397.
46. Gunzler, V., Majamaa, K., Hanauskeabel, H.M. and Kivirikko, K.I. (1986) Catalytically Active Ferrous-Ions Are Not Released from Prolyl 4-Hydroxylase under Turnover Conditions. *Biochim Biophys Acta*, **873**, 38-44.
47. Banerji, B., Conejo-Garcia, A., McNeill, L.A., McDonough, M.A., Buck, M.R., Hewitson, K.S., Oldham, N.J. and Schofield, C.J. (2005) The inhibition of factor inhibiting hypoxia-inducible factor (FIH) by beta-oxocarboxylic acids. *Chem Commun (Camb)*, 5438-5440.
48. Tsukiyama, F., Nakai, Y., Yoshida, M., Tokuhara, T., Hirota, K., Sakai, A., Hayashi, H. and Katsumata, T. (2006) Gallate, the component of HIF-inducing catechins, inhibits HIF prolyl hydroxylase. *Biochem Biophys Res Commun*, **351**, 234-239.
49. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol*, **276**, 307-326.

50. Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D*, **66**, 213-221.

Chapter 7

Conclusions

Prior to this research, structure-based inhibition studies on the AlkB subfamily of 2OG oxygenases were lacking. While structures of AlkB in complex with 2OG and DNA substrates were available (1-4), other than one structure of FTO in complex with NOG and 3meT (5), no structures of AlkB-inhibitor complexes were reported and the structure of ALKBH5 was unknown. The research described in this thesis has contributed an additional 11 crystal structures of the AlkB subfamily of 2OG oxygenases (AlkB, FTO and ALKBH5) in complex with inhibitors (6-9) to the Protein Data Bank (PDB) (see Table 7 – 1). These structures provide useful information on the binding modes of different classes of 2OG oxygenase inhibitors, and can be used for *in silico* inhibitor screening by the scientific and pharmaceutical communities.

Table 7 - 1. PDB IDs of structures determined in the research described in this thesis.

No.	PDB ID	Protein Name	Ligand	Reference
1	3T4V	<i>E. coli</i> alkylated DNA repair protein AlkB	<i>N</i> -(carboxycarbonyl)- <i>S</i> -(naphthalene-2-ylmethyl)- <i>L</i> -cysteine	(9)
2	3T3Y	<i>E. coli</i> alkylated DNA repair protein AlkB	<i>N</i> -[(3-hydroxypyridin-2-yl)carbonyl]glycine	(9)
3	4JHT	<i>E. coli</i> alkylated DNA repair protein AlkB	5-carboxy-8-hydroxyquinoline	(8)
4	4IDZ	Human fat mass and obesity associated protein, FTO	<i>N</i> -oxalylglycine	(6)
5	4IE0	Human fat mass and obesity associated protein, FTO	2,4-pyridine dicarboxylate	(6)
6	4IE4	Human fat mass and obesity associated protein, FTO	5-carboxy-8-hydroxyquinoline	(6)
7	4IE5	Human fat mass and obesity associated protein, FTO	<i>N</i> -[(3-hydroxypyridin-2-yl)carbonyl]glycine	(6)
8	4IE6	Human fat mass and obesity associated protein, FTO	<i>N</i> -[(1-chloro-4-hydroxyisoquinolin-3-yl)carbonyl]glycine	(6)
9	4IE7	Human fat mass and obesity associated protein, FTO	4,5-dihydroxy-9,10-dioxo-9,10-dihydroanthracene-2-carboxylic acid	(6)
10	4QH0	Human fat mass and obesity associated protein, FTO	(3-hydroxy-6-(naphthalene-1-yl)picolinoyl)glycine	-
11	4NJ4	Human AlkB homologue 5, ALKBH5	<i>N</i> -[(1-chloro-4-hydroxyisoquinolin-3-yl)carbonyl]glycine	(7)

The work to develop inhibitors of AlkB and FTO represents the first structure-based inhibition studies on the AlkB subfamily of 2OG oxygenases (6,9). As described in Chapter 2, proof-of-concept studies on AlkB demonstrated that AlkB can be inhibited by 2OG mimetics and that structure-guided inhibitor optimisation was applicable to AlkB (9). The results subsequently informed the design of inhibitors of FTO. For example, the inclusion of a naphthalene group onto a 2OG mimetic inhibitor to enhance inhibitor potency has since yielded positive results in both AlkB (Chapter 2 compound **18**) and also FTO (Chapter 3 compound **23**). The pyridyl glycinamide inhibitor, *N*-[(3-hydroxypyridin-2-yl)carbonyl]glycine (Chapter 2 compound **22**), which was inaccurately predicted to bind in AlkB with its pyridine ring near the substrate (10), was found to bind to FTO in the way it was originally predicted for AlkB (6,9). Thus, differences in the active sites in the AlkB subfamily of 2OG oxygenases have the potential to

affect inhibitor binding modes, and should be taken into consideration during structure-based inhibitor design.

An important aspect of the research was the successful development of a purification and crystallisation protocol for FTO that readily enables for its co-crystallisation with various classes of inhibitor (Chapter 3). The reported structures of FTO-inhibitor complexes and methods will benefit members of the scientific community who are interested in further developing improved FTO inhibitors, which may potentially be developed into a novel anti-obesity medicine. Building on from this research, efforts can now be directed towards *in silico* drug screening, rational inhibitor design, and organic synthesis, in order to obtain inhibitors that are more potent, selective and cell penetrable.

Structural studies were also conducted on a third 2OG oxygenase from the AlkB subfamily – ALKBH5 (Chapter 4). A crystal structure of ALKBH5 with IOX3 crosslinked to Cys200 was determined and published as a novel structure of human ALKBH5 (11). This structure is significant because ALKBH5 has recently been shown to be an m⁶A ‘eraser’, suggesting that ALKBH5 may play a role in the dynamic regulation of m⁶A on mRNA (12). Based on the structure of ALKBH5, the β IV-V loop of the DSBH is proposed to confer ssRNA selectivity, and a phenylalanine finger (Phe234) from this loop was predicted to induce the m⁶A base to flip into the active site of ALKBH5 (11). Similar conclusions were drawn by other groups that reported structures of ALKBH5 shortly after the ALKBH5-IOX3 crosslinked structure was published (13-15). Mutagenesis studies by Feng *et al.* (2014) on Phe234 to alanine and glutamate reduced the activity of ALKBH5 to 41% and 13.5% respectively (14), thus supporting the prediction that Phe234 is an important residue for the demethylation mechanisms of ALKBH5 (11,14). A

crystal structure of an ALKBH5-RNA substrate complex should provide a clearer answer to the above hypothesis. Additionally, comparison of the active site of ALKBH5 with that of FTO will aid in the design of inhibitors that are selective for FTO.

The function of OGFOD2 is yet to be defined. The results described in Chapter 5 have identified OGFOD2 as an endoplasmic reticulum localised protein and a potential 2OG oxygenase, as predicted by sequence comparison between OGFOD2 and the PLODs (16). Biophysical characterisation showed that OGFOD2 exists as both monomer and dimer in solution, and preliminary X-ray data enabled initial crystallographic characterisation of OGFOD2 crystals. The protein purification methods that were developed will be useful for future structural and biochemical studies in order to identify the substrate and to determine the structure of OGFOD2.

The fifth 2OG oxygenase investigated in the course of this research was the human collagen prolyl hydroxylase, which was also the first identified 2OG oxygenase (17). CPH is known to be challenging with regards to expression, purification and crystallisation (18-21). There were no reported *in vitro* assays that directly measure hydroxylation of collagen peptides by CPH, as most reported assays relied on indirect quantification of CO₂ release as a result of 2OG turnover (19,20,22-24). The structural and biochemical work carried out on CPH, conducted in collaboration with John Scotti, has advanced the capability in studying CPH through the development of methods to fully purify the CPH $\alpha\beta$ complex, to crystallise the CPH complex, and to directly quantify its hydroxylase activity (Chapter 6). Although a structure of the CPH $\alpha_2\beta_2$ complex has not been solved yet, SAXS solution data indicated that the purified protein is likely to be

the $\alpha_2\beta$ heterotrimer. This finding is unprecedented as CPH is so far reported to exist only as a tetramer or a dimer in solution (19,20,23-26). While this result does not invalidate the existence of a CPH tetramer, it does show that CPH can exist as a trimer. The biochemical assay developed in this research can now be used for *in vitro* screening of inhibitors for CPH that could potentially be developed into anti-metastatic agents.

Overall, the described structural and biochemical research represent important breakthroughs that will aid medicinal chemistry and *in vitro* studies of two groups of 2OG oxygenases – nucleic acid and collagen modifying 2OG oxygenases. Members of these two groups of 2OG oxygenases are united by their roles in human disease (27-30), as well as their being part of a larger family of 2OG oxygenases that have important physiological functions (31). It is hoped that the structural and biochemical data arising from this research will directly or indirectly expand our overall understanding of the versatile superfamily of Fe(II) and 2-oxoglutarate dependent oxygenases.

7.1 References

1. Yang, C.G., Yi, C., Duguid, E.M., Sullivan, C.T., Jian, X., Rice, P.A. and He, C. (2008) Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature*, **452**, 961-965.
2. Yi, C., Jia, G., Hou, G., Dai, Q., Zhang, W., Zheng, G., Jian, X., Yang, C.G., Cui, Q. and He, C. (2010) Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature*, **468**, 330-333.
3. Yu, B., Edstrom, W.C., Benach, J., Hamuro, Y., Weber, P.C., Gibney, B.R. and Hunt, J.F. (2006) Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature*, **439**, 879-884.
4. Yu, B. and Hunt, J.F. (2009) Enzymological and structural studies of the mechanism of promiscuous substrate recognition by the oxidative DNA repair enzyme AlkB. *Proc Natl Acad Sci USA*, **106**, 14315-14320.
5. Han, Z., Niu, T., Chang, J., Lei, X., Zhao, M., Wang, Q., Cheng, W., Wang, J., Feng, Y. and Chai, J. (2010) Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature*, **464**, 1205-1209.
6. Aik, W., Demetriades, M., Hamdan, M.K., Bagg, E.A., Yeoh, K.K., Lejeune, C., Zhang, Z., McDonough, M.A. and Schofield, C.J. (2013) Structural basis for inhibition of the fat mass and obesity associated protein (FTO). *J Med Chem*, **56**, 3680-3688.
7. Aik, W., Scotti, J.S., Choi, H., Gong, L., Demetriades, M., Schofield, C.J. and McDonough, M.A. (2014) Structure of human RNA N(6)-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res*, **42**, 4741-4754.
8. Hopkinson, R.J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K.H., Li, X.S., Kristensen, J.B.L., King, O.N.F., Chan, M.C. *et al.* (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem Sci*, **4**, 3110-3117.
9. Woon, E.C., Demetriades, M., Bagg, E.A., Aik, W., Krylova, S.M., Ma, J.H., Chan, M., Walport, L.J., Wegman, D.W., Dack, K.N. *et al.* (2012) Dynamic combinatorial mass spectrometry leads to inhibitors of a 2-oxoglutarate-dependent nucleic acid demethylase. *J Med Chem*, **55**, 2173-2184.
10. Chen, B.E., Ye, F., Yu, L., Jia, G.F., Huang, X.T., Zhang, X.J., Peng, S.Y., Chen, K., Wang, M.N., Gong, S.Z. *et al.* (2012) Development of Cell-Active N-6-Methyladenosine RNA Demethylase FTO Inhibitor. *Journal of the American Chemical Society*, **134**, 17963-17971.
11. Aik, W., Scotti, J.S., Choi, H., Gong, L., Demetriades, M., Schofield, C.J. and McDonough, M.A. (2014) Structure of human RNA N6-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res*.

12. Zheng, G., Dahl, J.A., Niu, Y., Fedorcsak, P., Huang, C.M., Li, C.J., Vagbo, C.B., Shi, Y., Wang, W.L., Song, S.H. *et al.* (2013) ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell*, **49**, 18-29.
13. Chen, W., Zhang, L., Zheng, G., Fu, Y., Ji, Q., Liu, F., Chen, H. and He, C. (2014) Crystal structure of the RNA demethylase ALKBH5 from zebrafish. *FEBS Lett*, **588**, 892-898.
14. Feng, C., Liu, Y., Wang, G., Deng, Z., Zhang, Q., Wu, W., Tong, Y., Cheng, C. and Chen, Z. (2014) Crystal structures of the human RNA demethylase alkbh5 reveal basis for substrate recognition. *J Biol Chem*, **289**, 11571-11583.
15. Xu, C., Liu, K., Tempel, W., Demetriades, M., Aik, W., Schofield, C.J. and Min, J. (2014) Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single stranded m6A RNA demethylation. *J Biol Chem*.
16. Aravind, L. and Koonin, E.V. (2001) The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol*, **2**, RESEARCH0007.
17. Hutton, J.J., Jr., Trappel, A.L. and Udenfriend, S. (1966) Requirements for alpha-ketoglutarate, ferrous ion and ascorbate by collagen proline hydroxylase. *Biochem Biophys Res Commun*, **24**, 179-184.
18. Neubauer, A., Neubauer, P. and Myllyharju, J. (2005) High-level production of human collagen prolyl 4-hydroxylase in *Escherichia coli*. *Matrix Biol*, **24**, 59-68.
19. Kivirikko, K.I. and Myllyharju, J. (1998) Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biology*, **16**, 357-368.
20. Kivirikko, K.I. and Pihlajaniemi, T. (1998) Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Adv Enzymol Relat Areas Mol Biol*, **72**, 325-398.
21. Pekkala, M., Hieta, R., Kursula, P., Kivirikko, K.I., Wierenga, R.K. and Myllyharju, J. (2003) Crystallization of the proline-rich-peptide binding domain of human type I collagen prolyl 4-hydroxylase. *Acta Crystallogr D Biol Crystallogr*, **59**, 940-942.
22. Kivirikko, K.I. and Myllyla, R. (1982) Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods Enzymol*, **82 Pt A**, 245-304.
23. Kukkola, L., Koivunen, P., Pakkanen, O., Page, A.P. and Myllyharju, J. (2004) Collagen prolyl 4-hydroxylase tetramers and dimers show identical decreases in Km values for peptide substrates with increasing chain length: mutation of one of the two catalytic sites in the tetramer inactivates the enzyme by more than half. *J Biol Chem*, **279**, 18656-18661.
24. Myllyharju, J. (2003) Prolyl 4-hydroxylases, the key enzymes of collagen biosynthesis. *Matrix Biol*, **22**, 15-24.
25. Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T. and Kivirikko, K.I. (1997) Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer. The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta2 tetramer. *J Biol Chem*, **272**, 17342-17348.

26. Helaakoski, T., Annunen, P., Vuori, K., MacNeil, I.A., Pihlajaniemi, T. and Kivirikko, K.I. (1995) Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase alpha-subunit isoform: formation of an alpha 2 beta 2 tetramer with the protein disulfide-isomerase/beta subunit. *Proc Natl Acad Sci U S A*, **92**, 4427-4431.
27. Thalhammer, A., Aik, W., Bagg, E.A.L. and Schofield, C.J. (2012) The potential of 2-oxoglutarate oxygenases acting on nucleic acids as therapeutic targets. *Drug Discovery Today: Therapeutic Strategies*, **9**, e91-e100.
28. Gilkes, D.M., Bajpai, S., Wong, C.C., Chaturvedi, P., Hubbi, M.E., Wirtz, D. and Semenza, G.L. (2013) Procollagen lysyl hydroxylase 2 is essential for hypoxia-induced breast cancer metastasis. *Mol Cancer Res*, **11**, 456-466.
29. Gilkes, D.M., Chaturvedi, P., Bajpai, S., Wong, C.C., Wei, H., Pitcairn, S., Hubbi, M.E., Wirtz, D. and Semenza, G.L. (2013) Collagen prolyl hydroxylases are essential for breast cancer metastasis. *Cancer Res*, **73**, 3285-3296.
30. Rose, N.R., McDonough, M.A., King, O.N.F., Kawamura, A. and Schofield, C.J. (2011) Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev*, **40**, 4364-4397.
31. Loenarz, C. and Schofield, C.J. (2011) Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci*, **36**, 7-18.

Appendix

Table A - 1. RNA co-purified with hFTO from over-production in *Escherichia coli* BL21 (DE3).

No.	RNA	Nucleotide residue number	Sequence (5' – 3')
1	23S ribosomal RNA	3287299 – 3287642 (of <i>Escherichia coli</i> BL21 (DE3) complete genome)	CTTCTCGGTTTCCCTTCGGCTCCCCTATTCGGTTAAC CTTGCTACAGAATATAAGTCGCTGACCCATTATACA AAAGGTACGCAGTCACACAGATTGCTCTGTGCTCCCA CTGCTTGTACGTACACGGTTTCAGGTTCTTTTCACT CCCCTCGCCGGGTTCTTTTCGCCTTTCCTCACGGT ACTGGTTCACATATCGGTCAGTCAGGAGTATTTAGCCT TGGAGGATGGTCCCCCATATTTCAGACAGGATACGAC GTGTCCCGCCCTACTCATCGAGCTCACAGCATGTGCA TTTTTGTGTACGGGGCTGTCACCCTGTATCGCGCGCC TTTCCAGACGCTTCC
2	23S ribosomal RNA	2593963 – 2594391 (of <i>Escherichia coli</i> BL21 (DE3) complete genome)	GTCCACTTTCGTGTTTGCACAGTGCTGTGTTTTTAAT AAACAGTTGCAGCCAGCTGGTATCTTCGACTGATTTTC AGCTCCATCCGCGAGGGACTTCACCTACATATCAGCG TGCCTTCTCCCGAAGTTACGGCACCATTTTGCCTAGT TCCTTCACCCGAGTTCTCTCAAGCGCCTTGGTATTCT CTACCTGACCACCTGTGTGCGTTTTGGGGTACGATTTG ATGTTACCTGATGCTTAGAGGCTTTTCCTGGAAGCA GGGCATTTGTTGCTTCAGCACCGTAGTGCCTCGTCAT CAGCCTCAGCCTTGATTTTCCGGATTTGCCTGGAAA ACCAGCCTACACGCTTAAACCGGGACAACCGTCGCCC GGCCAACATAGCCTTCTCCGTCCCCCTTCGCAGTAA CACCAAGTACAGGAATATTAACC
3	23S ribosomal RNA	2593730 – 2594114 (of <i>Escherichia coli</i> BL21 (DE3) complete genome)	TTCGGCGGGTACACTGCATCTTCACAGCGAGTTCAAT TTCAGTGTCTCGGGTGGAGACAGCCTGGCCATCAT TACGCCATTTCGTGCAGGTCGGAACCTTACCCGACAAGG AATTTTCGTACCTTAGGACCGTTATAGTTACGGCCGC CGTTTACCGGGGCTTCGATCAAGAGCTTCGCGTTACC GCTAACCCCATCAATTAACCTTCCGGCACCGGGCAGG CGTCACACCGTATACGTCCACTTTTCGTGTTTGCACAG TGCTGTGTTTTTAATAAACAGTTGCAGCCAGCTGGT ATCTTCGACTGATTTTCAGCTCCATCCGCGAGGGACTT CACCTACATATCAGCGTGCCTTCTCCCGAAGTTACGG CACCATTTTGCCTAGTTCCT

Table A - 2. RNA pulled down by recombinant hFTO from HEK293T cell lysate.

No.	RNA	Nucleotide residue number	Sequence (5' - 3')
1	RN7SK	178 – 226	CTTCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTGC TAGAACCTCCTACAAATGGACCTTGAGAGCTTGTT
2	RN7SK	17 – 161	TGGCTGCGACATCTGTCACCCCATTTGATCGCCAGGGTT GATTTCGGCTGATCTGGCTGGCTAGGCGGGTGTCCCCTT CCTCCCTCACCGCTCCATGTGCGTCCCTCCCGAAGCTGC GCGCTCGGTCAAGAGGACGACCATCCCCG
3	45S pre- ribosomal RNA	4918 – 5293	CCTCGTTCATGGGGAATAATTGCAATCCCCGATCCCCA TCACGAATGGGGTTCAACGGGTTACCCGCGCTGCCGG CGTAGGGTAGGCACACGCTGAGCCAGTCAGTGTAGCGC GCGTGACGCCCCGACATCTAAGGGCATCACAGACCTG TTATTGCTCAATCTCGGGTGGCTGAACGCCACTTGTCC CTCTAAGAAGTTGGGGGACGCCGACCGCTCGGGGGTGC CGTAACTAGTTAGCATGCCAGAGTCTCGTTCGTTATCG GAATTAACCAGACAAATCGCTCCACCACTAAGAACG GCCATGCACCACCACCCACGGAATCGAGAAAGAGCTAT CAATCTGTCAATCCTGTCCGTGTCCGGGCCGGGTGTGT
4	45S pre- ribosomal RNA	9281 – 9662	TTTTTACCGGATAAACTGCGTGCGGGGGGTGCGTCGG GTCTGCGAGAGCGCCAGCTATCCTGAGGGAACTTCGG AGGGAACCAGCTACTAGATGGTTCGATTAGTCTTTTCGC CCCTATACCCAGGTCGGACGACCGATTTGCACGTCAGG ACCGCTACGGACCTCCACCAGAGTTTCTCTGGCTTCG CCCTGCCCAGGCATAGTTCACCATCTTTCGGGTCCTAA CACGTGCGCTCGTGCTCCACCTCCCCGGCGCGGGCGG GAGACGGGCCGGTGGTGCGCCCTCGGCGGACTGGAGAG GCCTCGGGATCCACCTCGGCCGGCGAGCGCGCCGGCC TTCACCTTCATTGCGCCACGGCGGCTTTCGTGCGAGCC CCCGTC
5	RN7SK	143 – 291	AAGAGGACGACCATCCCCGATAGAGGAGGACCGGTCT TCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTGCTA GAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAG AACGTAGGGTAGTCAAGCTTCCAAGACTCCAGATCAC
6	RNU4-2	14 – 75	GGCAGTATCGTAGCCAATGAGGTTTATCCGAGGCGCG ATTATTGCTAATTGAAAACCTTTTCCTA
7	45S pre- ribosomal RNA	11681 – 11386	CTCCTGCCGTTTACCCGCGCTTCATTGAATTTCTTCAC TTTGACATTTCAGAGCACTGGGCAGAAATCACATCGCGT CAACACCCGCCGCGGGCCTTCGCGATGCTTTGTTTTAA TTAAACAGTCGGATTCCCCTGGTCCGCACCAGTTCTAA GTCGGCTGCTAGGCGCCGGCCGAGGCGAGGCGCCGCGC GGAACCGCGGCCCCGGGGGCGGACCCGGCGGGGGGAC CGGCCCCGCGGCCCTCCGCCGCCGCGCGCCGCGCGCG CGCCGAGGAGGAGGGGGGAACGGGGGN
8	45S pre- ribosomal	10205- 11067	CCGCCCCGACCCTTCTCCCCNCGCCGCCGCCCCNCGC GGCGCTCCCCCGGGAGGGGGGAGGACGGGGAGCGGG

	RNA		GGAGAGAGAGAGAGAGAGAGGGCGCGGGGTGGGGAGGGA GCGAGCGGCGCGCGCGGGGTGGGGCGGGGAGGGCCGC GAGGGGGGTGCCCCGGGCGTGGGGGGGCGGCGGCGCC TCGTCCAGCCGCGGCGCGCGCCAGCCCCGCTTCGCGCC CCAGCCCGACCGACCCAGCCCTTAGAGCCAATCCTTAT CCCGAAGTTACGGATCCGGCTTGCCGACTTCCCTTACC TACATTGTTCCAACATGCCAGAGGCTGTTACCTTGGA GACCTGCTGCGGATATGGGTACGGCCCGGCGCGAGATT TACACCCTCTCCCCGGATTTTCAAGGGCCAGCGAGAG CTCACCGGACGCCGCGGAACCGCGACGCTTTCCAAGG CACGGGCCCCTCTCTCGGGGCGAACCATTCCAGGGCG CCCTGCCCTTCACAAAGAAAAGAGAACTCTCCCCGGG CTCCCGCCGGCTTCTCCGGGATCGGTGCGGTTACCGCA CTGGACGCCTCGCGGCGCCATCTCCGCCACTCCGGAT TCGGGGATCTGAACCCGACTCCCTTTTCGATCGGCCGAG GGCAACGGAGGCCATCGCCCGTCCCTTCGGAACGGCGC TCGCCCATCTCTCAGGACCGACTGACCCATGTTCAACT GCTGTTACATGGAACCCTTCTCCACTTCGGCCTTCAA AGTTCTCGTTTGAATATTTGCTACTACCACCAAGATCT GCACCTGCGGCGGCTCCACCCGGGCCCCGCGCCCTAGGC TTCAAGGCTCACCGCAGCGGCCCTCCTAA
9	RNU4-2	25 – 77	AGCCAATGAGGTTTATCCGAGGCGCGATTATTGCTAA TTGAAAACTTTCCCA
10	45S pre- ribosomal RNA	4951 – 5031	ATCGGAATTAACCAGACAAATCGCTCCACCAACTAAG AACGGCCATGCACCACCACCCACGGAATCGAGAAAGAG CTATCA
11	16S mitochond ria rRNA	2238 – 2376 (of human mitochondri on complete genome)	CTAACATATAACTGAACTCCTCACACCCAATTGGACCA ATCTATCACCTATAGAAGAACTAATGTTAGTATAAG TAACATGAAAACATTCTCCTCCGCATAAGCCTGCGTCA GATTAAAACACTGAACTGACAATTAACA
12	RNU4-1	25 – 75	AGCCAATGAGGTCTATCCGAGGCGCGATTATTGCTAAT TGAAAACTTTCCCA
13	45S pre- ribosomal RNA	9707 – 9782	CTCAAACTTTAAATGGGTAAGAAGCCCGGCTCGCTGGC GTGGAGCCGGGCGTGGAATGCGAGTGCCTAGTGGGCC AGTTCTGCTTACCAAAAGTGGCC
14	RN7SK	143 – 291	AAGAGGACGACCATCCCCGATAGAGGAGGACCGGTCT TCGGTCAAGGGTATACGAGTAGCTGCGCTCCCCTGCTA GAACCTCCAAACAAGCTCTCAAGGTCCATTTGTAGGAG AACGTAGGGTAGTCAAGCTTCCAAGACTCCAGATCAC
15	18S rRNA	934 – 1091	GGGCATTTCGTATTGCGCCGCTAGAGGTGAAATTCCTGG ACCGGCGCAAGACGGACCAGAGCGAAAGCATTTGCCA AGAATGTTTTTCATTAATCAAGAACGAAAGTCGGAGGT TCGAAGACGATCAGATACCGTCGTAGTTCCGACCATAA ACGACGCCA
16	ATXN2L, transcript variant F,	953 – 1129	CTTGATTCAATCTCTCGAGCCAACTGGGCCGCACGCAG CTCTCGCTGACGAACTCTTCTGAGTTGTCCTTTTCTA AGGGCACCGTATAAGAAGAAAGACTGCTATCATAGGT

	mRNA		AGTCTTCACACCGTAGTTCTCCTCATTGAACTTGAACA TTTCATTGGGGTCCCATCCATTGGACTGGGGAGGGAGG AAGGTGGTTTTTAAAACCCAGCAGGAAAAGCAACGACA AGTCACA
17	RN7SK	38 – 287	CATTGATCGCCAGGGTTGATTTCGGCTGATCTGGCTGGC TAGGCGGGTGTCCCTTCCTCCCTCACCGCTCCATGTG CGTCCCTCCCGAAGCTGCGCGCTCGGTCTGAAGAGGACG ACCATCCCCGATAGAGGAGGACCGGTCTTCGGTCAAGG GTATACGAGTAGCTGCGCTCCCTGCTAGAACCTCCAA ACAAGCTCTCAAGGTCCATTTGTAGGAGAACGTAGGG TAGTCAAGCTTCCAAGACTCCAG
18	45S pre- ribosomal RNA	10543 – 10736	GGAGAGTTCTCTTTTCTTTGTGAAGGGCAGGGCGCCCT GGAATGGGTTTCGCCCCGAGAGAGGGGCGCGTGCCTTGG AAAGCGTCGCGGTTCCGGCGGCGTCCGGTGAGCTCTCG CTGGCCCTTGAAAATCCGGGGGAGAGGGTGTAATCT CGCGCCGGGCGGTACCCATATCCGCAGCAGGTCTCCAA GGTGA
19	45S pre- ribosomal RNA	5108 – 5287	AAGTGGCGTTCAGCCACCCGAGATTGAGCAATAACAG GTCTGTGATGCCCTTAGATGTCCGGGGCTGCACGCGCG CTACACTGACTGGCTCAGCGTGTGCCTACCCTACGCCG GCAGGCGCGGGTAACCCGTTGAACCCCATTCGTGATGG GGATCGGGGATTGCAATTATTCCTCCATGATA
20	RN7SK	10 – 287	CTGGAGTCTTGGAAGCTTGACTACCCTACGTTCTCCTA CAAATGGACCTTGAGAGCTTGTTTGGAGGTTCTAGCA GGGGAGCGCAGCTACTCGTATACCCTTGACCGAAGACC GGTCTCCTCTATCGGGGATGGTCGTCTCTTCGACCG AGCGCGCAGCTTCGGGAGGGACGCACATGGAGCGGTG AGGGAGGAAGGGGACACCCGCCTAGCCAGCCAGATCA GCCGAATCAACCCTGGCGATCAATGGGGTGACAGATG TCGCAGCCAGATCGCC

Table A – 3. Mascot searches of proteins pulled-down by N-terminal GFP-tagged OGFOD2 in 293T cells.

Access.	Score	emPAI	Description
P68371	4267	16.41	Tubulin beta-2C chain
Q13509	2687	5.18	Tubulin beta-3 chain
P68366	2364	3.01	Tubulin alpha-4A chain
Q9H3K6	156	1.38	BolA-like protein 2
P62829	200	1.27	60S ribosomal protein L23
Q9Y5L4	146	0.74	Mitochondrial import inner membrane translocase subunit Tim13
Q5VYJ4	56	0.74	Putative small nuclear ribonucleoprotein polypeptide E-like protein 1
P15531	91	0.71	Nucleoside diphosphate kinase A
P31689	510	0.63	DnaJ homolog subfamily A member 1
P51970	411	0.58	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8
P62913	137	0.58	60S ribosomal protein L11
Q06830	94	0.52	Peroxiredoxin-1
P27708	1213	0.51	CAD protein
P09455	20	0.47	Retinol-binding protein 1
Q5VTE0	310	0.46	Putative elongation factor 1-alpha-like 3
P62249	106	0.45	40S ribosomal protein S 16
P62857	50	0.45	40S ribosomal protein S 28
P07478	107	0.42	Trypsin-2
P63173	39	0.42	60S ribosomal protein L38
O60884	181	0.32	DnaJ homolog subfamily A member 2
Q92598	271	0.3	Heat shock protein 105 kDa
P05386	102	0.3	60S acidic ribosomal protein P1
P28072	154	0.28	Proteasome subunit beta type-6
P28066	61	0.27	Proteasome subunit alpha type-5
P62854	81	0.26	40S ribosomal protein S 26
Q15370	47	0.26	Transcription elongation factor B polypeptide 2
P60900	138	0.25	Proteasome subunit alpha type-6
P37108	77	0.23	Signal recognition particle 14 kDa protein
P35080	110	0.22	Profilin-2
P69905	71	0.22	Hemoglobin subunit alpha
P02100	53	0.21	Hemoglobin subunit epsilon
Q99460	312	0.2	26S proteasome non-ATPase regulatory subunit 1
P62277	68	0.2	40S ribosomal protein S 13
Q9GZV4	37	0.2	Eukaryotic translation initiation factor 5A-2
P24666	71	0.18	Low molecular weight phosphotyrosine protein phosphatase
P23528	67	0.18	Cofilin-1
P46783	29	0.18	40S ribosomal protein S 10
P55060	121	0.16	Exportin-2
P60842	117	0.15	Eukaryotic initiation factor 4A-I
P62820	90	0.15	Ras-related protein Rab-1A
Q9NR31	83	0.15	GTP-binding protein SAR 1a
P49721	67	0.15	Proteasome subunit beta type-2
P43686	106	0.14	26S protease regulatory subunit 6B
P17980	75	0.14	26S protease regulatory subunit 6A
P24534	119	0.13	Elongation factor 1-beta
P25787	55	0.13	Proteasome subunit alpha type-2
P31947	72	0.12	14-3-3 protein sigma
P60174	57	0.12	Triosephosphate isomerase
P25788	53	0.12	Proteasome subunit alpha type-3
P28074	52	0.12	Proteasome subunit beta type-5
Q96T66	29	0.12	Nicotinamide mononucleotide adenylyltransferase 3
O95757	77	0.11	Heat shock 70 kDa protein 4L
Q86Y56	102	0.11	HEAT repeat-containing protein 2
P25789	83	0.11	Proteasome subunit alpha type-4
Q8N0Y7	69	0.11	Probable phosphoglycerate mutase 4
Q86UV7	45	0.11	Tripartite motif-containing protein 73
P25786	27	0.11	Proteasome subunit alpha type-1
Q99614	60	0.1	Tetratricopeptide repeat protein 1
P22087	44	0.1	rRNA 2'-O-methyltransferase fibrillarin

Table A – 4. Mascot searches of proteins pulled-down by C-terminal GFP-tagged OGFO2 in 293T cells.

Accession	Score	Empirical PAI	Description
P04350	4201	13.56	Tubulin beta-4 chain
Q13885	4102	12.51	Tubulin beta-2A chain
Q13509	2852	6.01	Tubulin beta-3 chain
P68363	4096	4.12	Tubulin alpha-1B chain
Q9NY65	2008	1.92	Tubulin alpha-8 chain
P60842	866	1.8	Eukaryotic initiation factor 4A-I
P31689	552	1.48	DnaJ homolog subfamily A member 1
Q06830	195	1.32	Peroxisomal protein 1
P62829	105	1.27	60S ribosomal protein L23
P62820	152	0.73	Ras-related protein Rab-1A
P24666	138	0.65	Low molecular weight phosphotyrosine protein phosphatase
P51970	422	0.58	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 8
P62913	237	0.58	60S ribosomal protein L11
P27708	1590	0.57	CAD protein
P32119	190	0.53	Peroxisomal protein 2
P61026	100	0.51	Ras-related protein Rab-10
Q9UHV9	107	0.45	Prefoldin subunit 2
Q96KS0	175	0.43	Egln9 homolog 2
Q9HB71	154	0.43	Calcyclin-binding protein
Q8N726	56	0.4	Cyclin-dependent kinase inhibitor 2A, isoform 4
Q5VTE0	330	0.37	Putative elongation factor 1-alpha-like 3
Q71UM5	67	0.36	40S ribosomal protein S27-like
Q9H3K6	80	0.34	BolA-like protein 2
P62304	48	0.32	Small nuclear ribonucleoprotein E
P49721	68	0.31	Proteasome subunit beta type-2
Q92598	342	0.3	Heat shock protein 105 kDa
P05386	47	0.3	60S acidic ribosomal protein P1
P00492	112	0.29	Hypoxanthine-guanine phosphoribosyltransferase
Q9Y230	105	0.28	RuvB-like 2
P28072	99	0.28	Proteasome subunit beta type-6
P25787	57	0.27	Proteasome subunit alpha type-2
P07478	77	0.26	Trypsin-2
Q14818	124	0.25	Proteasome subunit alpha type-7
Q60884	139	0.23	DnaJ homolog subfamily A member 2
Q9NPJ4	46	0.22	Proline-rich nuclear receptor coactivator 2
P69891	67	0.21	Hemoglobin subunit gamma-1
P09455	16	0.21	Retinol-binding protein 1
Q00410	265	0.2	Importin-5
P46776	54	0.2	60S ribosomal protein L27a
P62249	39	0.2	40S ribosomal protein S16
P22392	73	0.19	Nucleoside diphosphate kinase B
P62937	65	0.19	Peptidyl-prolyl cis-trans isomerase A
P30050	28	0.19	60S ribosomal protein L12
P11441	22	0.19	Ubiquitin-like protein 4A
P23528	86	0.18	Cofilin-1
P46783	35	0.18	40S ribosomal protein S10
P63208	28	0.18	S-phase kinase-associated protein 1
Q43324	40	0.17	Eukaryotic translation elongation factor 1 epsilon-1
Q8N5M4	39	0.17	Tetratricopeptide repeat protein 9C
A6NIZ1	43	0.16	Ras-related protein Rap-1b-like protein
P52597	196	0.15	Heterogeneous nuclear ribonucleoprotein F
Q13263	161	0.15	Transcription intermediary factor 1-beta
Q9Y6B6	82	0.15	GTP-binding protein SAR1b
P61586	31	0.15	Transforming protein RhoA
P38646	139	0.14	Stress-70 protein, mitochondrial
Q95816	115	0.14	BAG family molecular chaperone regulator 2
P62826	95	0.14	GTP-binding nuclear protein Ran
P51149	76	0.14	Ras-related protein Rab-7a
Q15907	72	0.14	Ras-related protein Rab-11B
Q96L21	52	0.13	60S ribosomal protein L10-like
P24534	47	0.13	Elongation factor 1-beta
Q8IUX1	29	0.13	Transmembrane protein 126B
P55060	155	0.12	Exportin-2
P60174	82	0.12	Triosephosphate isomerase
P09661	60	0.12	U2 small nuclear ribonucleoprotein A'
P31947	34	0.12	14-3-3 protein sigma
Q86Y56	148	0.11	HEAT repeat-containing protein 2
P25789	62	0.11	Proteasome subunit alpha type-4
P15259	50	0.11	Phosphoglycerate mutase 2
Q9Y5Y2	32	0.11	Cytosolic Fe-S cluster assembly factor NUBP2
Q43707	94	0.1	Alpha-actinin-4
Q99614	79	0.1	Tetratricopeptide repeat protein 1
P52907	60	0.1	F-actin-capping protein subunit alpha-1
Q13123	45	0.1	Protein Red
P22087	42	0.1	rRNA 2'-O-methyltransferase fibrillarin

The following pages contain co-authored full articles and the associated licences/permissions for reprinting:

1. Full article reprinted with permission from Woon, E.C.Y., Demetriades, M., Bagg, E.A.L., Aik, W., Krylova, S.M., Ma, J.H.Y., Chan, M., Walport, L.J., Wegman, D.W., Dack, K.N., McDonough, M.A., Krylov, S.M. and Schofield, C.J. (2012) 'Dynamic Combinatorial Mass Spectrometry Leads to Inhibitors of a 2-Oxoglutarate-Dependent Nucleic Acid Demethylase. *J. Med. Chem.*, **55**, 2173-2184.' Copyright (2012) American Chemical Society.
2. Full article reprinted from Drug Discovery Today: Therapeutic Strategies, **9**, Thalhammer, A., Aik, W., Bagg, E.A.L. and Schofield, C.J., The Potential of 2-Oxoglutarate Oxygenases Acting on Nucleic Acids as Therapeutic Targets., e91-e100, Copyright (2012), with permission from Elsevier.
3. Full article reprinted from Current Opinion in Structural Biology, **22**, Aik, W., McDonough, M.A., Thalhammer, A., Chowdhury, R. and Schofield, C.J., Role of the Jelly-roll Fold in Substrate Binding by 2-Oxoglutarate Oxygenases., 691-700, Copyright (2012), with permission from Elsevier.
4. Full article reprinted with permission from Aik, W., Demetriades, M., Hamdan, M.K.K., Bagg, E.A.L., Yeoh, K.K., Lejeune, C., Zhang, Z., McDonough, M.A. and Schofield, C.J. (2013) Structural Basis for Inhibition of the Fat Mass and Obesity Associated Protein (FTO). *J Med Chem*, **56**, 3680-3688. (13) Copyright (2013) American Chemical Society.
5. Full article reproduced from Hopkinson, R.J., Tumber, A., Yapp, C., Chowdhury, R., Aik, W., Che, K.H., Li, X.S., Kristensen, J.B.L., King, O.N.F., Chan, M.C. *et al.* (2013) 5-Carboxy-8-hydroxyquinoline is a broad spectrum 2-oxoglutarate oxygenase inhibitor which causes iron translocation. *Chem. Sci.*, **4**, 3110 - reproduced by permission of The Royal Society of Chemistry.
6. Aik, W., Scotti, J.S., Choi, H., Gong, L., Demetriades, M., Schofield, C.J. and McDonough, M.A. Structure of human RNA *N*⁶-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res.* 2014, 42 (7): 4741-4754 by permission of Oxford University Press
7. Xu, C., Liu, K., Tempel, W., Demetriades, M., Aik, W., Schofield, C.J. and Min, J. (2014) Structures of human ALKBH5 demethylase reveal a unique binding mode for specific single stranded m⁶A RNA demethylation. *J Biol Chem.* (In press) by permission of the American Society for Biochemistry and Molecular Biology.