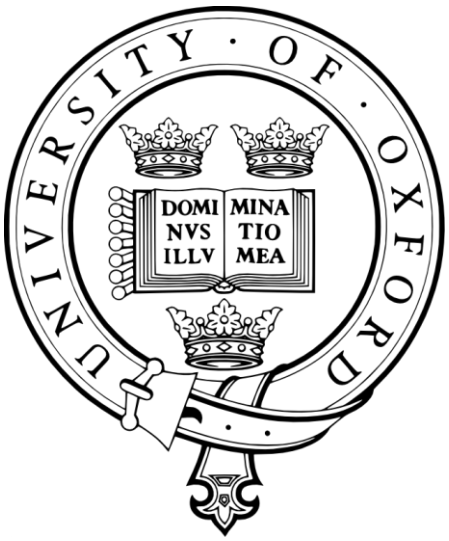
Roles of Oxygenases in Nucleic Acid Modification

Eleanor A. L. Bagg



Chemistry Research Laboratory

Department of Biochemistry

Lincoln College

Trinity Term 2011

A thesis submitted to the board of the Division of Medical Sciences of the University of Oxford in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Roles of Oxygenases in Nucleic Acid Modification

2-Oxoglutarate (2OG) and Fe(II) dependent oxygenases have a broad range of substrates, extending from histones to fatty acids. Several 2OG oxygenases have nucleic acid substrates, with members of the AlkB subfamily being responsible for nucleic acid modification and repair. The AlkB protein itself is part of the *Escherichia coli* adaptive response, protecting the DNA from methylation damage. Methyl lesions are repaired by a direct removal mechanism via a hydroxylated intermediate, with release of formaldehyde. Homologues of AlkB have been identified throughout the vertebrates, with nine known human homologues: AlkB homologue 1-8 (ABH1-8) and Fat, mass and obesity associated protein (FTO). ABH2, ABH3 and FTO catalyse similar reactions to AlkB, whereas ABH8 methylates then hydroxylates modified wobble-position uridines in tRNA. The remaining homologues are of unknown function.

The *FTO* gene is associated with obesity in humans, a link confirmed by mouse models; mice lacking FTO are thinner than wildtype individuals, whereas overexpression of FTO leads to increased mass. Investigation of recombinant FTO identified a novel *C*-terminal helical domain which appears to mediate protein dimerisation *in vitro*. A loss of function mutation in this *C*-terminal domain produces a lean phenotype in mice, emphasising the importance of this domain for the protein's function *in vivo*. The FTO protein was further studied in cells, and localisation of several protein variant constructs were studied by immunofluorescence. Cell lysis and immunoprecipitation techniques were developed that enable proteomic analyses of proteins with which FTO may interact in cells. No protein interactors were confidently identified, suggesting that FTO may not interact with specific proteins in cells, and instead may preferentially interact with nucleic acids.

Studies were initiated on two further members of the ABH family, ABH1 and ABH7. Recombinant proteins were prepared and characterised as 2OG oxygenases, however initial attempts to identify potential histone or nucleic acid substrates were not successful. Both proteins were found to be localised in the mitochondria, however proteomic analysis was unable to identify proteins interacting with either protein in cells.

Selective inhibitors are required for *in vivo* inhibition of the ABH proteins. AlkB and ABH2 proteins were purified and characterised, and a formaldehyde dehydrogenase-coupled assay was developed to follow activity of these DNA demethylases. A dynamic combinatorial mass spectrometry method was employed to identify novel inhibitor scaffolds for AlkB, leading to the successful discovery of the first series of potent and selective inhibitors for this class of enzymes. Crystal structures of AlkB in complex with the most potent compounds were obtained, rationalising the inhibition observed. This work therefore suggests that therapeutic inhibition of this family of 2OG oxygenases is likely to be tractable.

Contents

Acknowledgements

Firstly, many thanks to Prof. Chris Schofield for supervision over the past four years in this exciting area of research, for his guidance, enthusiasm for scientific research, and countless suggestions and ideas.

I am very grateful to the two Germans that provided guidance at different stages of my DPhil. Firstly Thomas, who initiated studies on FTO and reignited the lab's interest in the DNA demethylases, and later Alex W, who introduced me to the art of cell culture, and his eternally optimistic outlook on lab work and research.

Within the CJS group many people have been willing to help and share their expertise. In particular I would like to thank Esther and Marina for AlkB DCMS; Wei Shen, Mike, and Jerome for crystallography; Akane and Nathan for advice on assays; Jasmin for mass spectrometry; Anders and Armin for oligo synthesis; and my Part II student Alex D for initiating studies on ABH1 and ABH7. Finally, thanks to Emily, Akane, Wei Shen and Adam for proof reading this thesis.

Many thanks to our collaborators on the FTO project within Oxford: Chris Church in Prof. Roger Cox's lab at Harwell, and James McTaggart, working with Prof. Fran Ashcroft in Physiology. Particular thanks also to Holger for valuable discussions, and his help in analysing (hundreds!) of proteomic tryptic digest samples.

Further thanks to our external collaborators: Sergey and Lana Krylov from Toronto for their capillary electrophoresis studies, and to the Pfizer team from Sandwich, coordinated by Kevin Dack, for interesting discussions and providing compounds.

Thank-you to everyone else in the CJS lab, in particular to Wendy for invaluable technical assistance, and Adam for patiently dealing with all of our computer queries. Additional thanks to Aman, Anna R, Celia, Christoph, Ed, Louise, Luc, Maggie, Monica, Olly, Inga, Rashed, Rok, Rubén, Richard, Sarah, Tom and all other CJS lab members from the past five years that have helped make working in the lab so enjoyable.

Outside of the lab a big thank-you to all my friends, in particular my fantastic housemates and other members of my favourite orchestra OMO, for providing a welcome break from the lab and science!

Last, but by no means least, a huge thank-you to my parents and sister for their support through (another) degree!

Abbreviations

1meA	1-methyladenine
1meG	1-methylguanine
2HG	2-hydroxyglutarate
2OG	2-oxoglutarate
2TY	2 x Tryptone/Yeast extract medium
3meC	3-methylcytosine
3meT	3-methylthymine
3meU	3-methyluracil
5hmeC	5-hydroxymethylcytosine
5meC	5-methylcytosine
ABH	AlkB human homologue
Amp^R	Ampicillin resistant
APS	Ammonium Persulfate
AU	Absorbance Units
BER	Base Excision Repair
BLAST	Basic Local Alignment Search Tool
BMI	Body Mass Index
bp	Base pairs (of DNA sequence)
BSA	Bovine Serum Albumin
Cam^R	Chloramphenicol resistant
CD	Circular Dichroism
ChIP	Chromatin Immunoprecipitation
CIP	Calf Intestinal Phosphatase
CMV	Cytomegalovirus
CoA	Coenzyme A
COS	CV-1 Origin, with SV40 (Green African monkey origin)
CV	Column volumes
Cy3	Cyanine dye 3
Da	Dalton (unit of molecular mass)
DAPI	4',6-diamidino-2-phenylindole
DCMS	Dynamic Combinatorial Mass Spectrometry
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DR	Direct Reversal
dNTP	Deoxyribonucleotide 5'-triphosphate
ds	Double stranded
DSBH	Double stranded β-helix
DSF	Differential Scanning Fluorimetry
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
EGFP	Enhanced Green Fluorescent Protein

EPR	Electron Paramagnetic Resonance
ESI	Electrospray Ionisation
FBS	Fetal Bovine Serum
FDH	Formaldehyde Dehydrogenase
FIH	Factor Inhibiting HIF
FPLC	Fast Protein Liquid Chromatography
FTO	Fat Mass and Obesity Associated; name of human FTO protein
GFP	Green fluorescent protein
HEK	Human Embyonic Kidney
HEPES	*N*-[2-Hydroxyethyl]piperazine-*N'*-[2-ethanesulphonic acid]
hFTO	*Homo sapiens* FTO
HIF	Hypoxia Inducible Factor
HPLC	High-pressure Liquid Chromatography
HREs	Hypoxia Response Elements
HRP	Horseradish Peroxidase
IC_{50}	Half-maximal inhibitory concentration
i.d.	Internal diameter
IMAC	Immobilised Metal Affinity Chromatography
I.M.A.G.E.	Integrated Molecular Analysis of Genomes and their Expression
IP	Immunoprecipitation
IPTG	Isopropyl-β-D-thiogalactopyranoside
Kan^R	Kanamycin resistant
LB	Luria-Bertani medium
LBA	Luria-Bertani agar
LCMS	Liquid Chromatography-Mass Spectrometry
MALDI	Matrix Assisted Laser Desorption/Ionisation
MES	2-(4-morpholino)-ethane sulfonic acid
mFTO	*Mus musculus* Fto
MMR	Mis-match repair
MMS	Methyl methanesulphonate
MS	Mass Spectrometry
MWCO	Molecular Weight Cut Off
n/a	Not applicable
NAD^+	Nicotinamide Adenine Dinucleotide
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NER	Nucleotide Excision Repair
NMR	Nuclear Magnetic Resonance
NOG	*N*-oxalylglycine
NP40	Nonidet P40 (octylphenoxypolyethoxyethanol)
NTA	Nitrilotriacetic acid
OD	Optical Density
PAGE	Polyacrylamide Gel Electrophoresis

PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	RCSB Protein Data Bank
PEG	Poly(ethylene glycol)
PEI	Poly(ethylenimine)
PHD	Prolyl Hydroxylase Domain Containing
pI	Isoelectric Point
ppm	Parts per million
PVDF	Polyvinylidene difluoride
pVHL	von Hippel-Lindau tumour suppressor protein
QToF	Quadrupole Time of Flight
RFU	Relative Fluorescence Units
RNA	Ribonucleic acid
ROS	Reactive oxygen species
rpm	Revolutions per minute
SAM	*S*-adenosyl-L-methionine
SDS	Sodium Dodecyl Sulphate
Sm^R	Streptomycin resistant
SOC	Super Optimal Broth Catabolite Repression
ss	Single stranded
T2D	Type II Diabetes
TAE	Tris-Acetate-EDTA
TBST	Tris Buffered Saline Tween
TED	*Tris*-carboxymethyl Ethylene Diamine
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TET	Ten-eleven translocation
TLS	Translesion Synthesis
T_m	Melting temperature
Tris	*Tris*-(hydroxymethyl)aminomethane
UV/vis	Ultraviolet/visible spectroscopy
v/v	Volume by volume
WT	Wildtype
w/v	Weight by volume
YFP	Yellow Fluorescent Protein

Chapter 1 – Introduction

1.1 Fe(II)/2OG dependent oxygenases

1.1.1 General features and mechanism of 2OG-dependent oxygenases

Members of the 2-oxoglutarate (2OG) and non-haem Fe(II) dependent oxygenase family catalyse a broad spectrum of synthetically challenging reactions, including hydroxylation, demethylation, desaturation, and oxidative ring closure[1]. In humans, 2OG oxygenases have roles in the regulation of gene expression at the DNA, RNA, transcription factor and protein levels[2]. In fact, 2OG oxygenases possibly catalyse the most diverse range of reactions of any enzyme family[3].

2OG oxygenase family members contain a typical double-stranded beta-helix (DSBH), or "jelly roll" fold (Figure 1-1), initially revealed by the crystal structure of isopenicillin-N synthase (IPNS)[4]. The DSBH core is surrounded by additional α-helices, β-strands, and loops, with the insertions and extensions differing for each protein. These help to define unique substrate binding sites and establish protein-protein contact surfaces[5]. The proteins have a common Fe(II) binding motif, in which the metal is coordinated by a 2-His-1-carboxylate facial triad, with the HXD/E residues normally found at the end of β2, and the second His on β7[6].

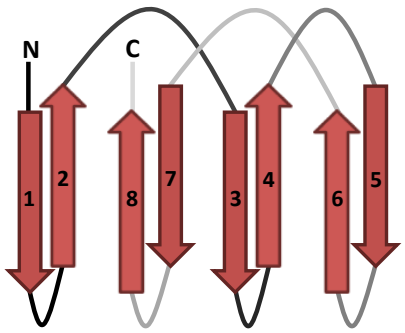


Figure 1-1: Schematic diagram of the typical double stranded β-helix (DSBH) fold. *N*- and *C*-termini are indicated, and the strands are labeled with their position in the sequence.

The required co-substrates for the reactions catalysed by the vast majority of the family members are the Krebs cycle intermediate 2OG, and molecular oxygen, with the byproducts of the reaction being succinate and carbon dioxide[2]. Typically these enzymes hydroxylate the primary substrate, during which one oxygen atom from the

dioxygen co-substrate is incorporated into the alcohol product[7], with the other oxygen atom incorporated into the succinate co-product[8]. Initially 2OG binds to the metal centre, followed by the substrate, then dioxygen. 2OG is stoichiometrically consumed in the formation of the Fe(IV) oxo species, which oxidises the substrate by insertion of an oxygen atom into a C-H bond. A general mechanism for the reactions of 2OG oxygenases is outlined in Figure 1-2, using the reaction of the DNA demethylase AlkB as an example.

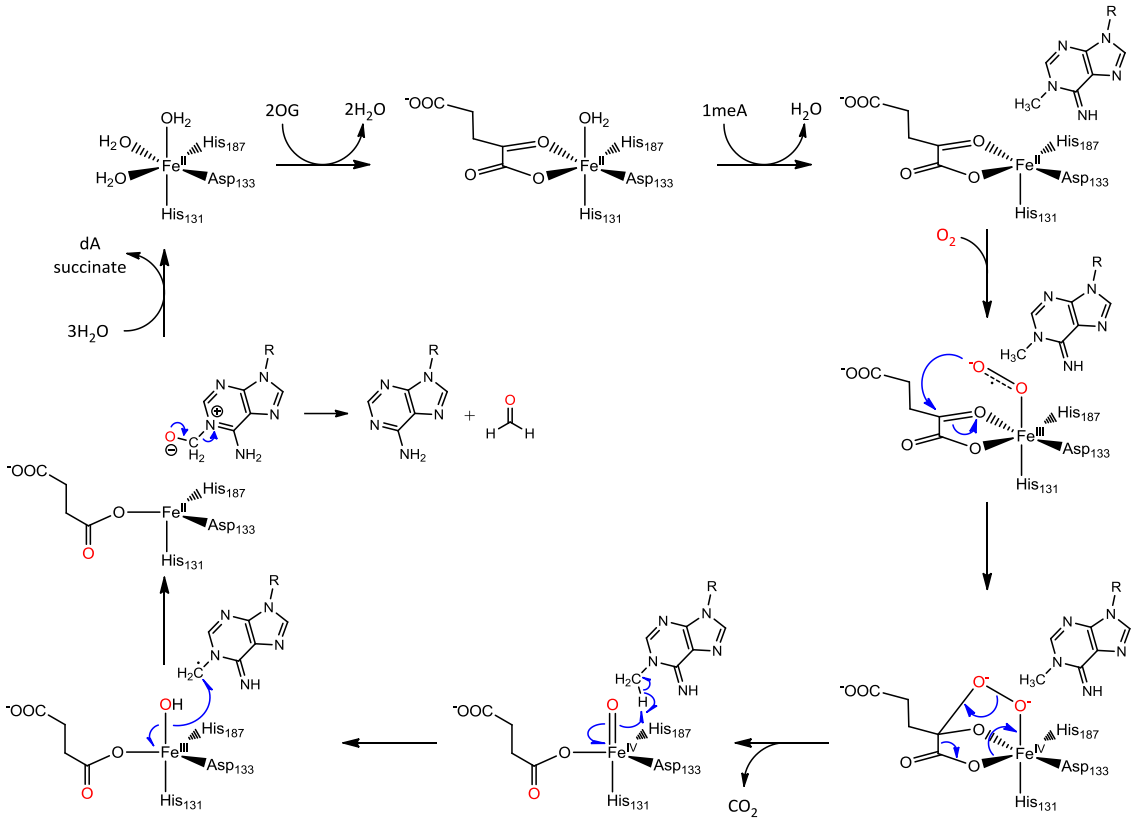


Figure 1-2: Consensus mechanism for the catalytic cycle of the 2OG oxygenase AlkB, showing demethylation of 1meA to yield A. 2OG binds to the metal centre, followed by binding of the 1meA substrate, then dioxygen. Reaction of dioxygen with the bound 2OG, then release of CO_2, produces a Fe(IV) oxo species. The oxo species can abstract a hydrogen from the 1meA methyl group, leading to a radical mechanism in which oxygen is inserted into a C-H bond of the 1meA species, producing the hydroxylated intermediate. This is concomitant with reduction of the Fe(IV) centre to Fe(II), and is followed by release of succinate from the Fe centre. The hydroxymethyl group collapses, recovering the demethylated base, and formaldehyde.

1.1.2 Post translational protein modifications

The first 2OG oxygenases identified were those involved in the biosynthesis of collagen, the major protein component of connective tissue. The three oxygenases

involved catalyse hydroxylation of Pro[9-10] and Lys[11] residues, thus stabilising the secondary structure of collagen fibrils and fibres[12].

1.1.3 Oxygen sensing

Hypoxia Inducible Factor (HIF) is a heterodimeric transcription factor, comprising α- and β-subunits. Upon dimerisation, the complex binds to hypoxia response elements (HREs) enabling transcription of many genes, including erythropoietin and vascular endothelial growth factor[13]. HIF levels are tightly regulated under normoxic conditions through post-translational hydroxylation by four oxygen-dependent oxygenases[14-16].

Factor Inhibiting HIF (FIH) is a HIF-α asparaginyl hydroxylase catalysing hydroxylation of N803 in the *C*-terminal transcription activation domain of human HIF-1α[17]. This hydroxylation blocks the interaction between HIF and its transcriptional coactivator p300, inhibiting the hypoxic response[18]. A family of three *trans*-4-prolyl hydroxylases (PHDs) hydroxylate HIF-1α at P402[19] or P564[20] in two oxygen-dependent degradation domains[21]. Hydroxylation at either site targets HIF-1α for ubiquitinylation through capture by pVHL and subsequent proteasomal degradation[22-23]. Under hypoxic conditions the PHDs are inhibited by limiting oxygen concentrations. Non-hydroxylated HIF-1α can therefore complex with HIF-1β and p300, enabling binding to DNA and activation of the hypoxic response cascade. This pathway is summarised in Figure 1-3.

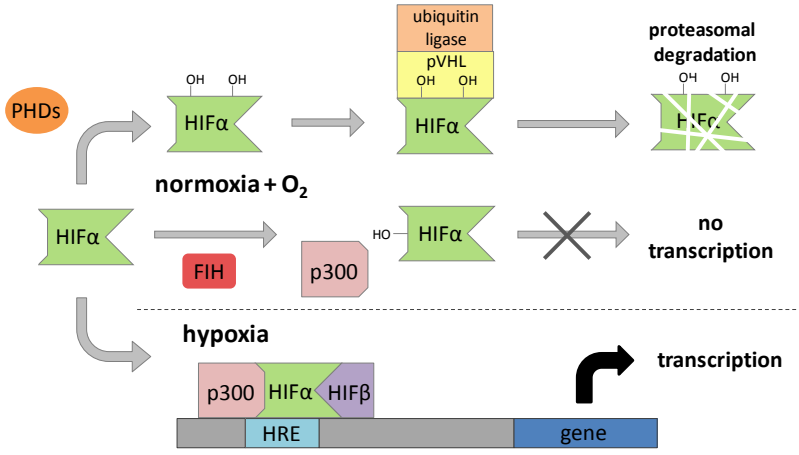


Figure 1-3: Schematic representation of the HIF (hypoxia inducible factor) hypoxic response pathway. In normoxic conditions HIF is hydroxylated leading to either proteasomal degradation or blocking of the interaction with p300. Under hypoxic conditions hydroxylation is prevented, leading to formation of the p300-HIFα-HIFβ complex, which initiates transcription from hypoxia response elements (HREs).

1.1.4 Histone modifications

Histones are required for the packaging of DNA into chromatin. 2OG oxygenases have important roles in epigenetic modifications to the *N*-terminal tails of histones, which are subject to many post-translational modifications including methylation, phosphorylation and acetylation[24].

Methylation of lysine or arginine residues in histones has a complex effect, and may regulate transcription sterically, possibly blocking recruitment of other histone binding or modifying proteins[24]. However, these modifications are not static and several histone demethylases have been identified, many of which are members of the 2OG oxygenase family[25-26]. These demethylases have differing specificities for the spectrum of lysine methylation states, the structural rationalisation for which is currently under investigation[27]. The demethylation of lysines occurs in a similar manner to the demethylation of DNA, with an initial hydroxylation producing an unstable intermediate, which collapses to release formaldehyde, and reveal the unmodified lysine[25].

1.1.5 RNA splicing

JMJD6 catalyses C^5 lysyl hydroxylation in arginine-serine rich domains of RNA splicing-related proteins and histones[28], with recent crystal structures rationalising the preference of JMJD6 for C^5-hydroxylation rather than lysyl N^{ϵ}-demethylation[29], and displaying possible binding of ssRNA[30]. The requirement of JMJD6 for molecular oxygen suggests that oxygen availability may regulate RNA splicing[28].

1.1.6 Fatty acid metabolism

Carnitine is a carrier molecule that transports long chain dietary fatty acids from the cytosol to the mitochondria for β-oxidation. The biosynthesis of carnitine requires two 2OG oxygenases, N^{ϵ}-trimethyllysine hydroxylase and γ-butyrobetaine hydroxylase (BBOX)[31]. A recent structural and mechanistic study of BBOX has revealed the mechanism of its inhibition by trimethylhydrazine-propionate[32], a compound that has

been used clinically to treat cardiac and cerebral ischaemia by preventing accumulation of cytotoxic fatty acid oxidation intermediates[33].

Phytanoyl-CoA hydroxylase (PAHX) also has a role in fatty acid metabolism. Phytanic acid is an intermediate in the digestion of chlorophyll and cannot be processed by the normal β-oxidation mechanism, as this is blocked by the presence of a β-methyl group. Instead, PHYH converts phytanic acid-CoA to hydroxy-phytanoyl-CoA[34], allowing subsequent removal of the α-carbon atom to form pristanic acid, which may be metabolised by the β-oxidation mechanism. Refsum's disease is characterised by an accumulation of phytanic acid[35], often caused by mutations in the *pahx* gene, many of which have been structurally rationalised[36].

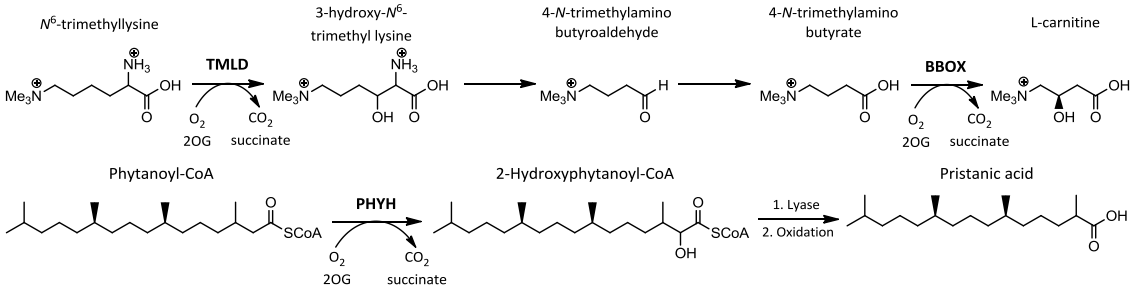


Figure 1-4: Reaction schemes highlighting the roles of 2OG oxygenases involved in fatty acid biosynthesis. TMLD: N^{ε}-Trimethyllysine hydroxylase; BBOX: γ-Butyrobetaine hydroxylase; PHYH: Phytanoyl-CoA hydroxylase.

1.1.7 Nucleic acid modifications

In addition to the reactions catalysed on protein and small molecule substrates, several 2OG oxygenases from a range of organisms have nucleic acid-based substrates, both acting on individual bases, or on oligonucleotide sequences.

Thymine 7-hydroxylase (T7H) is found in several species of fungi, including *Neurospora crassa* and *Rhodotorula glutinis*[37], as well as in bacteria (*Pseudomonas*), and kinetoplastids such as *Trypanosoma brucei*. *N. crassa* and *R. glutinis* can grow using thymine as their sole nitrogen source, via a thymine salvage pathway. The primary reaction of T7H is the three consecutive oxidations of the thymine methyl group, ultimately forming the carboxylic acid (Figure 1-5A)[38]. T7H also catalyses a

range of further reactions with thymine and its analogues, including demethylation, epoxidation and sulphur oxidation[39].

Fungi may also oxidise purines using xanthine hydroxylase, which oxidises xanthine to uric acid (Figure 1-5B)[40]. Discovery of this activity identified that fungi such as *Aspergillus nidulans* have two mechanisms for xanthine metabolism[40].

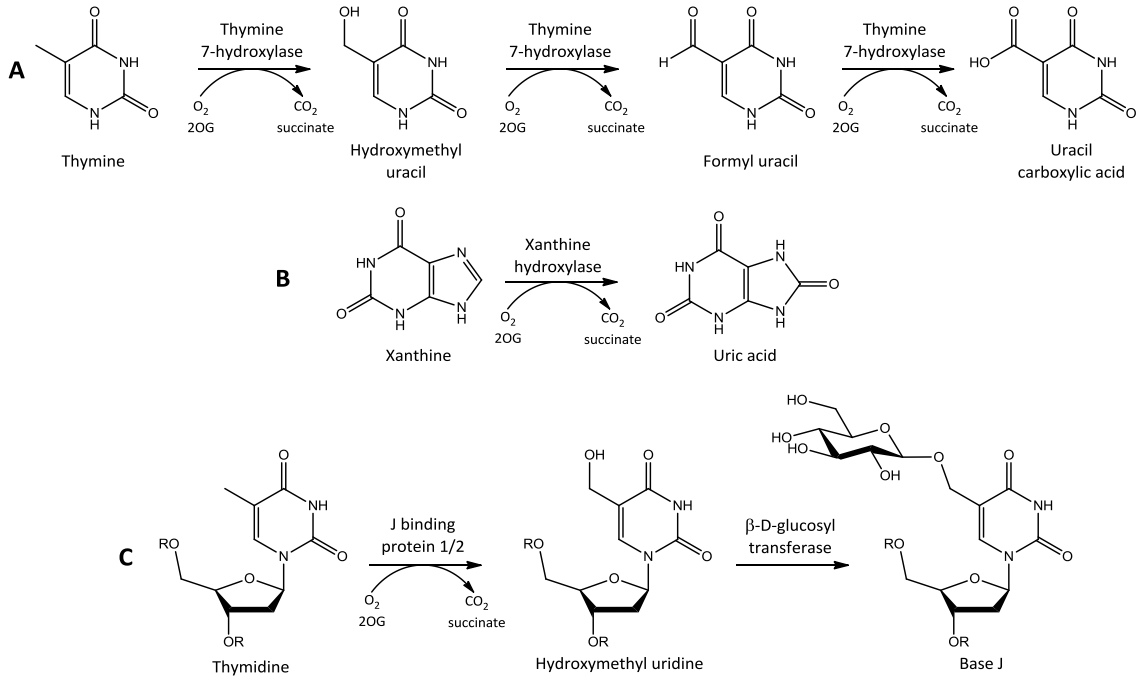


Figure 1-5: Summary of selected nucleic acid base modifications catalysed by 2OG oxygenases. (A) Oxidation of thymine to uracil carboxylic acid by thymine 7-hydroxylase; (B) Oxidation of xanthine to uric acid by xanthine hydroxylase; (C) Base J biosynthesis, involving the key hydroxylation step catalysed by J binding protein 1/2.

β-D-Glucopyranosyloxymethyluracil (base J) was the first hypermodified base identified in eukaryotic DNA, and is found in the bloodstream form of kinetoplastids such as *T. brucei*[41]. The biosynthesis of base J occurs via hydroxylation of thymidine by the 2OG oxygenases J binding protein 1 and 2 (JBP1 and 2) to give hydroxymethyl uracil[42], followed by glucosylation to yield base J (Figure 1-5C). *T. brucei* causes African Sleeping Sickness in humans[43], and is able to evade the host's immune system by expression of a variable surface glycoprotein (VSG) coat, encoded by subtelomeric sequences[44]. During infection of a mammalian host, trypanosomes are able to simultaneously switch off one VSG site and turn on another, a process known as antigenic variation[45]. Base J is usually telomeric, and it is thought that its presence may

induce compaction of chromatin, either leading to gene silencing[46], or stabilisation of chromosomes with respect to rearrangements of the subtelomeric VSG sequences[47].

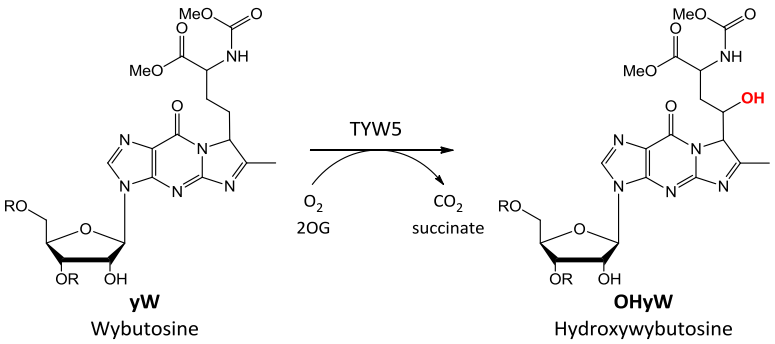


Figure 1-6: Hydroxylation of wybutosine by the human 2OG oxygenase TYW5, yielding hydroxywybutosine.

Human 2OG oxygenases modifying nucleic acids have also been identified. As predicted bioinformatically[48] TYW5 catalyses a hydroxylation in the biosynthesis of hydroxywybutosine (OHyW) at position 37 of $tRNA^{Phe}$ (Figure 1-6), although the function of this modification is currently unknown[49].

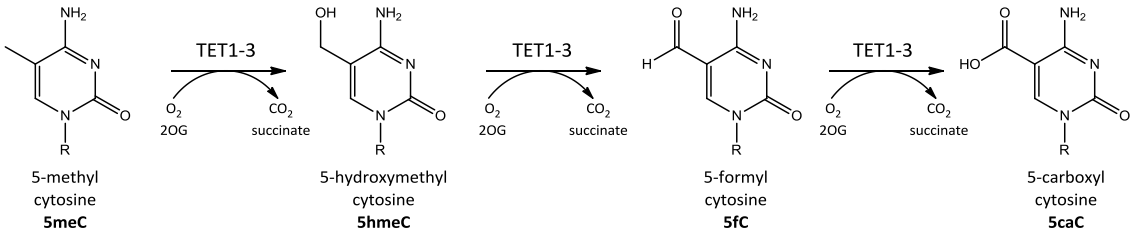


Figure 1-7: Reaction scheme for the oxidation of 5-methylcytosine by the TET 2OG oxygenases.

The most recently identified family of human 2OG oxygenases comprises the Ten-Eleven-Translocation proteins, TET1-3. Interest in the TET proteins was initially stimulated by discovery of a link between mutations in TET2 and myelodysplastic syndromes[50]. The TETs have greatest homology to the JBPs, and were found to catalyse a similar hydroxylation, that of the epigenetic marker 5meC to 5hmeC (Figure 1-7)[51-52]. Subsequently, it was found that the TETs are also able to further oxidise 5meC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), raising the exciting possibility that the reactions of the TETs may be involved in the demethylation of 5meC[53]. Additionally, reduced hydroxylation of 5meC has been observed for myeloid cancers

which carry mutant TET2, suggesting a possible rationalisation of the TET2 cancer link[54].

The final major class of nucleic acid modifying 2OG oxygenases includes the *E.coli* enzyme AlkB and its human homologues ABH1-8 and FTO. AlkB is a DNA demethylase, which repairs methylation damage to bases in oligonucleotide sequences via a direct reversal mechanism. Members of this 2OG oxygenase subgroup have been studied in this work, and these proteins will be discussed in more detail below.

1.1.8 2OG oxygenases in E. coli

In addition to AlkB, bioinformatic studies have identified the existence of the 2OG oxygenases YcfD, YbiX, YbiU, CsiD, and TauD in *E. coli*, although this list may not be exhaustive. TauD is known to be required for utilisation of taurine as a sulphur source under conditions of sulphate starvation[55], converting taurine to sulphite and aminoacetaldehyde[56]. Many mechanistic studies have been carried out on the TauD protein, which has led to much of the current understanding of the reaction mechanism of the 2OG oxygenases. Little is currently known about the functions of the remaining 2OG oxygenases, although the production of CsiD has been observed to be induced by carbon starvation[57], and YcfD has been recently assigned as an arginine hydroxylase (Dr Chia-Hua Ho, CJS laboratory, unpublished results).

1.2 Nucleic acid structure and modifications

Nucleic acids are essential for life; deoxyribonucleic acid (DNA) acts as a long-term store of genetic information, and ribonucleic acid (RNA) has central roles in gene expression and protein synthesis. Maintenance of the integrity of nucleic acids is therefore vital. As early as the 1930s, before the chemical structure of double stranded DNA (dsDNA) was elucidated in 1953[58], scientists were aware that ionising and UV radiations were damaging to the genetic material of cells, but that recovery from this damage was possible[59].

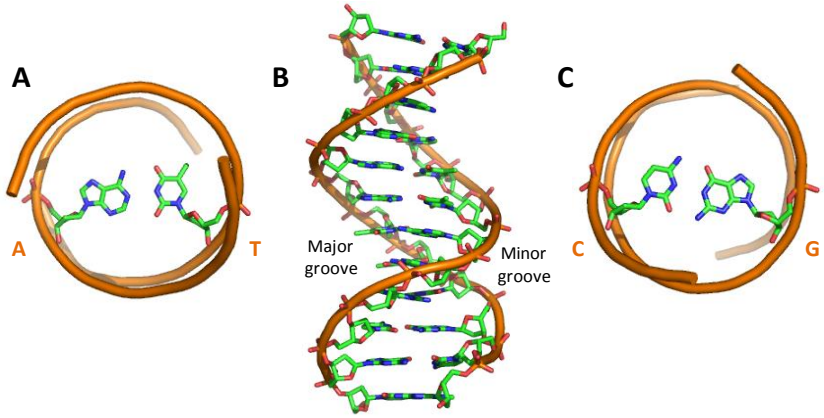


Figure 1-8: The structure of double-stranded B-DNA (dsDNA). (A) View down the 5'-3' axis highlighting a TA base pair and its 5' phosphates. (B) The double helix viewed perpendicular to the 5'-3' axis; (C) Alternate view down the 5'-3' axis highlighting a CG base pair and its 5' phosphates. Figure prepared from X-ray crystal structure Protein Data Bank (PDB) ID: 1BNA[60].

The most common B-type structure of dsDNA is outlined in Figure 1-8, highlighting complementary base pairing and formation of the double helix. DNA is proposed to encode for the "genome" of most organisms, and is more stable than RNA due to the presence of deoxyribose in the sugar-phosphate backbone, which is less susceptible to hydrolysis than the ribose sugar of RNA.

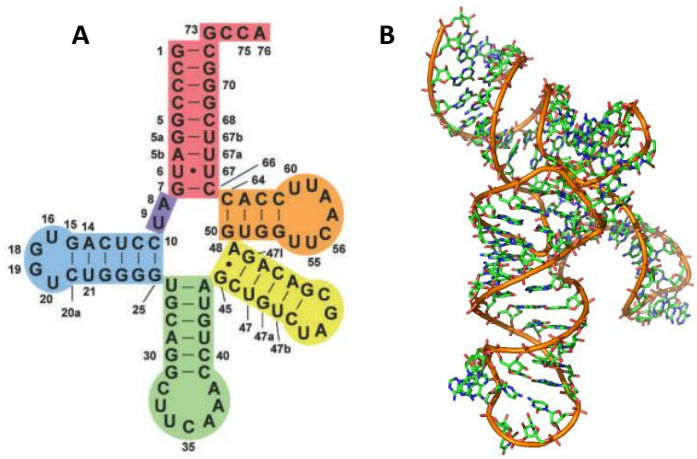


Figure 1-9: Structure of human tRNA^{Sec} represented as (A) Cloverleaf model, with numbering of residues. Acceptor arm: red, AD linker: purple; D arm: blue; anticodon arm: green; extra arm: yellow; and T arm: orange. (B) X-ray crystal structure, PDB ID: 3A3A[61].

Although RNA is usually singly stranded, the complex three dimensional structures it adopts are more varied than those of DNA, reflecting its functional diversity. Whereas messenger RNA (mRNA) encodes the transcribed DNA sequence, and transfer RNA (tRNA, representative structure in Figure 1-9) carries amino acids to the ribosome facilitating protein transcription[62], ribosomal RNA (rRNA) has a catalytic role in

protein synthesis as a peptidyl transferase[63]. Short RNA sequences have regulatory roles, with micro RNA (miRNA) and small interfering RNA (siRNA) regulating gene expression[64]. *In vivo*, tRNA is very stable, whereas mRNAs can have very different half-lives. Furthermore mRNA half-lives can vary significantly between kingdoms, typically ~15 min in *Escherichia coli*[65], but several days in humans[66].

Modifications to the chemical structure of nucleic acid bases are common, either through intentional structural modifications or damage from endogenous or exogenous sources.

1.2.1 Controlled DNA modifications

Controlled DNA methylation is widespread in prokaryotes and eukaryotes as a regulatory mechanism. Methylation commonly occurs through methyl transfer from *S*-adenosyl-L-methionine (SAM) by specific methyl transferases[67].

Prokaryotic DNA is subject to several methylations, forming the modified bases 5meC, N^4meC, and N^6meA (Figure 1-10, see Appendix A for explanation of structure nomenclature). These methylations may have evolved to protect bacterial DNA from endonuclease digestion by viruses, however they also fulfil several other roles[68]. Establishment of a specific methylation pattern allows host-pathogen recognition, with incorrectly methylated DNA subject to endonuclease degradation. This appears to be the major role of both 5meC and N^4meC modifications in prokaryotes, whereas N^6meA has multiple additional functions[69]. N^6meA methylation occurs shortly after DNA synthesis, ensuring that immediately after replication only the parent strand will be methylated. This enables distinction from the daughter strand allowing any replication repairs to be corrected using the appropriate strand as a template. The N^6meA base is also present in fungi and protists, and is important for signalling sites of protein-DNA interactions, and for the control of genome replication[70].

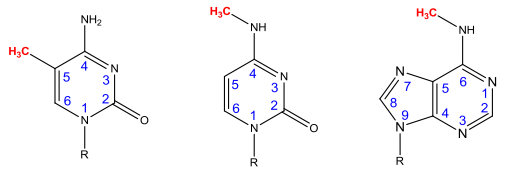


Figure 1-10: Structures of modified bases used as regulatory signals. All three modified bases are used by prokaryotes, whereas 5meC is additionally found in eukaryotes.

The only identified regulatory DNA methylation observed in higher eukaryotes is 5meC. In plants the 5meC modification occurs in a range of sequence contexts[71], but in vertebrates 5meC methylation is specific to CpG sites[72]. CpG dinucleotides are present throughout the genome, but also cluster at "CpG islands" which are common near transcriptional start sites[73].

Methyl-DNA binding proteins recognise 5meC and bind to the methylated DNA leading to recruitment of histone modifiers, and transcriptional silencing[74]. Furthermore, the presence of the 5meC methyl groups can prevent binding of transcription factors to DNA, further preventing transcription[75]. Away from CpG islands 5meC appears to be involved in maintenance of the pluripotent state in embryonic stem cells[76]. DNA methylation is a long-term form of repression, and is used for silencing of imprinted genes, the inactive X chromosome, and transposons[77]. However, aberrant cytosine methylation has been linked to many human diseases, including cancer and psychiatric disorders[78].

Hydroxylated 5meC (5hmeC) was first detected in rat DNA in the 1970s[79], but was only recently detected in human DNA[80], and has since been identified in many tissues[81]. Currently the role of 5hmeC is unclear, however it could be an intermediate in the demethylation of 5meC, or may be associated with regulation of gene expression, possibly by alteration of local chromatin structure[77].

1.2.2 Controlled RNA modifications

RNA modifications are widespread; to date over 100 modified RNA nucleotides have been described in prokaryotes and eukaryotes[82]. The creation of mature mRNA capable of translation requires the addition of a 5' "cap" following transcription (Figure 1-11).

This involves addition of a guanosine via an unusual 5' to 5' triphosphate linkage, followed by methylation to yield 7meG. This cap is also the substrate for further methylations, including N^6meA, 2O'meG, and 2O' methylation of the penultimate base. These modifications provide significant resistance against 5' exonucleases, and allow regulation of nuclear export, and intron excision[83]. 2O' methylation also appears important in distinguishing self from non-self RNA[84].

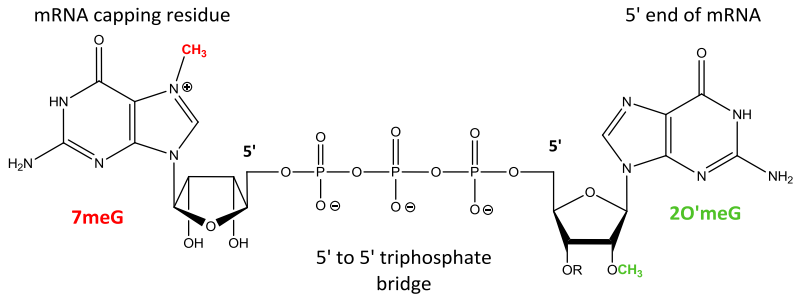


Figure 1-11: Structure of the 5' mRNA cap required for mature mRNA. A terminal guanosine is added via a 5'-5' triphosphate link, and methylated to form 7meG. The penultimate guanosine is commonly also methylated on the ribose ring, giving 2O'meG.

The broadest range of modifications is found in tRNA, and modifications to tRNA are crucial for three dimensional structure formation and the prevention of translational defects, which have been implicated in human disease[85].

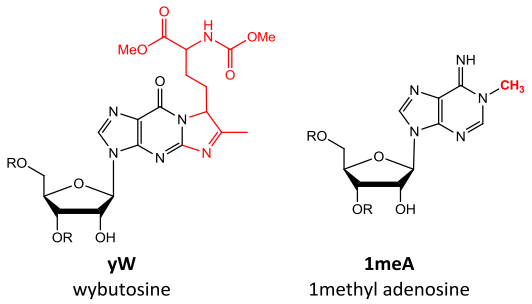


Figure 1-12: Structures of RNA bases commonly modified in tRNA.

Wybutosine (Figure 1-12) is a hypermodified guanosine required for stabilisation of codon-anticodon pairing maintaining the reading frame during translation[86]. The modified adenosyl residue 1meA (Figure 1-12) is formed by the post-translational methylation of adenine in RNA by methyl-1-adenosine transferase[87], and is involved in directing tRNA cloverleaf folding[88] (see Figure 1-9), most likely due to the prevention of standard Watson-Crick base pairing[89]. 1meA also alters the structure of dsDNA[90].

~25% of tRNAs have 1meA at position 58 in the T loop, and 1meA is also commonly found at position 14 in the D loop[91].

1.2.3 DNA damage

1.2.3.1 Endogenous DNA damage

DNA damage occurs continuously in cells, and may be caused by endogenous agents. The methyl group donor SAM is a weak methylating agent and can non-enzymatically methylate DNA. In the presence of 40 μM SAM ~600 cytotoxic 3meA lesions (Figure 1-13) are formed per day[92], a level similar to that expected from continuous exposure of cells to 20 nM methyl methanesulfonate (MMS), a potent methylating agent. In comparison to this a significantly greater quantity of purine bases (2000-10000 per day) are turned over due to depurination and subsequent repair[93]. A very low rate (about 10 per day) of glucose addition to the amino groups of bases has also been observed[94]. Products of lipid peroxidation commonly react with DNA forming exocyclic adducts, such as 1,N^6-ethenoadenine (εA) and 3,N^4-ethenocytidine (εC), shown in Figure 1-13[95]. These bulky lesions block base pairing, and are highly mutagenic[96].

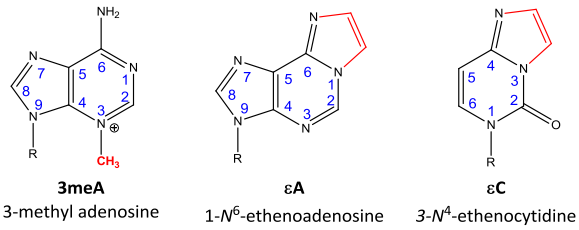


Figure 1-13: Structures of modified bases commonly formed by endogenous alkylation damage.

1.2.3.2 Methylation

Methylation of DNA is also caused by exogenous reagents, which are classified according to the regiochemical outcome of the methylation[97]. S_N1 alkylating reagents include *N*-methyl nitrosourea, the anti-cancer drug Temozolomide, and chemicals such as the potent carcinogen 4-(methylnitrosoamino)-1-(3-pyridyl)-1-butanone (NNK) found in tobacco smoke[98] (Figure 1-14). These methylate DNA at nitrogen and oxygen atoms (including those in the phosphate backbone), as shown in Figure 1-15. In

contrast, S_N2 alkylating reagents include methyl halides, methyl methanesulfonate (MMS) and SAM (Figure 1-14); these methylate more commonly on nitrogen atoms[97].

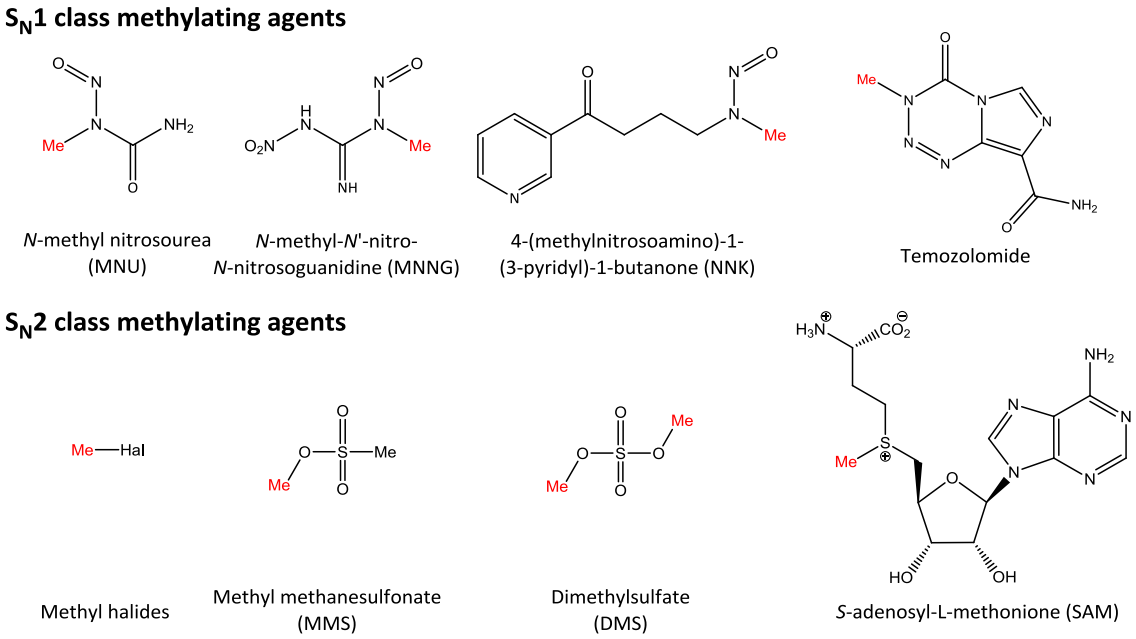


Figure 1-14: Common nucleic acid methylating agents, classified by mechanism. S_N1 class reagents preferentially react at oxygen atoms, but also react at nitrogen, whereas S_N2 reagents tend to react at nitrogen atoms.

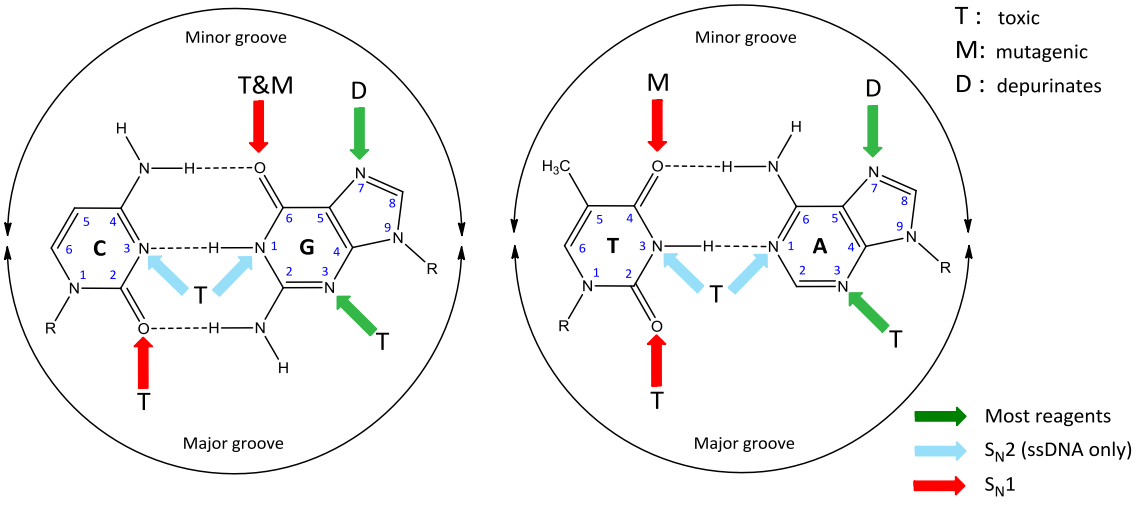


Figure 1-15: Regiochemistry of S_N2 and S_N1 methylating reagents. Arrows indicate positions of methylation by different reagents. Watson-Crick hydrogen bonds are indicated as dashes. Figure adapted from[99], with effect of methylation from[97].

The major adduct formed in dsDNA is 7meG (Figure 1-16). This does not block replication or cause miscoding[100], but addition of the methyl group at position 7 makes the *N*-glycosidic bond more susceptible to cleavage as the methylated 7meG base is a

better leaving group than G alone, leading to depurination[101]. O^6meG (Figure 1-16) is formed to a lesser extent, but is highly mutagenic and genotoxic as it readily mispairs with T, inducing mutations[102], and is considered to be the most carcinogenic methylated base formed[100]. As this lesion is formed predominantly by S_N1-type reagents, this explains the higher toxicity of this class of reagents, and their exploitation in cancer therapies[103]. Other lesions have a lower incidence of formation, therefore have been studied in less detail. However, O^4meT (Figure 1-16) has a similar effect to O^6meG [104], and 3meA (Figure 1-13) slows replication[105].

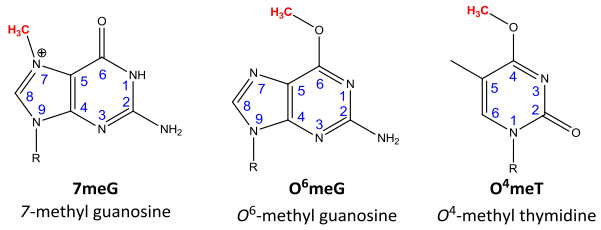


Figure 1-16: Structures of modified bases commonly formed by exogenous methylating agents.

Significantly different methylation patterns are observed on methylation of ds compared to ssDNA, due to the relative availability of different reaction sites in the absence of hydrogen-bonding between bases (Figure 1-15)[97]. 1meA and the analogous methylations (Figure 1-17) are formed much more commonly in ssDNA than dsDNA. The nucleophilicity of the N1-A and N3-C atoms is higher than that of N1-G and N3-T, due to the adjacent amino (rather than carbonyl) groups, therefore these are more reactive to methylating agents. In *E. coli*, 3meC was found to be the most mutagenic of this family of lesions, although all of them block replication[106]. Methylation of RNA apparently follows a similar pattern to that observed in ssDNA[107].

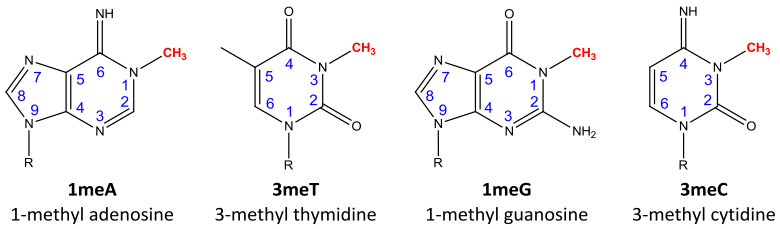


Figure 1-17: Structures of modified bases commonly formed by exogenous methylating agents in single stranded nucleic acids.

1.2.3.3 DNA strand breaks

Breaks in the DNA sequence can occur in one or both strands of the DNA helix, with dsDNA breaks being most harmful as no template is present for their repair. The broken strands may freely diffuse away from each other, leading to broken chromosomes, genomic rearrangements, or cell death[108]. dsDNA breaks are necessary intermediates in certain cellular processes, such as V(D)J recombination, which occurs in developing B- and T-lymphocyctes[109]. However they may also be formed when replication is blocked, which can be caused by oxidative lesions or the presence of ss breaks. Additionally breaks may be caused by ionising and UV radiations, or mechanical stress[108]. DNA strand breaks are also caused by methylation, with both MMS and MNNG shown to induce sister chromatid exchanges[110]. These may be initiated by otherwise relatively innocuous lesions such as 3meA, due to stalled replication forks or repair intermediates[111].

1.2.3.4 Further modifications

Oxidation of bases by reactive oxygen species (ROS) is common, particularly in the mitochondria where concentrations of these species are highest. Spontaneous hydrolytic deamination of bases occurs *in vivo*, particularly converting C to U, and 5meC to T[94]. On treatment of DNA with bisulfite *in vitro* the C to U conversion occurs preferentially, a reaction which is exploited in the bisulfite DNA sequencing protocol[112].

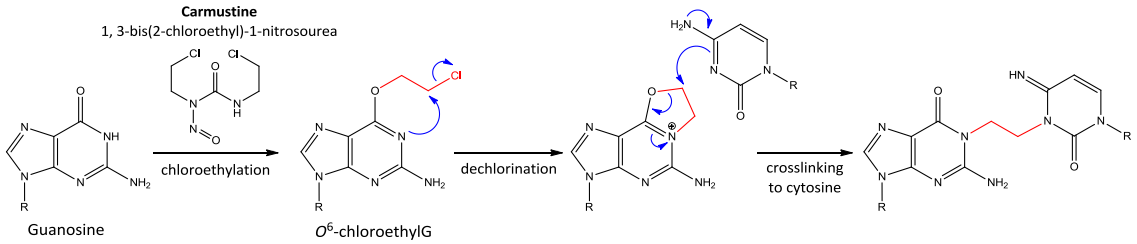


Figure 1-18: Reaction scheme for formation of crosslinks between G and C bases, initiated by reaction with Carmustine.

DNA cross links are also commonly formed, and may be caused by UV light or the chemotherapy agents cisplatin and nitrogen mustards. Cross links are also formed by bifunctional chloroethylating agents such as 1, 3-bis(2-chloroethyl)-1-nitrosourea

(Carmustine). These initially form O^6-chloroethylG, which dechlorinates then crosslinks C1-G to C3-C (Figure 1-18)[113].

1.3 Repair pathways

1.3.1 Base excision repair

The base excision repair (BER) pathway (Figure 1-19) responds to relatively minor disturbances to the overall DNA helical structure caused by small modifications to individual bases including oxidation, methylation, deamination, and depurination[114]. The process is initiated by one of a number of specific DNA glycosylases that recognise the altered base and cleave the glycosylic bond via a ring-flipping mechanism, leaving an abasic (AP) site. This is processed by an AP endonuclease followed by a diesterase creating a free 3'-OH terminus, to which polymerase β may add a new base, then the nick is sealed by a DNA ligase[115]. This mechanism is commonly employed in the repair of 3meA, either by AlkA in *E. coli*[116], or by 3meA-DNA-glycosylase (AAG) in humans[117]. 8-oxoG, formed by ROS, is also repaired by this mechanism, by the OGG1 glycosylase[118].

1.3.2 Nucleotide excision repair

Nucleotide excision repair (NER, Figure 1-19) removes a broad spectrum of DNA lesions that cause structural deformation of the DNA helix, including large hydrocarbon adducts, and pyrimidine dimers caused by UV irradiation. This mechanism is the primary defence against exogenous DNA mutagens[119] and in mammals involves over 25 proteins[120]. The deformation of the helix is recognised by a protein complex, which causes excision of a single-stranded portion of the DNA sequence (usually 25-32 nucleotides), followed by replacement of the missing sequence by DNA polymerases and DNA ligases[120].

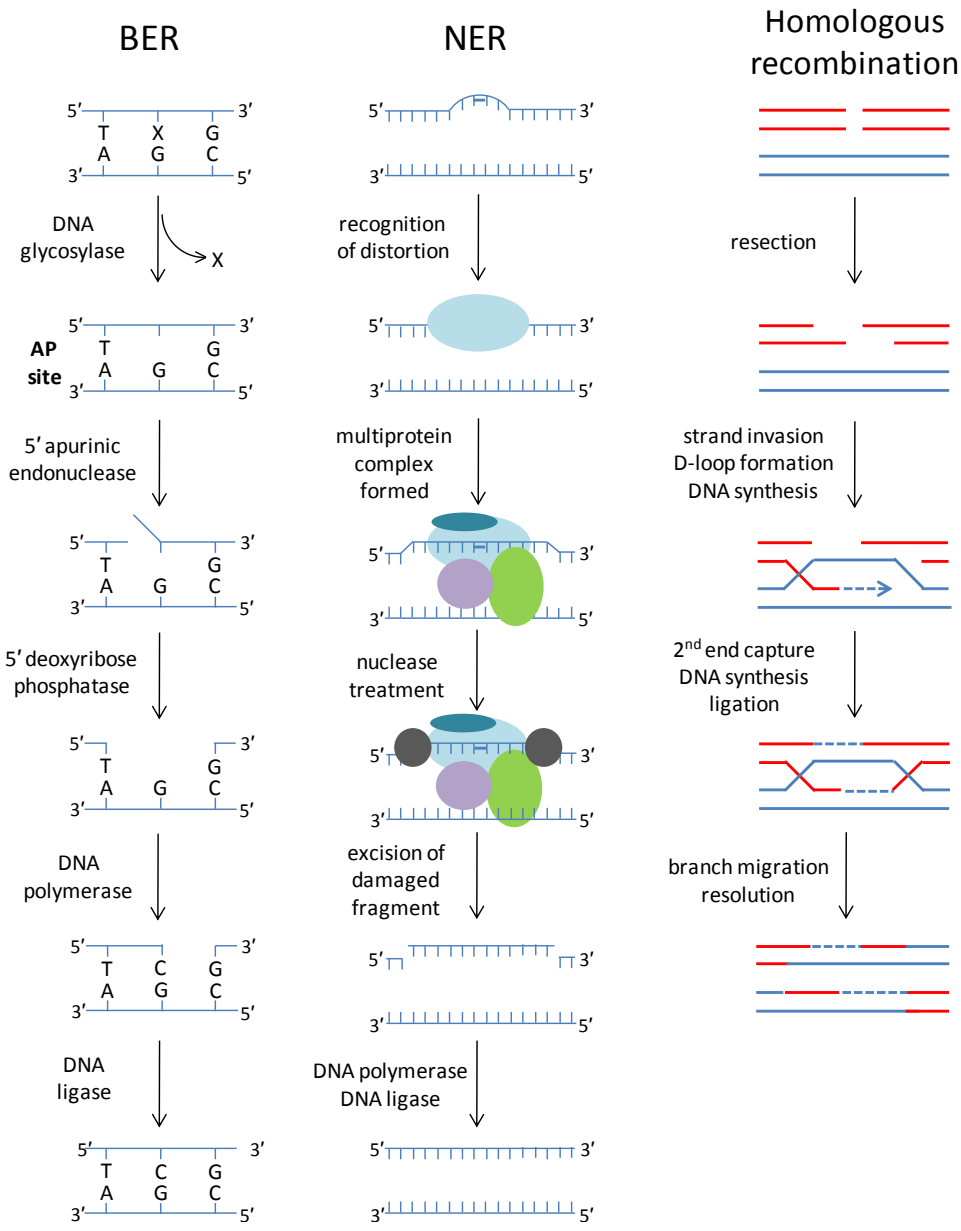


Figure 1-19: Schematic representations of three major classes of DNA repair mechanism. BER (base excision repair) of a base with a small chemical lesion, e.g. 3meA; NER (nucleotide excision repair) of a larger DNA modification e.g. thymine dimer; and homologous recombination repair of a double strand DNA break.

1.3.3 Mismatch repair

Mismatch repair (MMR) occurs post-replication, with repair proteins binding directly to single base mismatches and insertion/deletion loops in the DNA sequence[121]. The repair process is similar to that of BER, with DNA nicked, the lesion excised, repaired, then the incision closed by a ligase[121].

1.3.4 Translesion synthesis

Translesion synthesis (TLS) was recently identified as a distinct repair mechanism carried out by novel polymerases that are capable of synthesising DNA without the normal requirement for hydrogen bonding to a template[122]. Lesions normally recognised by BER or NER may be repaired, including thymine dimers and 8-oxoG, therefore TLS is a useful rapid alternative to classical repair methods, and may be particularly useful in S-phase to prevent ds-break formation, but is low-fidelity[123].

1.3.5 Recombinational repair

Repair of disruptive dsDNA breaks occurs via recombination, in either a homologous or non-homologous manner. Homologous recombination uses homologous sequences (e.g. from a sister chromatid) as a template for synthesis (Figure 1-19)[124]. Non-homologous end-joining is more rapid, but is error-prone, as the broken ends are held together without the guidance of a template sequence[125].

1.3.6 Direct reversal

Direct reversal (DR) repair mechanisms are somewhat different from the other processes discussed, as they do not involve removal of the damaged nucleotide. Instead, repair proteins act directly on the damaged base, removing the lesion and restoring the native base. Enzymes acting by this mechanism include photolyases[126] (repairing UV induced pyrimidine dimers), Ada and MGMT (repairing O^6MeG), and members of the AlkB family (removing small alkylation lesions)[127].

1.3.7 Sanitation of the dNTP pool

Damage may also occur to the unprotected dNTP precursors required for DNA synthesis by DNA polymerases, and mechanisms are required to prevent incorporation of incorrect and potentially mutagenic dNTPs into DNA. These damaged dNTPs are hydrolysed to the monophosphates, which are unable to be incorporated into DNA. In particular 8-oxo dGTP and 2-OH-dATP (Figure 1-20) are generated by oxidation of DNA precursors in the nucleotide pool. These are hydrolysed by the

pyrophosphohydrolase MTH1[128], which is found in both the cytosol and mitochondria[129].

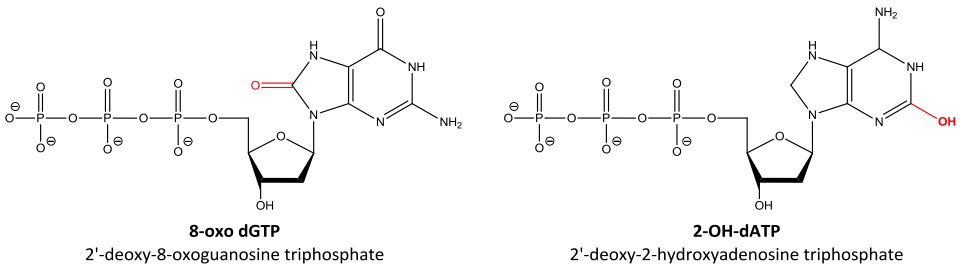


Figure 1-20: Structures of oxidised dNTPs generated in the nucleotide pool, and repaired by MTH1.

1.4 Comparison of nuclear and mitochondrial DNA repair

Mitochondria contain a separate autonomously replicating DNA genome (mtDNA). This has high organisational economy with no introns, and very few non-coding bases between coding sequences[130]. The mitochondrial genome appears essential, as cells lacking mtDNA do not respire[131], and high levels of mtDNA mutations promote apoptosis and premature aging[132]. Maintenance of the sequence integrity of this separate genome through damage repair is therefore important, and it is likely that mtDNA repair takes place in a complex associated with the mitochondrial inner membrane, possibly as a component of the mitochondrial nucleoid[133]. In general, it appears that similar repair mechanisms exist in mitochondria to those utilised in the nucleus, but that the mitochondrial mechanisms are simpler, using fewer proteins.

Of the mechanisms discussed above, some classes of DR and BER have been definitively identified in mitochondria. A splice variant of MGMT carries out DR of O^6meG and a limited range of related lesions[134], however no photolyases or homologues of the AlkB family have been identified. The BER mechanism is particularly important in mitochondria due to the presence of high levels of ROS including superoxide, which is produced as a byproduct of respiration at ~0.1 % of the respiratory rate[135]. The most common oxidised lesion is 8-oxoG (Figure 1-20), repaired by a mitochondrial-directed transcript of OGG1[136].

The NER pathway has not yet been observed in mitochondria, which may suggest that a subset of larger lesions are not repaired, but as many copies of mtDNA are present in cells (~3650 per cell in skeletal muscle[137]) this may not be as problematic as for nuclear DNA. An additional mechanism existing in mitochondria involves binding of the mitochondrial transcription factor A (TFAM) to damaged mtDNA forming a physical shield[138], possibly protecting the DNA from further damage[139].

1.5 E. coli *adaptive response*

Bacteria use an inducible "adaptive response" to enhance resistance to fluctuating levels of methylating agents in the environment. This response has been studied in greatest detail for *E. coli*, where the adaptive response involves increased expression of four genes: *ada*, *alkA*, *alkB* and *aidB*[140].

The Ada protein is multifunctional, and is both a positive regulator of the adaptive response and a DNA demethylase. Ada and AlkB are expressed from the same operon, with the first adenine of the *alkB* ATG initiation codon overlapping with the final adenine of the *ada* TAA termination codon[141]. -10 and -35 promoter sequences appear to be present upstream of the *alkB* gene, but these are separated from the gene by a greater distance than is normal in *E. coli*, so activity, if any, from this promoter would be likely to be very low. Further studies using an *alkB*/*β*-galactosidase fusion showed high *β*-galactosidase activity due to the action of the *ada* promoter, suggesting that *ada* and *alkB* are cotranscribed[141] Under normal circumstances both proteins are present at the same low expression level of about 2-4 molecules per cell[142]. The *C*-terminal domain of Ada demethylates toxic and mutagenic O^6meG and O^4meT lesions (Figure 1-16) by rapid, but suicidal, transfer of the methyl group to a buried cysteine residue[143]. Alongside this reaction, the *N*-terminal domain of Ada demethylates methylphosphotriesters, which have no other known repair mechanism in *E.coli*, with the methyl group transferred to a cysteine residue in the protein active site. HUMAN Methylation of this domain acts as a sensor for DNA damage converting Ada to a transcriptional coactivator, increasing its binding to the promoters of the *ada-alkB*

operon, and the *alkA* and *aidA* genes[144]. Following repair of all methylphosphotriesters, non-methylated Ada accumulates in the cell, inhibiting *ada* gene activation, and turning off the adaptive response[145].

AlkA is a DNA glycosylase with a broad substrate range, repairing bases with alkyl (most commonly 3meA[116]) and other more complex lesions[146]. Damaged bases are recognised by their altered charge distribution[147]. However, slow cleavage of the glycosyl bond of the undamaged A and G bases is also observed when AlkA is overexpressed, showing the importance of inducing AlkA expression only when necessary[148].

AidB is the least well understood component of the adaptive response, and its function is currently unclear. AidB binds DNA[149] and RNA[150], although a crystal structure has shown that the nucleic acid binding pocket is distantly located from the likely active site[151]. Limited dehydrogenase activity is observed[152] and the crystal structure shows a distinctive FAD binding site[151]. Based on the small size of the active site it has been postulated that AidB may play a protective role by detoxification of alkylating reagents[151], although initial investigations did not identify turnover of MNNG[153]. In addition to its regulation by Ada, AidB also has an autoregulatory function, and represses its own synthesis under conditions of normal cell growth[149]. AidB also shows tight binding to flavodoxin and acyl carrier protein, with possible electron transfer between these species[153]. Most recently, AidB has been found to bind to UP elements, transcription enhancer sequences upstream of many highly expressed genes. It appears that through this action the AidB nucleic acid binding domain may target the protein to specific genes, preferentially protecting them from alkylation damage[150].

The final enzyme in the adaptive response is AlkB, a Fe(II) and 2OG dependent oxygenase that removes lesions from bases via a novel DR mechanism. This process will be described in more detail below.

1.6 AlkB

1.6.1 AlkB is a nucleic acid demethylase

The *alkB* gene was first identified through control of the sensitivity of *E. coli* to MMS, with alkB mutants formed by MNNG treatment unable to reactivate MMS treated λ phage[154]. Mapping of the gene on the *E. coli* genome showed that it encodes a protein of approximately 24 kDa[155]. The AlkB protein was purified, but no activity was observed in the standard activity assays carried out, testing for ATPase, glycosylase and methyltransferase activities, Ada proteolysis and Ada promoter binding[141].

The enzymatic basis for the survival advantage conferred by AlkB, and therefore the protein's function, was a mystery for many years, although many pieces of evidence were gathered suggesting the likely enzymatic activity. Recombinant AlkB binds more strongly to ssDNA than dsDNA, and preferentially binds DNA treated with methylating agents[156]. Furthermore, AlkB repairs damage to ss phage introduced by S_N2 agents, but not by S_N1 agents or γ rays, and AlkB mutants are less sensitive to MMS in stationary phase, when less ssDNA is present. As 1meA and 3meC lesions are induced by MMS more readily in ssDNA these were therefore posed as possible candidates for AlkB activity, but no observation of active removal of radiolabelled 1meA or 3meC was observed using traditional assay methods[156].

The final piece of evidence required came from bioinformatic studies carried out using a protein fold-recognition method. These suggested that AlkB characterises a novel family of Fe(II) and 2OG dependent oxygenases, and is likely to function through catalysis of oxidative detoxification of alkylated DNA bases, and possibly also by action on ssRNA[157]. This work therefore revealed the unexpected cofactor requirements, and AlkB was found to reverse 1meA and 3meC DNA lesions through an unprecedented oxidative demethylation mechanism (Figure 1-21), with the methyl lesion released as formaldehyde[158-159].

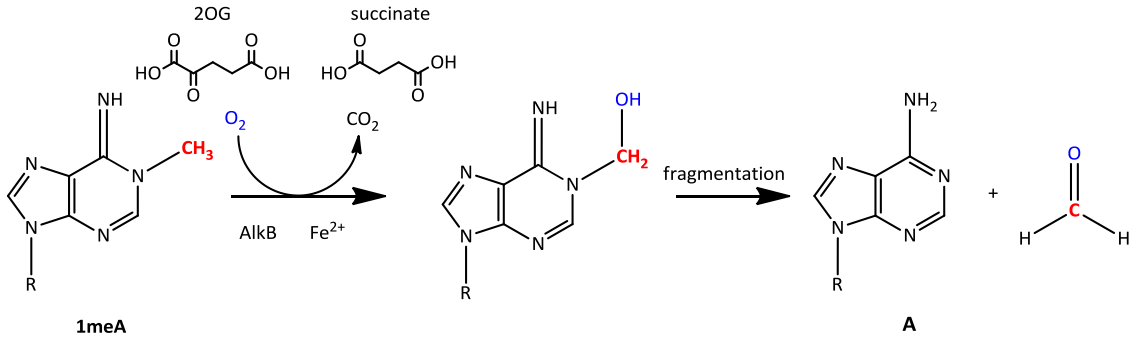


Figure 1-21: Reaction scheme for the demethylation of 1meA nucleic acid lesions in DNA by AlkB. The methyl group is initially hydroxylated by AlkB, then the unstable hydroxylated intermediate fragments to release formaldehyde and the repaired native base.

AlkB homologues have also been identified in plant viruses, despite their limited genome size[157]. Representative viral AlkB proteins from members of the *Flexiviridae* family have highest activity on 1meA lesions, but also repair 3meC[160]. Interestingly these proteins have a marked preference for repair of ssRNA, the first proteins of this family with such a preference, although as the plant viral genome is composed of ssRNA it is perhaps unsurprising. No AlkB homologues have been identified in Archaea, instead the representative family member *Archaeoglobus fulgidus* repairs 1meA and 3meC lesions by an AlkA homologue via the BER mechanism[161].

1.7 Human AlkB homologues

The first homologue of AlkB in *H. sapiens* (ABH) was identified in 1996[162]. Following determination of the enzymatic function of AlkB, further bioinformatic searches led to the identification of two further homologues, ABH2 and ABH3[163]. The remaining members of the family were found using a complementary "phylogenomic" analysis, whereby large scale searches were carried out in an attempt to make functional predictions via phylogenetic analysis. This approach classified a further five homologous proteins, present throughout the vertebrates (ABH4-8)[164], and allowed analysis of the phylogenetic, and therefore possibly functional, relationship between the family members (Figure 1-22).

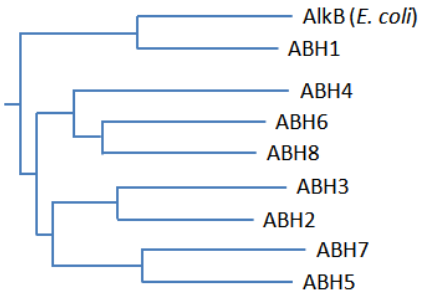


Figure 1-22: Maximum likelihood phylogenetic tree of the AlkB/ABH family, adapted from[164]. ABH2 and 3, ABH5 and 7, and ABH4, 6 and 8 are likely to be in-paralogous.

The final member of the family was not initially determined as a related 2OG oxygenase due to its significantly different sequence and therefore is still referred to by its original name of Fat Mass and Obesity associated protein (FTO)[165-166], although it has been termed by some as "ABH9" due to its sequence homology, and related function. Each of the family members retain the conserved iron, 2OG and substrate binding residues, although sequence analysis suggests that ABH8 and FTO contain additional protein domains (Figure 1-23).

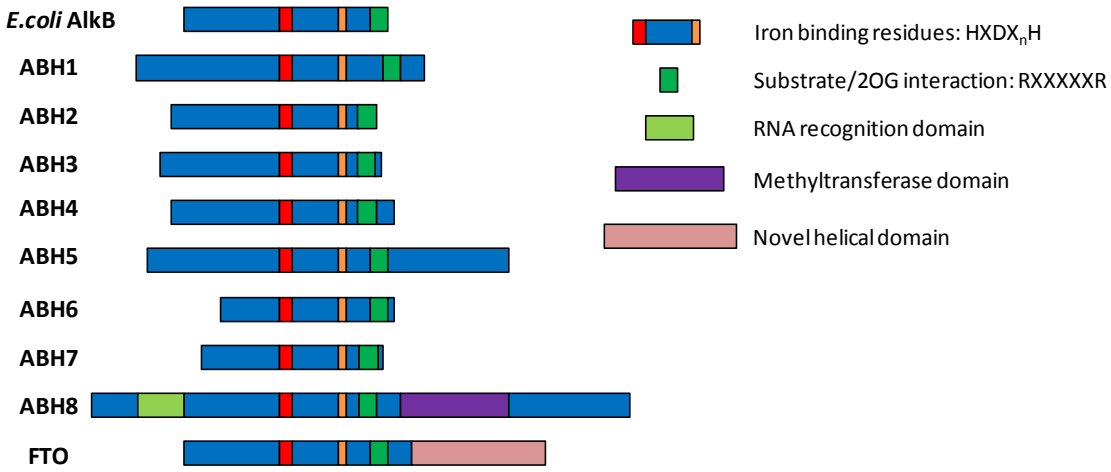


Figure 1-23: Cartoon alignment of ABH family showing the conserved and unique features. Adapted from[127].

1.7.1 ABH2 and ABH3

Studies of ABH2 and ABH3 revealed that both proteins have a similar function to AlkB, catalysing oxidative demethylation of 1meA and 3meC lesions[163,167]. ABH2 and ABH3 have been identified in many human tissues, with highest expression of ABH2 in the liver and bladder, and good expression levels in cultured HeLa cells. ABH3 has

highest expression in the spleen, prostate, bladder and colon, but is expressed at a very low level in cultured HeLa cells[163]. ABH3 is usually localised in the nucleoplasm, but is sometimes found as distinct spots in the nucleoli, or dispersed throughout the cytoplasm, whereas ABH2 is a nuclear protein, localising to replication foci in S phase, and dispersing throughout the nucleoplasm outside of S phase[167]. At replication foci ABH2 colocalises with PCNA, in an interaction mediated by a novel binding motif found in residues 1-7 of the ABH2 protein sequence[168].

Unsurprisingly, due to their roles in the repair of mutagenic and toxic DNA lesions, both ABH2 and ABH3 have been tentatively linked to cancer. Point mutations of either ABH2 or ABH3 have been identified in paediatric brain tumours (glioma, ependymomas and medulloblastomas)[169], although this study was too small to determine whether these mutations had any effect on the progression of the disease or the survival prognosis. However, both proteins were upregulated (ABH2 to a higher extent) in tumour tissues. The elevation of ABH2 may explain the resistance of some gliomas to photodynamic therapy treatment (photosensitisers injected into tumours are excited by light to produce ROS and thus cell death), through induction of ABH2 mRNA, which mediates repair of the DNA damage caused[170]. ABH2 induction appears to be mediated by the transcription factor tumour protein 53 (TP53), which is induced by photosensitiser treatment, and binds -190 bp to the ABH2 transcription start site[170]. Cell growth and viability in non-small cell lung cancer cell lines was unaffected by ABH2 knockdown, but when knocked down in tandem with cis-platin treatment, cell viability and proliferation rate was affected to a greater extent than cis-platin treatment alone, suggesting that combination therapies may be effective in lung cancer treatment[171]. Elevation of ABH3 protein levels has been observed in human prostate carcinomas[172], to an extent which the authors believe may enable its use as a novel prostate cancer-specific marker.

The murine homologues of ABH2 and 3 (mABH2 and 3) have been studied, and their reactions are similar to those of their human counterparts[173]. Mice in which *Abh2*, *Abh3*

or both of the *Abh2* and *Abh3* genes have been deleted appear to be healthy, with no obvious phenotype[174]. However, analysis of liver genomic DNA revealed that ABH2 deletion mice accumulate a detectable quantity of 1meA by 4 months of age, with about 150-500 1meA present in the genome at 8 months. In contrast, 8 month old ABH3 deletion mice have less than 120 1meA per genome[174]. Furthermore, the repair of MMS-induced 1meA lesions only occurred for nuclear extracts containing ABH2, but not if only ABH3 was present. It thus appears that ABH2 is the primary enzyme for the *in vivo* repair of 1meA. The accumulation of 1meA in the mouse genome also shows that a substantial degree of aberrant methylation occurs under normal physiological conditions.

1.7.2 Comparison of AlkB to ABH2 and ABH3

AlkB has a broader substrate range than ABH2 and 3, which is unsurprising as AlkB is the only homologue found in *E. coli.* The three proteins preferentially repair 1meA and 3meA lesions in DNA. AlkB and ABH3 also repair 1meA and 3meC in RNA[167], with very slow repair by ABH2; although this is more efficient when the ssRNA is annealed to a complementary DNA strand[175]. To date AlkB has been reported to also repair the DNA lesions 3meT[176-177], 1meG[177], 1ethylA[163], εA[178-179], εC[179] and ᴇA[180], summarised in Figure 1-24. ABH2 and ABH3 both repair 1ethylA[163] and 3meT[176], and εA[181] is repaired by mABH2. As discussed in Section 1.8.2 FTO is a further 3meT demethylase, possibly acting in a repair manner. The human BER enzyme AAG (see section 1.3.1) has also been found to repair εA[182], ᴇA[183], and 1meG[184]. These are all substrates of AlkB, and εA is a weak substrate of mABH2, therefore it appears that in humans some redundancy in reactivity may be observed. AlkB[159] and ABH3 both show a preference for ssDNA, whereas ABH2 prefers dsDNA substrates[167], although this preference is only seen either in the presence of magnesium[175] or at high NaCl concentrations[173].

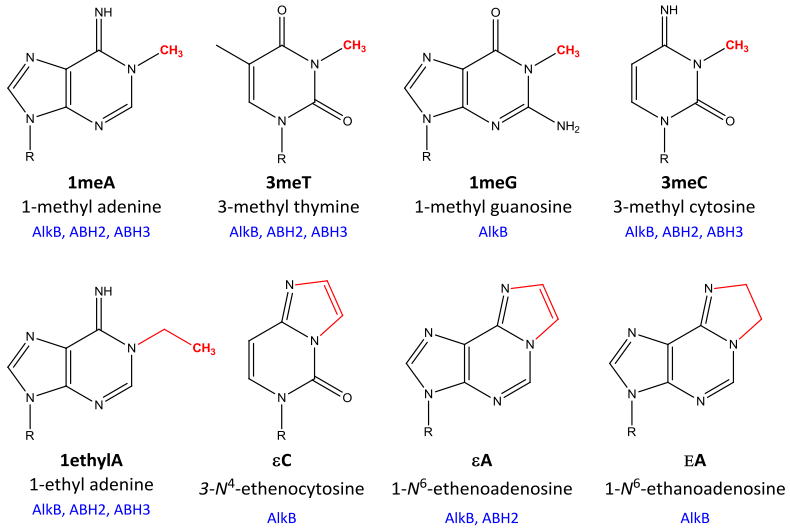


Figure 1-24: Summary of DNA lesions reported to be repaired by AlkB, ABH2 and ABH3.

Replacement of Fe in the active site of AlkB by either Co(II) or Cu(II) does not cause structural alterations, although it inhibits demethylation[185]. Ni(II) inhibits demethylation by ABH2 and ABH3. Replacement of Fe with Ni(II) does not prevent the binding of 2OG at the ABH2 metal centre[186-187], but it appears that on binding of 2OG and substrate no open coordination site remains to allow oxygen binding and activation[188].

Several studies have followed the activities of AlkB, ABH2 and ABH3 via *in vivo* assays, in several cases following the demethylation of MMS treated phage or plasmids, demonstrating that the demethylation reaction occurs on "real" sequences[156,163]. More sophisticated studies have also been developed, showing *in vitro* reactivation of methylated mRNA and tRNA by AlkB and ABH3[189], and the repair by *E. coli* of specific lesions in synthesised single stranded vectors, which additionally allowed assessment of the relative mutagenicity of the lesions[106,179].

A chemical crosslinking methodology has been developed to facilitate investigation of transient DNA-protein interactions, utilising DNA strands incorporating a modified cysteine residue, which forms a disulfide bond with an active site cysteine[190]. This was successfully employed with ABH2 and ABH3, confirming the preference of ABH2 for binding dsDNA[191].

Several crystal structures of the proteins have been published, revealing much about their observed specificities. AlkB[192] and ABH3[193] were initially crystallised without

DNA, but crosslinking of proteins to DNA has enabled structures of AlkB and ABH2 in the presence of bound DNA[194]. These structures do not show any obvious structural features discriminating between DNA or RNA substrates. However crystallisation of AlkB bound to a bulky caged DNA lesion shows that larger lesions may be accommodated in the active site by reorganisation of the flexible hairpin and substrate recognition loops[195]. Mutational analysis of the AlkB structure shows that the protein exists in two states: "search" and "repair". Individual residues have different roles, with Y76 acting as a gating residue, which is initially open to allow access of the DNA, then closes to lock the DNA in place to react[196]. The method used for locating the damaged bases requiring repair is currently unclear. The most likely method is that the proteins scan along DNA strands to detect any unusual or unstable bases or base pairs, and where these are observed the damaged base is "flipped out" from the base-stacked, hydrogen bound, intrahelical position to a solvent exposed, extrahelical position to enable repair[197].

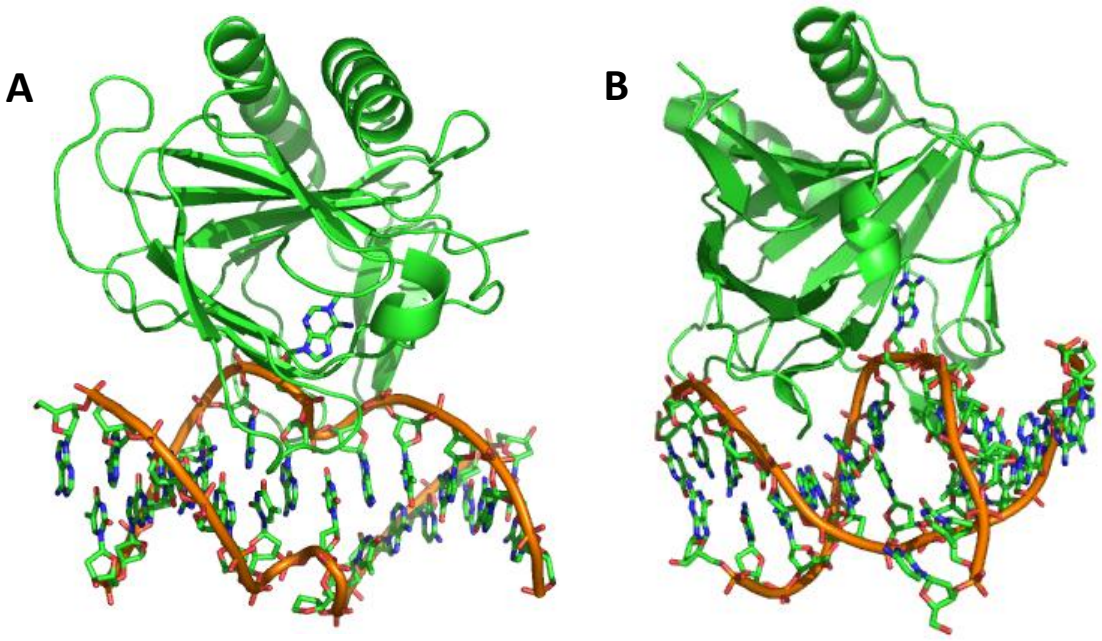


Figure 1-25: Cartoon representations of (A) AlkB (PDB ID: 3BIE[194]) and (B) ABH2 (PDB ID: 3BUC[194]) bound to double stranded DNA, highlighting the flipping of the damaged 1meA base into the protein active site.

AlkB and ABH2 appear to have very different base flipping mechanisms (Figure 1-25). AlkB makes contact mainly with the damaged base and the directly flanking sequence.

It then appears to kink the DNA backbone to invert the damaged base by 180°, simultaneously squeezing together the flanking bases to maintain base-stacking between these non-adjacent bases (Figure 1-25A)[194]. In contrast, ABH2 maintains contact with both strands of the dsDNA substrate, and on flipping of the base the vacant area in the DNA base-stack is occupied by the hydrophobic F102 residue from the β2-β3-hairpin, without causing significant distortion of the DNA double helix (Figure 1-25B)[194].

This β2-β3-hairpin is also responsible for the preference of ABH2 and ABH3 for double or single stranded DNA sequences respectively. In ABH2 this hairpin is composed of hydrophobic residues, whereas the hairpin of ABH3 contains charged or polar residues. Two independent studies have shown that exchanging either the full β-hairpin region, or just selected residues, is sufficient to switch the specificity of the proteins to that of the donor protein[198-199].

Comparison of AlkB crystal structures obtained to date shows the existence of two conformational states altering the relative accessibility of the oxygen diffusion tunnel, possibly to reduce ROS generation[200]. An NMR structure also showed conformational variation, with the structure of AlkB with succinate appearing more dynamic than that with 2OG. This may suggest that on decarboxylation of 2OG, a transition to an open dynamic state occurs, allowing release of succinate (and its replacement with fresh 2OG), and weakening the affinity for the DNA product[201].

EPR studies on AlkB-Cu(II) complexes showed that the metal coordination site is tetragonal with an elongated axial axis, with two equivalent histidine residues bound. In the presence of succinate one free binding site can be observed, as succinate binds in a monodentate manner[202]. Mechanistic studies of AlkB using density functional theory are in general agreement with the consensus reaction mechanism (Figure 1-2). The rate limiting step is predicted to be abstraction of a hydrogen atom from the methyl group, as the subsequent steps are very exoergic[203-204].

Some experimental characterisation of the oxidation intermediates of AlkB has been undertaken, using AlkB/substrate complexes crystallised anaerobically with Fe(II) and

2OG, followed by exposure to air to initiate reactions[205]. Structures were obtained for each of the major classes of AlkB substrates; neutral 3meT, positively charged 3meC and bulky εA. With the neutral 3meT base, the relatively stable hemiaminal intermediate may be observed in the crystal structure[205]. The electron density in the crystal structure with 3meC is not sufficient to definitively observe a reaction intermediate, but use of a carbon analogue for which the intermediate species is stable enough for detection reveals a zwitterionic intermediate. It is therefore likely that a zwitterionic species is also formed in the reaction with 1meA[205]. As has been previously predicted, a glycol intermediate is observed in the reaction with εA, likely formed from a rapidly opened epoxide species.

1.7.3 ABH1

Within the ABH family ABH1 has the greatest homology to AlkB (52 % similar, 23 % identical)[162] but the function of ABH1 has been elusive, and to date publications discussing the likely activity of ABH1 are not conclusive. Initial ABH1 studies were unable to observe *in vitro* activity related to that of AlkB[163,167], and attempts to crosslink dsDNA to ABH1, under the same conditions used for crosslinking to ABH2 and ABH3, were unsuccessful[191], suggesting that DNA may not be a substrate. Additionally, ABH1 was not observed to be induced in skin fibroblasts on MMS treatment[162], or to reactivate MMS treated ssDNA phage in *alkB* deficient *E. coli*[163], although its expression slightly increased the survival rate of MMS treated *E. coli alkB* mutants[162]. Furthermore, identification in mice of an interaction of ABH1 with Mrj, a protein which also interacts specifically with histone deacetylases, suggested a possible role for ABH1 in epigenetic regulation of gene expression, possibly through histone modification[206].

Two mouse ABH1 gene deletion studies have yielded contrasting results. The first study found the highest expression of ABH1 mRNA in the developing placenta, and ABH1$^{-/-}$ mice have interuterine growth retardation and placental defects. In a subsequent larger-scale mouse ABH1 knockout study it was found that mice lacking one or both copies of ABH1 are born at a lower frequency than wildtypes, with a sex-ratio skewed

towards males. Therefore, ABH1 appears to mediate gene regulation in spermatogenesis, during which expression of ABH1 is elevated 10-fold[207]. Interestingly, many ABH1 null mice in this study had small or absent eyes, particularly on the right-hand side, a phenotypic similarity to a knockout mouse model of the 2OG oxygenase JMJD6[208].

Subcellular localisation data for ABH1 are also inconsistent, although this may reflect the use of ABH1 constructs tagged with different fusion proteins, which may interfere with subcellular targeting. An ABH1 *N*-terminal EGFP fusion showed distribution of the transfected protein throughout the cell[209], and *N*-terminally Myc-tagged ABH1 is localised in the nuclear euchromatin, the site of localisation of transcribed genes[206]. However, a subsequent study found that an ABH1 *C*-terminal EYFP fusion, and endogenous ABH1, are localised in the mitochondria of HeLa cells[210]. Therefore it is likely that mitochondrial targeting may occur via the ABH1 *N*-terminal region. This study used recombinant ABH1 protein truncated by three amino acids at the *N*-terminus, which was observed to slowly repair 3meC lesions on ssDNA[210], similarly to the previously identified activity of ABH3.

ABH1 has also been observed to have DNA lyase activity on unmodified dsDNA, causing double-strand DNA breaks[211]. Unusually, this activity requires no co-substrates, therefore is postulated to use a novel active site located away from the iron binding DSBH core. Despite the high sequence similarity of the core region of ABH1 to AlkB, ABH1 has additional sequence at both the *N*- (~70 residues) and *C*-termini (~40 residues), therefore one of these regions may facilitate this activity. It is unclear how the formation of damaging dsDNA breaks from undamaged DNA is beneficial to cells, but it may be that this is a weak residual activity, and that the true substrate for the lyase activity is an as yet unidentified damaged base. However, the collection of conflicting data published to date suggests that the true role of ABH1 is still unclear.

1.7.4 ABH8

The structure of ABH8 is unique within the ABH family, with the protein architecture containing an *N*-terminal RNA recognition motif, and a *C*-terminal SAM dependent methyltransferase (MT) domain in addition to the 2OG oxygenase core.

Modified nucleosides in the tRNA anticodon loop can maintain the correct reading frame, or stabilise codon-anticodon pairing[212]. Specifically, modification of uridines in the wobble position of tRNAs can regulate translation of codons enriched in mRNAs encoding DNA-damage response proteins[213]. The function of ABH8 was identified through homology of the ABH8 MT domain to that of *Saccharomyces cerevisiae* Trm9, a MT catalysing the final step in formation of 5-methylcarboxymethyl uridine (mcm^5U) at the wobble position of tRNAs[214].

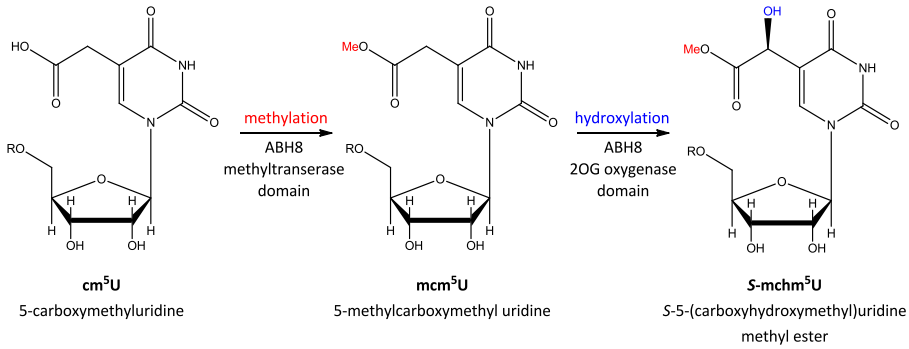


Figure 1-26: Reactions catalysed by ABH8. The methyltransferase domain of ABH8 initially methylates cm^5U to give mcm^5U, which may then be stereospecifically hydroxylated by the 2OG oxygenase domain of ABH8, yielding *S*-$mcmh^5U$.

The MT domain of ABH8 was found to also catalyse the methylation of 5-carboxymethyl uridine (cm^5U) to mcm^5U [215-216] (Figure 1-26). This reaction appears to take place specifically in certain tRNAs containing wobble uridines, namely $tRNA^{Lys(UUU)}$ and $tRNA^{Sec(UGA)}$ [216]. ABH8 interacts *in vivo* with a small accessory protein, TRM112, an interaction which is required for its *in vivo* MT activity[215]. This methylation step is required for downstream production of further modified bases, including mcm^5s^2U and mcm^5Um.

Further to this, the 2OG oxygenase domain of ABH8 hydroxylates mcm^5U at the α position, forming (*S*)-$mchm^5U$ [217] (Figure 1-26), a hypermodified nucleoside previously identified in the silkworm *Bombyx mori*[218]. Further investigations determined that both

the *R*- and *S*- diastereomers of mchm^{5}U are found in mammalian tRNA, but ABH8 is responsible for the stereoselective synthesis of only the *S*- diastereomer[219]. The biosynthetic route leading to formation of the *R*- diastereomer is as yet unknown, and epimerisation appears unlikely, therefore it is postulated that another member of the ABH family may be responsible for the final hydroxylation reaction. The MT and oxygenase domains were found to function independently on individual expression of domain-deleted proteins in ABH8 knockout mice[219]. Full length tagged ABH8 and various truncated mutants are all located in the cytoplasm[209,216], consistent with the observed RNA modification activity.

ABH8$^{-/-}$ mice do not contain mcm^{5}U, mcm^{5}s^{2}U or mcm^{5}Um, but have elevated levels of cm^{5}U [215]. Despite this, the mice have no obvious phenotype to 20 months of age. However, the lack of ABH8 means that tRNA$^{Sec(UGA)}$, important for the recoding of the UGA stop codon to code for selenocysteine (Sec), is not correctly modified. As a result, a slight decrease in recoding was observed as compared to wildtype mice. In contrast, overexpression of ABH8 protein has been observed in many cancer cell lines, with siRNA knockdown in bladder cancer cell lines leading to apoptosis[220], suggesting that tRNA modification may be essential for the survival of cancer cells.

1.7.5 Other ABH family members

Little is currently known about the remainder of the ABH family members (ABHs 4, 5, 6 and 7). In HeLa cells *N*-terminally GFP tagged ABHs 4, 6 and 7 were found to be dispersed throughout the cells, and GFP-ABH5 was present throughout the cell, however some cells contained bright foci in the cytoplasm, although these did not correspond to colocalisation with any specific organelles[209]. However, as observed for ABH1, these localisation results may not correspond to the true subcellular localisations as the large GFP fusion present at the *N*-terminus of the protein may mask any localisation signals present at this terminus of the protein sequence.

The metal binding characteristics of ABH4 have been studied spectroscopically and by EPR[221]. The conserved 2OG oxygenase HXD….H motif was found to bind iron, and

although bioinformatic analysis identified a cluster of cysteine residues in the *N*-terminal domain unique to ABH4, this was found not to bind iron[221].

Uniquely amongst the ABH family, ABH5 is directly upregulated by HIF-1α in hypoxia, via a HRE 5' to the ABH5 gene initiation site[222]. In this study direct detection of ABH5 using an antibody specific to the ABH5 protein showed localisation predominantly in the nucleus, which was depleted on ABH5 siRNA knockdown. Only sparse staining was observed through the cytoplasm, in contrast to previous observations[209]. No 2OG turnover was observed for recombinant ABH5 incubated with methylated oligonucleotides turned over by ABH2 and ABH3[222]. Severe hypoxia induces a DNA damage response and significant ROS damage occurs on reoxygenation, but this damage was not affected by overexpression of ABH5 in cells. This therefore suggests that ABH5 is unlikely to play a role in DNA damage response or repair.

1.8 Fat, mass and obesity associated protein

1.8.1 FTO and obesity

The *Fto* gene was first identified from a mouse mutant arising from a deletion of a 1.6 Mb cluster from chromosome 8[223], with a phenotype characterised by fused toes, loss of left-right asymmetry and premature death (but not elevated weight), which was initially postulated to arise due to dysfunctional programmed cell death[224]. This deletion involves removal of the Iroquis B gene cluster, coding for three homeodomain containing transcription factors[225], and three additional genes termed Fused toes 1, m and o (*Ft1*, *Ftm* and *Fto*)[223]. When the *Fto* gene itself was cloned it was found to have extremely long intron sequences, with a gene of 250 kb coding for a 502 amino acid protein[226], earning the gene the initial name "Fatso".

The human *FTO* gene initially attracted attention when a large scale (> 30,000 candidates) genome-wide association study[227] searching for common single nucleotide polymorphisms (SNPs) in type II diabetes (T2D) patients found that an identified association with a cluster of SNPs in the first intron of the *FTO* gene was mediated by elevated BMI[228]. The representative SNP studied, rs9939609, was found to be either

adenine or thymine. The A allele, present in 39 % of cases, was found to be linked to increased BMI. The SNP was not associated with birth weight, but a difference became apparent in children from 7 years of age. On average, an individual homozygous for the A risk allele is likely to be 3 kg heavier than a T homozygote. This was therefore the first evidence for a specific genetic link to increased body mass, and has since been verified by a number of studies, carried out for a range of ethnicities[229,230]. As a result of this work the gene was renamed as "fat, mass and obesity associated". More recent studies suggest that in some ethnic groups FTO may increase T2D susceptibility independent of the effect on weight[231], possibly suggesting the influence of environmental or other genetic factors. The FTO risk alleles are also associated with structural brain atrophy in the healthy elderly, independent of elevated BMI[232].

In humans, cohort studies suggest that the link to increased weight may be due to increased energy intake, as individuals with the risk allele report impared satiety responsiveness and increased calorific consumption[233-234], particularly of fatty, energy dense foods[235]. In the mouse brain FTO colocalises with the satiety mediator oxytocin, and appears to cause oxytocin upregulation[236].

GWAs can be very powerful as the studies are hypothesis free, thus allowing identification of unexpected pathways and connections[227]. However many common complex disorders, including obesity, have polygenic associations. To date GWAs for obesity have identified associations with variants in the FTO, INSIG2 and MC4R genes, although it appears that the at-risk variants only lead to obesity in an obesity-promoting environment[237]. Therefore direct comparison between studies analysing the effect of variants of a single gene in polygenic disorders, and *in vitro* functional investigations of the proteins involved may be difficult to interpret. To more fully understand the cumulative effect of the identified variants on obesity, combined functional studies would be required.

1.8.2 FTO is a nucleic acid demethylase

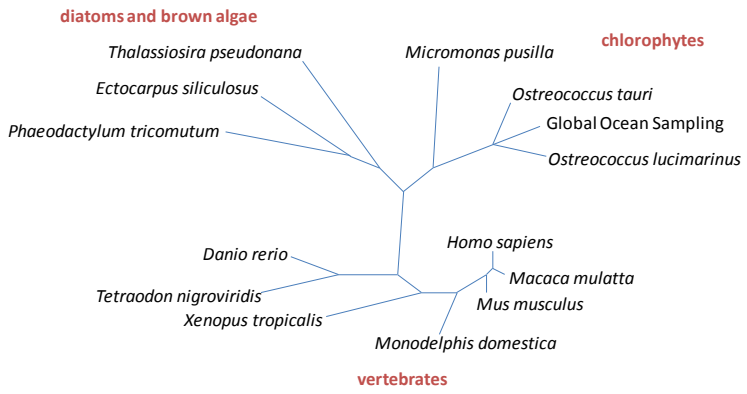


Figure 1-27: Maximum likelihood phylogenetic tree showing the distribution of the FTO gene. FTO is identified in three major clades: vertebrates, diatoms and brown algae, and chlorophytes. Figure adapted from[238].

Studies on recombinant murine Fto (mFTO) demonstrated DNA demethylation, with a preference for 3meT lesions[165]. Further studies on human FTO protein (hFTO) showed a slow turnover of 3meT and 3meU lesions in DNA and RNA respectively[239]. The FTO protein is localised in the cell nucleus[165] and is expressed throughout mice, with highest expression in the hypothalamus, supporting a role in energy homoeostasis[165,240]. The *FTO* gene is predicted to be present in vertebrates, and also in evolutionarily diverse marine eukaryotic algae including diatoms, brown algae and Prasinophyceae, but has not yet been identified in plants, fungi, invertebrates (Figure 1-27)[238].

1.8.3 In vivo *FTO studies*

Several mouse models have been generated to attempt to dissect the role of FTO *in vivo*, and support the findings of the obesity association studies.

Global loss of FTO produced a lean phenotype, albeit alongside a high mortality rate and growth retardation[241], whereas a mouse with a missense mutation in the FTO sequence (I367F) showed a less substantial weight reduction, via reduced fat mass, but without the more detrimental effects of the full knockout (Figure 1-28)[242]. In both cases, in contrast to the human studies, calorific consumption was unaltered between wildtype and mutant mice, with the mass difference attributed to an increased metabolic rate in mice lacking FTO. Mice expressing either one or two additional copies of *FTO* were found to have increased body mass, which was exacerbated when mouse were fed a

high fat diet[243]. Additionally, mice expressing two additional copies of FTO were found to have reduced glucose tolerance, a precursor of T2D.

Figure 1-28: A point mutation in the sequence coding for the FTO protein gives rise to a mouse with reduced body weight, reduced fat mass, increased energy expenditure, and unchanged physical activity. Left: control wildtype mouse (average fat mass 16.64 g; represented by lard); Right: homozygous I367F FTO mouse (average fat mass 11.05 g).

Targeted knockdown of FTO in the arcuate nucleus of rats showed an increase in food consumption, with the reverse effect observed for the corresponding overexpression[244], in contrast to the mass differences observed for whole-body cases. The effect of FTO on energy balance was further investigated by fasting studies, with FTO mRNA expression reduced in the arcuate nucleus of 48 h fasted mice[165], but increased in the hypothalamus of 48 h fasted rats[245]. However, this difference may be due to the increased sensitivity of mice to starvation. Despite these variations in mRNA level, FTO protein levels appear unaffected by feeding or fasting in mouse and rat brains[246]. Conditional knockout of FTO in the neural system resulted in a similar phenotype to that observed for the full body knockout, supporting the importance of FTO in the central nervous system[247].

Limited data for the effects of alterations to FTO in humans are available. A mutation causing replacement of the conserved 2OG binding arginine 316 with glutamine has been found in a large Palestinian Arab consanguineous family, causing postnatal growth retardation and early lethality as observed for knockout mice, but individuals additionally display brain deficits and facial dysmorphism[248]. *In vitro,* 2OG turnover is ablated in both hFTO R316Q[248] and mFTO R313A[165] (the equivalent 2OG binding residue in the mouse sequence) modified proteins. Therefore this interaction appears vital for the function of the FTO protein, and suggests that *in vivo* the R316Q protein may have toxic gain of function or dominant negative effects.

Various other mutations to *FTO* have been identified in humans, with only those either directly binding 2OG or in the substrate recognition lid having an effect on the catalytic activity[249]. Interestingly, mutations were found at a 3-fold higher frequency in the *C*-terminal portion of FTO, compared to the DSBH-containing *N*-terminal region, and individuals with *C*-terminal mutations are more likely to be obese. A patient has been described with partial trisomy of 16q, including the region coding for FTO, showing many symptoms including mental retardation, anatomical facial abnormalities and obesity[250]. However, no elevation of FTO mRNA was observed, making it unclear whether FTO is involved in the patient's obesity.

1.8.4 Cellular studies of FTO

The *FTO* gene shares a CpG island with the adjacent gene *RPGRIP1L* (*Ftm* in mice), which is transcribed in the opposite direction to *FTO*, therefore it is possible that the obesity association may instead be with *RPGRIP1L*/*Ftm*[251]. Ftm is located at the ciliary basal body and has been found to be necessary for establishment of left-right asymmetry, and patterning of the neural tube and limbs[252]. In mice, *Fto* is expressed at a higher level than *Ftm*, with the transcriptional activity of *Fto* and *Ftm* induced by the transcription factor CUTL1, via a regulatory site in the first intron of *Fto*[251]. Different cleavage forms of CUTL1 have different preferences for the SNP alleles present, with the predominant form in cells (p200) having a higher affinity for the risk allele, stimulating its expression. In human cell lines it appears that the obesity associated SNPs in the first intron of *FTO* affect the primary transcript levels of FTO only, with increased copies of *FTO* in the presence of the risk alleles[253]. However, a higher level of FTO CpG methylation has been observed for sequences containing the risk alleles[254].

The mRNA expression of *FTO* is greater in the subcutaneous adipose tissue of obese individuals[255], and appears to correlate with the expression of the transcription factor HIF-1α, suggesting a possible link to hypoxia[256]. *FTO* mRNA expression is also increased in the skeletal muscle of T2D patients, and overexpression of FTO in cultured myotubes increases the rate of lipogenesis in cells, and thus the production of ROS.

Subsequent microarray analysis of genes regulated by FTO overexpression found a reduction for several mitochondrial proteins, leading to a reduction in oxidative metabolism and induction of a shift from lipid oxidation to accumulation of lipid in myotubes[257]. This lipid accumulation may contribute to the obesity phenotype, but it is currently unclear how this effect may caused by a DNA repair enzyme. Additionally, the observed association between FTO variants and poorly controlled satiety responsiveness[233] has yet to be explained functionally.

1.9 Objectives

The work described in this thesis investigates the roles of members of the 2OG oxygenase superfamily in the modification of nucleic acids.

- The unique sequence of FTO, with a novel *C*-terminal domain
- Interaction of FTO with other species in cells
- Initial work towards understanding the roles of ABH1 and ABH7 in cells
- The activity and inhibition of AlkB.

Chapter 2 - FTO *C*-Terminal Domain

2.1 Introduction

2.1.1 Link of FTO to obesity

Obesity is a growing problem for society. Recent figures suggest 1.5 billion people globally are overweight (Body Mass Index (BMI) $\geq$ 25 kg m^{-2}), of which nearly 500 million are obese (BMI $\geq$ 30 kg m^{-2})[258]. Obesity reduces life expectancy[259], and increases the risk of suffering from conditions including Type II Diabetes (T2D)[260]. Obesity and T2D are complex, heritable[261-263] conditions caused by environmental and genetic factors. A genome-wide association study by Frayling et al. identified a strong association between SNPs in the first intron of *FTO* and T2D, mediated by increased BMI[228]. This was the first evidence for a specific genetic link to increased body mass, which has been replicated in several population studies[229-230].

2.1.2 Previous in vitro *studies of FTO*

Figure 2-1: Sequence alignment of FTO sequences from *Mus musculus* (mFTO) and *Homo sapiens* (hFTO). Residues highlighted in red: Fe(II) binding, residues highlighted in blue: 2OG binding, Red cylinders: α-helices, Blue arrows: β-strands. Light blue Roman numerals identify strands forming the DSBH core of the 2OG oxygenase domain. Figure prepared using TCoffee and GeneDoc. Structural assignment from hFTO crystal structure (PDB ID: 3LFM)[264].

Recombinant FTO from *Mus musculus* (mFTO) has been identified as a nucleic acid demethylase, with preference for 3-methyl thymidine (3meT) in DNA[165], and 3-methyl uridine (3meU) in RNA. Beyond this initial result, little was known about the properties of recombinant FTO. Understanding more about the properties of the recombinant

protein may aid studies of the *in vivo* role of FTO. There is a high parity between FTO sequences from *Mus musculus* and *Homo sapiens* (85 % amino acid identity, Figure 2-1), therefore it is likely that the proteins from the two species will carry out similar roles in the cell.

2.1.3 A missense mutation in the C*-terminal domain of mFTO*

In order to understand more about the *in vivo* role of FTO, studies in mice were carried out by a collaborating group at MRC Harwell. A major resource at MRC Harwell is a 1-ethyl-1-nitrosourea (ENU) induced mutagenesis archive[265], which consists of a large library of DNA from mutagenised mice. ENU is an ethylating agent, and a highly potent mutagen[266], most commonly causing A → T transitions[267]. Chemical mutagenesis using ENU produces many functional changes, including hypomorphs or gain-of-function. This complements and expands upon traditional gene inactivation by homologous recombination (knockouts), which usually result in non-functional/null alleles, therefore can be very useful in revealing previously unknown aspects of gene function[268].

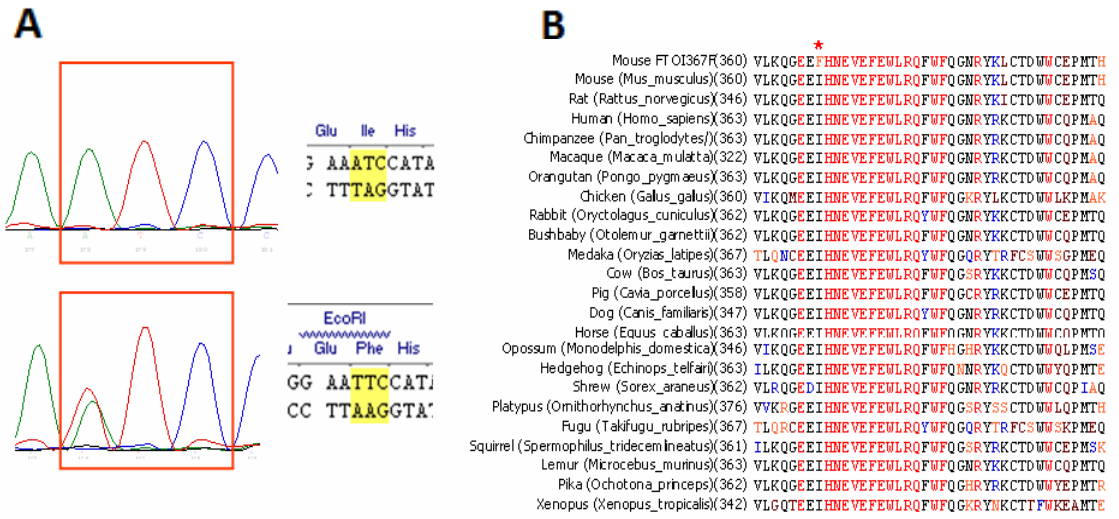


Figure 2-2: An ENU-induced missense mutation I367F in exon 6 of mFTO. (A) Sequencing reveals an A to T mutation in exon 6 of mFTO that leads to conversion of the isoleucine at residue 367 to phenylalanine. (B) Sequence alignment showing isoleucine 367 (asterisk) and surrounding residues are highly conserved in vertebrates. Colour code for amino acids indicates a consensus residue derived from a block of similar residues (black), a completely conserved residue (red), a residue weakly similar to a consensus residue (orange), a consensus residue derived from the occurrence of greater than 50% of a single residue (blue) and non-similar residues (dark-red). Figure prepared by Dr Chris Church (MRC Harwell).

A screen of DNA from 6,624 mice from the Harwell ENU-induced mutagenesis archive identified an adenosine-to-thymidine mutation in exon 6 of murine FTO (mFTO), leading to substitution of a phenylalanine for isoleucine-367 (I367F) in the *C*-terminal region of mFTO (Figure 2-2A). Although this substitution lies outside the double-stranded β-helix (DSBH) and associated elements that form the conserved "catalytic core" of 2OG oxygenases[269], the block of ~20 amino acids surrounding I367 is conserved in FTO throughout vertebrate evolution (Figure 2-2B) suggesting this region may have a physiological role. Mouse lines heterozygous and homozygous for the I367F mutation were developed by MRC Harwell researchers, led by Prof. Roger Cox.

2.1.4 Objectives

This chapter describes studies on recombinant mouse and human FTO aiming to ascertain the function of FTO. A number of new constructs were developed, and proteins were purified and characterised. The *C*-terminus of FTO was studied in more detail, and was found to comprise a novel discrete domain, likely causing dimerisation of the purified protein. Further work was carried out on recombinantly prepared FTO I367 variants, to correlate *in vitro* findings with studies in the mutant mouse model.

2.2 Protein construct production

2.2.1 Full length FTO constructs

Initial cloning of the mouse FTO sequence was carried out by Dr Thomas Gerken. The coding sequence of murine FTO (mFTO) was amplified from I.M.A.G.E. clone #4237261, and cloned into the NheI and EcoRI sites of pET-28a(+), with addition of a *C*-terminal TAG stop codon. This construct expresses a fusion protein facilitating protein purification, with addition of a poly-histidine tag (His-tag) at the *N*-terminus, separated from the protein by a thrombin cleavage peptide sequence (construct referred to as mFTOa). A purification protocol for the mFTOa protein was developed by Dr Thomas Gerken. BL21 (DE3) *E. coli* was used to express protein from this construct, but unfortunately problems were encountered during initial purifications. Attempted cleavage of the His-tag from the protein using thrombin led to a further internal protein

cleavage due to the presence of an additional, internal, thrombin site in the protein sequence.

Despite attempts to optimise the purification of uncleaved mFTO, uncleaved His-tagged protein could only be obtained at a maximum of 80 % purity, as judged by SDS-PAGE. The likely reason for the difficulty in further purification was the presence of *E. coli* chaperone proteins[270]. These may bind to the thrombin cleavage site-containing fusion peptide linking the His-tag to the FTO sequence (Dr Thomas Gerken, unpublished observations). Additionally, as thrombin cleavage of the His-tag while retaining intact FTO had been observed to be unfeasible, this fusion peptide was deleted from the original pET-28a(+) protein construct, producing a new construct with the His-tag linked directly to the protein sequence. This directly-fused construct therefore has a non-cleavable *N*-terminal His-tag, and is the construct which is used throughout the work described here (referred to as mFTO).

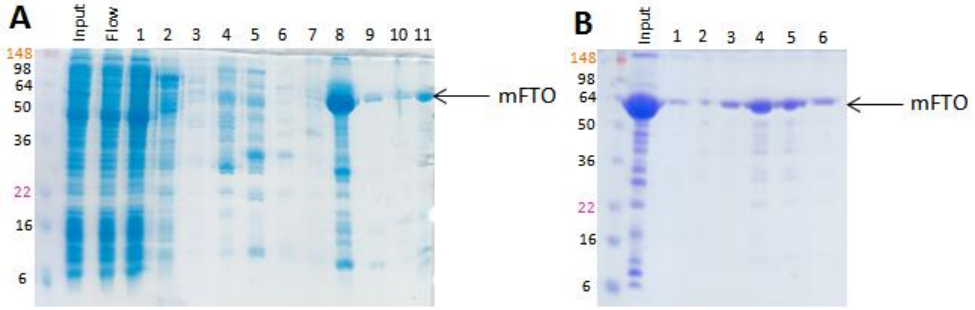


Figure 2-3: SDS-PAGE gels depicting a typical mFTO purification. (A) SDS-PAGE gel showing nickel affinity purification of mFTO. Input: cell lysate applied to column; Flow: protein flowing through column; 1-11: Fractions eluted from column (fractions 8-11 selected). (B) SDS-PAGE gel showing gel filtration purification of mFTO. Input: protein from nickel affinity column applied to column; 1-6: Fractions eluted from column (fractions 3-6 selected).

Optimal expression of soluble mFTO protein was determined (by Dr Thomas Gerken) to occur from *E. coli* grown for 4 hours at 18 °C post induction (addition of 0.5 mM IPTG). Purification of the His-tagged mFTO construct was achieved through nickel affinity chromatography. Proteins were eluted with an imidazole gradient, and SDS-PAGE analysis revealed appropriate fractions to be pooled and concentrated (Figure 2-3A). Further purification by gel filtration chromatography allowed buffer exchange and removal of proteins of different mass from the protein of interest[271]. Further

analysis by UV/vis spectroscopy and SDS PAGE allowed identification of the fractions containing the highest purity protein (Figure 2-3B), which was then concentrated and exchanged into a buffer of 50 mM Tris, 50 mM NaCl, 1 mM DTT pH 7.5 for storage at -80 °C. A typical yield of mFTO was 2.7 mg per litre of bacterial culture, at a purity of ~90 % judged by SDS-PAGE (Figure 2-4).

A summary of all FTO constructs produced for the work described in this chapter, with the expression conditions used, is included in Appendix C.

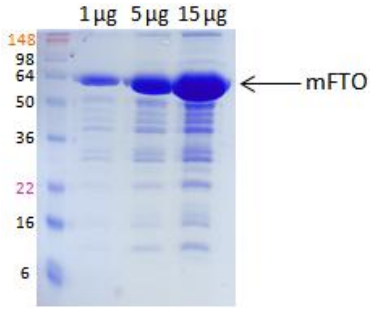


Figure 2-4: SDS-PAGE gel of purified mFTO, quantities as indicated.

In a complementary line of investigation, a construct encoding the human FTO protein was also prepared. At the time of preparation no I.M.A.G.E. clone was available for the sequence of human FTO. Therefore a synthetic gene encoding the human FTO protein, sequence-optimised for expression in *E. coli*, was commercially synthesised by GeneArt. Cloning of this gene into pET-28a(+) via the NdeI and BamHI sites was performed by Dr Celia Webby, with the construct produced referred to as hFTOa. However, similar problems were observed during purification as with the original murine FTO construct. Therefore the corresponding fusion peptide was deleted using site directed mutagenesis (process outlined in Figure 2-5), carried out using the Stratagene Quikchange® site-directed mutagenesis kit. In order to increase the yield of colonies containing the mutated insert the DpnI digestion was allowed to proceed at 37 °C for 2 h rather than the recommended 1 h. Primers used to effect the deletion are detailed in Appendix B. The sequence integrity of the resulting deletion was verified by DNA sequencing, and the produced construct is referred to here as hFTO.

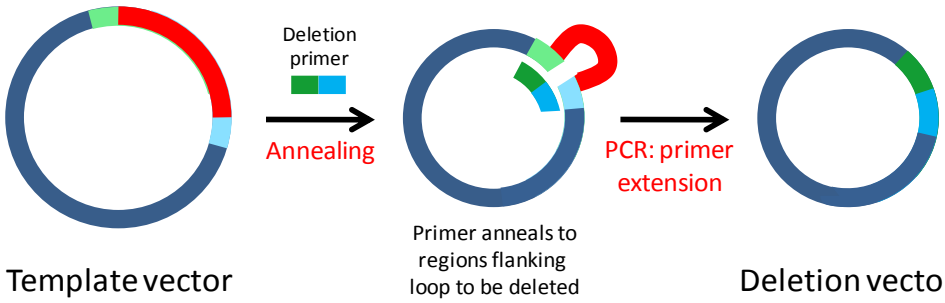


Figure 2-5: Scheme depicting deletion of DNA sequences from plasmid DNA using site-directed mutagenesis. For simplicity, only one DNA strand of the plasmid is shown. Figure adapted from[272].

Small scale bacterial growths carried out under a range of incubation conditions identified the greatest soluble protein expression after 15°C incubation for 16 hours following induction with 0.5 mM IPTG. Recombinant protein was purified by nickel affinity chromatography and gel filtration as for mFTO. Typically 6.9 mg hFTO protein was obtained per litre of bacterial culture, at a purity of ~90 %, judged by SDS-PAGE (Figure 2-6).

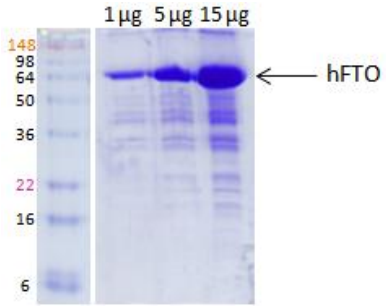


Figure 2-6: SDS-PAGE gel depicting purified hFTO, quantity as indicated.

2.2.2 C-Terminal truncations

The *N*-terminal domain of FTO, comprising the catalytic DSBH domain, has significant homology to the corresponding domains of AlkB and the ABHs. By contrast, the *C*-terminal region of the FTOs is not homologous with that of AlkB or the ABHs. Additionally, BLAST searches yield no homologues for this sequence in non-FTO sequences. Secondary structure predictions for the *C*-terminus suggest this region comprises a discrete helical domain. To investigate the importance of the *C*-terminal region of mFTO, a series of *C*-terminally truncated constructs were produced, to attempt to determine whether this region of the protein is required for activity. Although

not homologous, the *C*-terminus of FIH is necessary for hydroxylation activity, with *C*-terminally truncated constructs unable to oligomerise, and having significantly reduced activity[273]. Substitution of a single residue in the hydrophobic *C*-terminal region of FIH for a hydrophilic residue is sufficient to block dimerisation[273]. However the *C*-terminal region is less important for PHD2, as *C*-terminally truncated PHD2 constructs retain their activity and substrate preference[274].

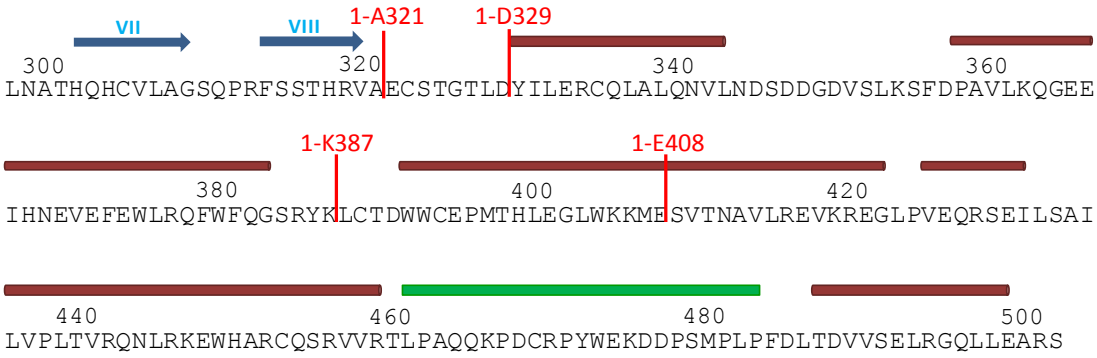


Figure 2-7: Identification of truncation sites in the *C*-terminal sequence of FTO. Secondary structure prediction obtained using PredictProtein[275]. Dark red cylinders: α-helices, Blue arrows: β-strands. Light blue Roman numerals identify strands forming the DSBH core of the 2OG oxygenase domain. Green rectangle: loop; Red markings/text identify sites of *C*-terminal truncations.

At this point no crystal structure of FTO was published, therefore the structure prediction programme PredictProtein[275] was used to identify likely secondary structure in the *C*-terminal domain of FTO, shown in Figure 2-7. This identified similar general features to those observed in the later published hFTO crystal structure[264] (identification of secondary structure present in crystal structure shown in Figure 2-1). From this information, possible sites for truncation of the FTO sequence were identified, as detailed in Figure 2-7 and Table 2-1. Two constructs were designed to remove the whole *C*-terminal domain, cleaving either immediately after the predicted end of the final β-strand comprising the DSBH (1-A321 mFTO), or immediately prior to the first predicted secondary structure in the *C*-terminal region (1-D329 mFTO). Two further constructs (1-K387 mFTO and 1-E408 mFTO) were designed to remove the most *C*-terminal portion of the protein, while retaining the initial helices predicted to be present following the DSBH, as these may be required for correct protein folding.

Truncations at these positions through the *C*-terminal domain were attempted, to create mFTO 1-E408, mFTO 1-K387, mFTO 1-D329 and mFTO 1-A321. These constructs were cloned from the wildtype pET-28a(+) mFTO template by mutation of the codon following the new *C*-terminal residue to a stop codon, by site directed mutagenesis, carried out following the Agilent Technologies QuikChange Site-Directed Mutagenesis protocol (Figure 2-8), again employing a 2 h digestion with DpnI.

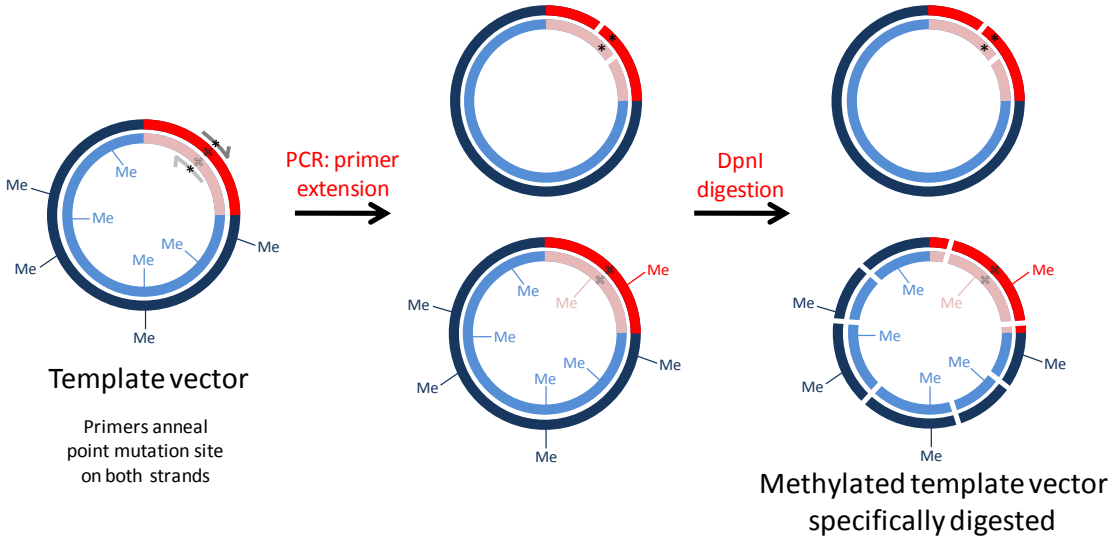


Figure 2-8: Scheme outlining steps in site-directed mutagenesis. Figure adapted from[272].

Although wildtype mFTO could be efficiently produced in a soluble form, expression trials showed poor expression levels for *C*-terminally truncated forms of wildtype mFTO, with only low amounts of largely insoluble protein (summarised in Table 2-1) produced in the presence of other highly expressed species, most likely protein chaperones[270].

C-terminal truncation	Mass /kDa	Explanation	Insoluble protein expression	Soluble protein expression
mFTO 1-E408	48144.2	Middle of second major helix in *C*-terminus	Low	Very low
mFTO 1-K387	45525.2	Following first major helix in *C*-terminus	Low	Low
mFTO 1-D329	38637.5	Likely break between two mFTO domains	Low	Very low
mFTO 1-A321	37830.7	Predicted end of DSBH core	Low	None

Table 2-1: Summary of mFTO *C*-terminal truncation constructs and their expression levels.

As a further expression test for mFTO 1-D329, a small-scale purification was attempted using HisBind™ resin (Novagen). Cells expressing the construct were lysed as in a typical protein purification protocol, then the lysate was mixed with precharged resin. Following incubation to allow His-tagged proteins to chelate the Ni(II) present the resin was washed and mixed with SDS-PAGE loading buffer, eluting protein that remained specifically bound to the beads. Insoluble, soluble and bound protein fractions were analysed by SDS-PAGE (Figure 2-9). A control protein (CmFTO, see Section 2.2.3) bound preferentially to the resin, observed as a significant band at ~ 20 kDa in the eluted protein lane. However, no bands were observed in the eluted protein fraction from cells expressing mFTO 1-D329 (~39 kDa), suggesting that, as no binding to the resin had occurred, soluble His-tagged protein had not been significantly expressed. As had been observed during attempted expression of each of the *C*-terminally truncated proteins, significant bands were observed at 40-45 kDa in the insoluble protein fractions, most likely being due to highly expressed chaperone proteins, such as DnaJ (which has a molecular mass of ~41 kDa)[270,276], labelled X in Figure 2-9.

The difficulties encountered here in expressing the truncated proteins may suggest that the *C*-terminal domain is important for the correct folding of the expressed protein, and therefore potentially important for the effective activity of FTO.

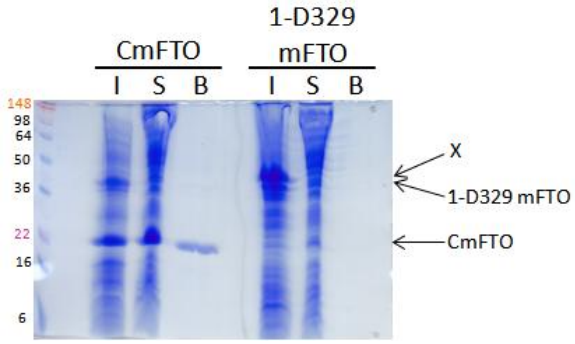


Figure 2-9: SDS-PAGE gel showing expression and His binding test for 1-D329 mFTO and CmFTO. I: Insoluble protein; S: Soluble protein; B: Protein bound to HisBind resin; X: Highly expressed protein (chaperone).

2.2.3 C-Terminal domain

Due to difficulties encountered with *C*-terminal truncations, an alternative approach was employed, producing the *C*-terminal portion separately to allow studies on this discrete

domain. Residues 1-328 from the mFTO construct (with non-cleavable His-tag) were deleted by site-directed mutagenesis, to create the construct D329-S502 mFTO, referred to throughout as CmFTO. Expression trials identified 16 hour growth of BL21 (DE3) *E. coli* at 15°C following induction with 0.5 mM IPTG as optimal for soluble CmFTO expression (Figure 2-10).

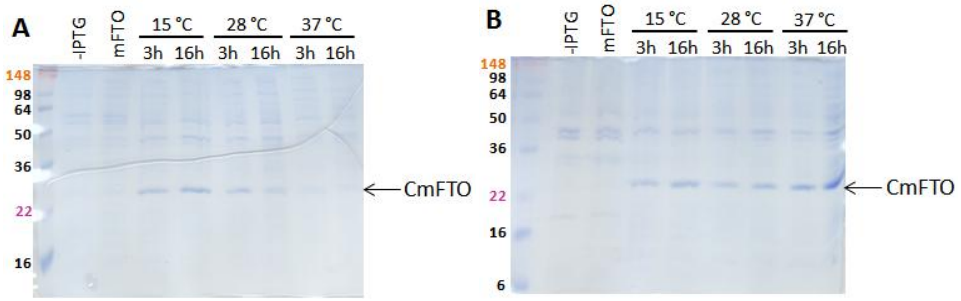


Figure 2-10: Expression trials of CmFTO in BL21 DE3 *E. coli*. (A) Soluble protein, (B) Insoluble protein. Expression induced with 0.5 mM IPTG, then cells grown under the temperature and time conditions indicated.

Recombinant protein was purified by nickel affinity chromatography and gel filtration (Figure 2-11). Approximately 9.44 mg CmFTO at >95% purity, judged by SDS-PAGE (Figure 2-12A), was obtained per litre of culture.

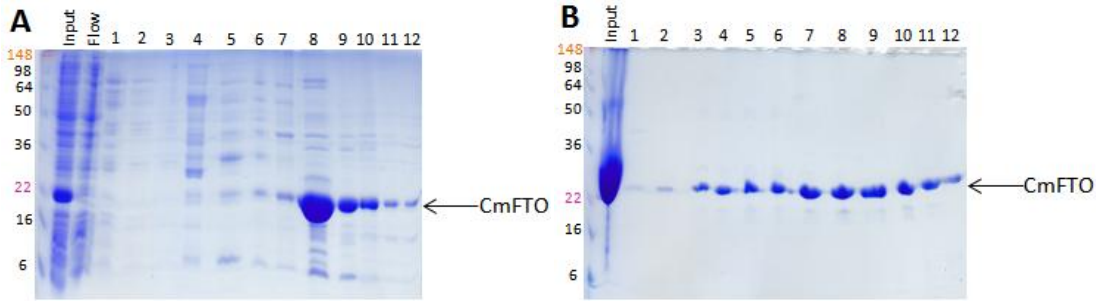


Figure 2-11: (A) SDS-PAGE gel showing Nickel affinity purification of CmFTO. Input: cell lysate applied to column; Flow: protein flowing through column; 1-12: Fractions eluted from column (fractions 8-10 selected). (B) SDS-PAGE gel showing gel filtration purification of CmFTO. Input: protein from nickel affinity column applied to column; 1-12: Fractions eluted from column (fractions 3-12 selected).

The corresponding *C*-terminal domain of human FTO was prepared by the same method. Expression of ChFTO was optimal under conditions of growth for 3 hours at 15 °C following induction with 0.2 mM IPTG, with an approximate yield of 0.81 mg protein per litre of bacterial culture, at a purity of ~90 % (Figure 2-12B).

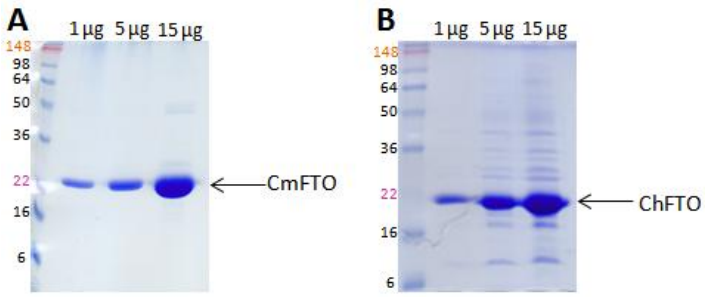


Figure 2-12: SDS-PAGE gels of purified (A) CmFTO and (B) ChFTO, quantities as indicated.

2.2.4 I367 point mutants

Given the value of the development of modified FTO mouse models by collaborators at MRC Harwell, *in vitro* analyses of the mouse I367F FTO sequence was a priority. Therefore the corresponding point mutation was made in both full length and *C*-terminal mFTO by site directed mutagenesis. Unfortunately, expression of both I367F mFTO and I367F CmFTO was poor and largely insoluble, meaning that effective purification of the proteins was not possible using the methods previously used for wildtype FTO constructs. This poor soluble expression suggests that the variant protein may not fold correctly. Therefore this residue, despite being away from the DSBH core of the 2OG oxygenase, may play an important role in the folding of FTO. Although both isoleucine and phenylalanine are hydrophobic residues, the significant difference in size (Figure 2-13), and the aromaticity of the benzyl ring present in phenylalanine may mean that the difference is too significant for maintenance of the 3D structure of the domain.

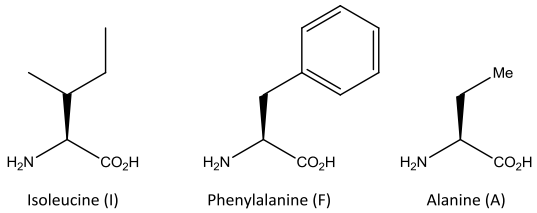


Figure 2-13: Comparison of the side chains of isoleucine (I), phenylalanine (F) and alanine (A).

Instead, the decision was taken to attempt replacement of the I367 residue with alanine, a substitution commonly used to ablate metal binding, and prevent activity[165,173,217]. It was hoped that replacement of isoleucine with the smaller alanine (Figure 2-13) would allow normal folding to occur, allowing soluble protein expression. However, any

specific features of the wildtype isoleucine residue would hopefully be removed, through substitution with the more innocuous alanine, allowing understanding of the role of the isoleucine residue in this protein domain.

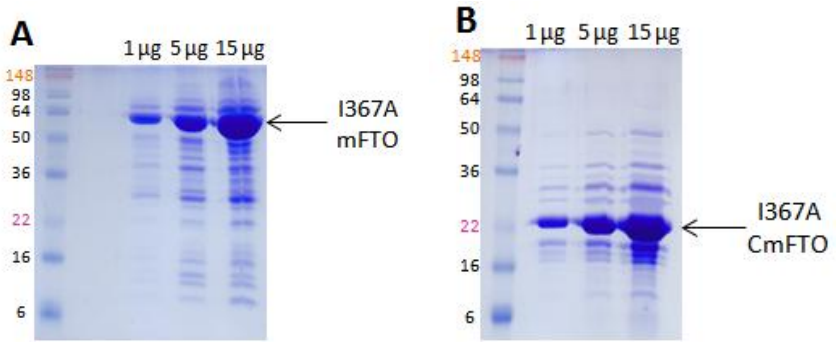


Figure 2-14: SDS-PAGE gels of purified (A) mFTO I367A and (B) CmFTO I367A, quantities as indicated.

The I367A mutation was again formed by site directed mutagenesis, in both mFTO and CmFTO. These constructs yielded soluble I367A mFTO and I367A CmFTO proteins. Expression was found to be optimal when cells were grown at 15°C for 16 hours (I367A mFTO) or 4 hours (I367A CmFTO), after induction of protein expression with 0.2 mM IPTG. Proteins were purified following the protocol previously described for mFTO, with 0.56 mg I367A mFTO (Figure 2-14A) and 0.63 mg I367A CmFTO (Figure 2-14B) produced per litre of bacterial culture. Purity of these proteins was slightly decreased compared to the corresponding wildtype sequences, possibly due to the reduced expression of the mutant proteins, or changes produced in the protein's structures due to the presence of the variant residue.

2.2.5 Alternate mouse amino acid sequences

When initially sourcing the *M. musculus* FTO gene for construct production it was observed that two different sequences exist. The nucleotide sequence of the commercially available I.M.A.G.E. clone differs in several positions from the sequence published by the National Center for Biotechnology Information (NCBI). This produces three coding differences (referred to as X, Y, Z) in the resultant amino acid sequence (Figure 2-15, Table 2-2), with each of the mutations falling within the *C*-terminal portion of the protein sequence. These differences are due to population variations

between multiple inbred mouse strains[277-278], with the I.M.A.G.E. sequence originating from one mouse strain (FVB/Balbc/DBA) and the published online NCBI sequence from an alternate strain (C57BL/6J). Mouse studies at MRC Harwell were carried out using C57BL/6J x C3H/HrH crosses.

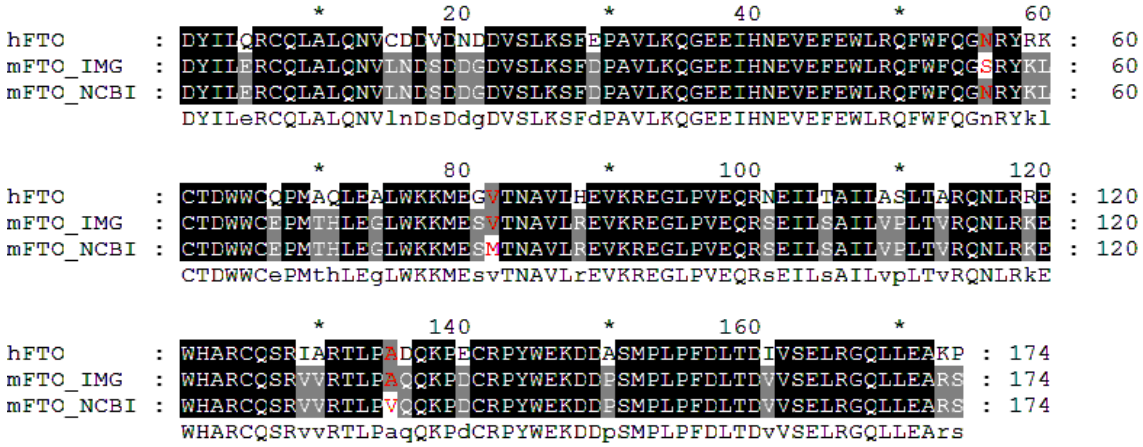


Figure 2-15: Sequence alignment between sequences of human and mouse CFTO, with residues differing between the I.M.A.G.E. (mFTO_IMG) and NCBI (mFTO_NCBI) sequences highlighted in red (prepared using TCoffee and GeneDoc). Residue 1 in the figure corresponds to residue 332 in hFTO and 329 in mFTO.

Source	X (383)	Y (409)	Z (462)
H. sapiens	Asn	Val	Ala
M. musculus I.M.A.G.E.	Ser	Val	Ala
M. musculus NCBI	Asn	Met	Val
P. trogloytes	Asn	Val	Ala
C. lupus familiaris	Asn	Val	Val
B. taurus	Ser	Leu	Ala
R. norvegicus	Asn	Val	Ala

Table 2-2: Comparison of residues present in different FTO sequence constructs. Numbering corresponds to residue position in mouse sequence.

Alignment with known FTO sequences from other species showed that residues at the three variable positions are not highly conserved (Figure 2-15, Table 2-2). At two of the variable sites (Y and Z) the residue present in the mouse I.M.A.G.E. sequence is the same as in the human FTO sequence, whereas at position X the mouse NCBI sequence is identical to the human sequence. At positions Y and Z most sequences listed (including the mouse I.M.A.G.E. sequence) have small hydrophobic residues (alanine, valine and leucine), however at position Y the mouse NCBI sequence has a rather

different methionine residue. Although at position X the I.M.A.G.E. mFTO sequence is different from the majority of the other FTO sequences, the serine found here is also present in the *B. taurus* sequence.

Due to the overall greater similarity of the I.M.A.G.E. mFTO sequence to the other published sequences, particularly to the *H. sapiens* sequence, the I.M.A.G.E. clone has been used throughout the work described here, unless specifically mentioned otherwise. However, to understand more about the role of the different residues in the *C*-terminal domain, and whether the differences between sequences from different species affects the properties of the proteins, the appropriate mutations (S383N, V409M and A462V) were introduced into the previously prepared I.M.A.G.E. mFTO construct, to produce the NCBI mFTO sequence. This was carried out using the Quikchange Multi site-directed mutagenesis kit (Agilent Technologies), as depicted in Figure 2-16.

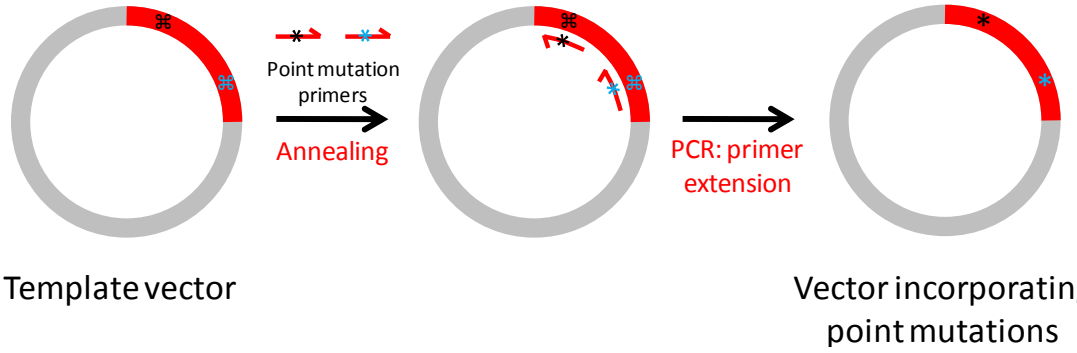


Figure 2-16: Scheme depicting simultaneous incorporation of two point mutations in plasmid DNA using Multi site-directed mutagenesis. Figure adapted from[279].

Following successful cloning of the NCBI constructs, through introduction of the three point mutations, expression trials were attempted for each of the proteins: mFTO NCBI, CmFTO NCBI, I367A mFTO NCBI and I367A CmFTO NCBI.

The only protein for which obvious expression was observed was CmFTO NCBI (Figure 2-17), with optimal soluble protein expression in *E. coli* grown at 15 °C for 16 h following induction with 0.5 mM IPTG.

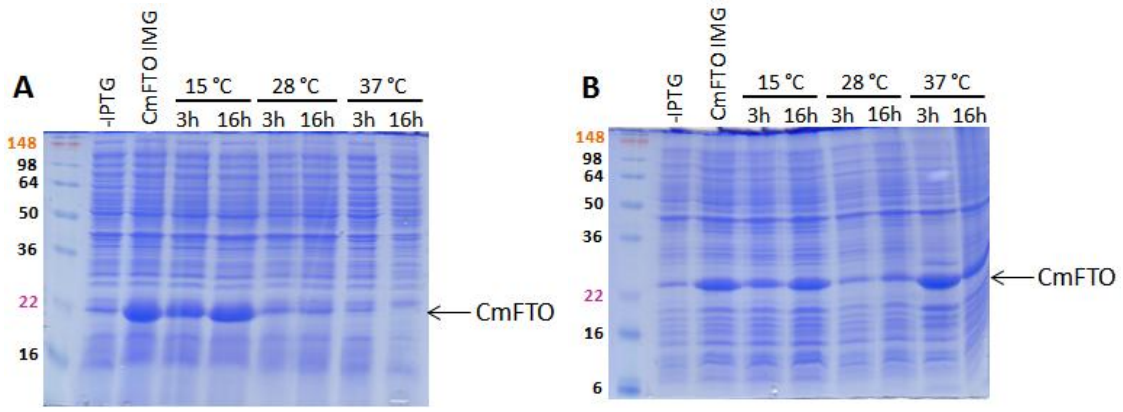


Figure 2-17: Expression trials for CmFTO NCBI in BL21 (DE3) *E. coli*. (A) Soluble protein; (B) Insoluble protein. Induction of expression with 0.5 mM IPTG, then cells grown under the time and temperature conditions indicated. Positive control for comparison of expression: CmFTO I.M.A.G.E. (CmFTO IMG) grown for 16 h at 15 °C, after induction with 0.5 mM IPTG.

Purification of CmFTO NCBI by affinity and size exclusion chromatography was successful, although at much lower yield (3.7 mg/litre culture), and at lower purity (~80 % by SDS-PAGE analysis, Figure 2-18) than for the corresponding I.M.A.G.E. sequence (Figure 2-12).

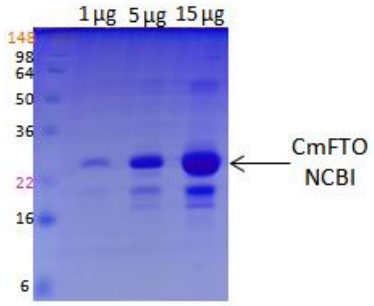


Figure 2-18: SDS-PAGE gel of purified CmFTO NCBI, quantities as indicated.

Due to the low levels of protein expressed, and the difficulty in determining whether FTO constructs were present amongst the many proteins in the *E. coli* cell lysate, western blots were attempted to investigate whether the NCBI constructs were expressed, and which conditions were optimal for growth. Bands at the expected protein masses were observed for each of the proteins expressed (shown for I367A mFTO NCBI and I367A CmFTO NCBI in Figure 2-19) with optimal, although still low, expression conditions identified in Table 2-3. Bands at reduced molecular mass are likely due to protein degradation.

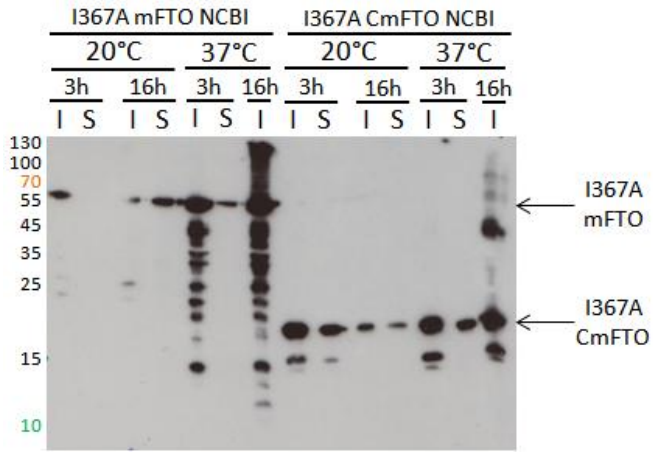


Figure 2-19: Western blot detecting expression of His-tagged FTO NCBI constructs in *E. coli* cell lysate. Expression conditions indicated above lanes. 5 µl loaded (0.05 % total culture). Detected with mouse α-His (1 in 1000) and α-mouse-HRP (1 in 1000) antibodies. Exposure 1 min.

Expression and purification of the further NCBI constructs was therefore attempted based on the very low levels of expression seen by western blot, but successful purification of sufficiently pure protein was not possible. Due to the significantly elevated expression of the mFTO I.M.A.G.E. sequences compared to the NCBI constructs, and the sequence-based arguments presented previously, further work on the NCBI sequences was not pursued.

Construct	IPTG /mM	Time /h	Temperature /°C
mFTO NCBI	0.5	3	15
I367A mFTO NCBI	0.5	16	20
CmFTO NCBI	0.5	16	15
I367A CmFTO NCBI	0.5	3	20

Table 2-3: Summary of mFTO NCBI protein expression conditions.

2.3 Dimerisation state

During initial purification of the mFTO and CmFTO proteins it was observed that protein elution from the gel filtration column occurred over a broad volume range. As this method causes separation due to molecular mass, this suggested the presence of species of different molecular masses, possibly due to the existence of more than one oligomerisation state. A number of techniques were therefore employed to investigate this possibility.

2.3.1 Native PAGE

In denaturing SDS-PAGE the electrophoretic mobility of a protein is governed primarily by its molecular mass. However for PAGE run in a non-denaturing, or "native", environment the mobility is dependent on its hydrodynamic size, and the protein's intrinsic charge. Therefore any protein-protein interactions occurring in the native protein state, for example oligomerisation, may be investigated[280].

To give an indication of molecular mass, a standard of analytical grade BSA was run alongside samples. This runs as multiple bands under non-denaturing conditions, forming a molecular mass scale of 67 kDa (monomer), 133 kDa (dimer) and 200 kDa (trimer)[281]. As the 200 kDa band can be weak, a further molecular weight standard of β-amylase (from *Ipomoea batatas*, sweet potato) was run (Figure 2-20B), which runs with a molecular mass of approximately 206 kDa[282].

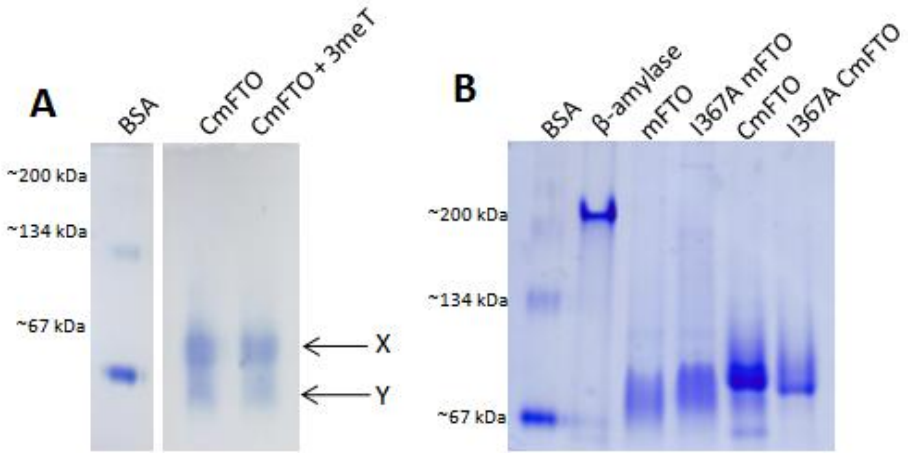


Figure 2-20: Native PAGE gels. (A) 4 µg proteins. CmFTO + 3meT with 8 µg 3meT oligonucleotide. (B) 5 µg BSA and β-amylase, 4 µg FTO. Molecular weight guides based on protein standards.

Initial native PAGE (Figure 2-20A) of CmFTO suggested that the protein exists as more than one species in the native state. The species with the higher mobility (marked Y) may be a monomer, and the species with lower mobility (marked X) a dimer. Comparison to the BSA standard suggests these species have approximate masses of ~50 and ~80 kDa. However, as previously discussed, the migration distance of species in native gels is not always easily relatable to the molecular mass of the protein, therefore the motion of CmFTO through the gel may be retarded relative to that of BSA.

Addition of 3meT 18mer oligonucleotide to CmFTO was attempted to identify whether the proportion of the two species present was altered in the presence of an identified FTO substrate. However there did not appear to be a resolvable difference between the protein oligomerisation states in this case (Figure 2-20A, Lane 2 c.f. Lane 3).

Comparison of the different mFTO constructs prepared (Figure 2-20B) gives unclear results. The full length mFTO appears to migrate more rapidly through the gel than the *C*-terminal constructs. There also appears to be little difference between the migration pattern of wildtype and I367A proteins. A more sensitive and quantitative technique is therefore required for investigation of possible oligomerisation.

2.3.2 Western blot

For use in mouse studies an α-FTO antibody was commercially prepared by Eurogentec, overseen by Dr Chris Church (MRC Harwell). The polyclonal antibody was raised in rabbits to recombinant (His-tagged) mFTO, then affinity purified using further recombinant mFTO. At MRC Harwell this antibody was used primarily for western blot detection of full length murine FTO in cell lysates. However, to determine the range and extent of recognition of the antibody, a western blot was carried out using a range of murine and human recombinantly prepared proteins. Following optimisation of antibody concentrations used to probe the membrane, the western blot in Figure 2-21 shows that each of the constructs are recognised by the custom antibody. Therefore this antibody is able to recognise not only the recombinant murine FTO, but also the closely related human protein, and the FTO *C*-terminal domain of both of these proteins. Detection of I367A mFTO mutants is also successful.

However, multiple FTO bands are observed in this western blot (which are also observed when probing with an anti-His antibody – data not shown), despite Coomassie staining of SDS-PAGE gels of purified proteins revealing essentially only single bands (e.g. Figure 2-12A), or a few less intense additional bands at lower mass (e.g. Figure 2-6). This more sensitive technique has therefore been able to detect the much lower levels of protein present across the molecular mass range. The multiple bands at

molecular masses lower than that of the full length purified proteins are likely due to protein degradation, for example by contaminating proteases. However, bands are also observed at increased mass compared to that of the wildtype protein, for example at ~200 kDa for full length FTO constructs, and ~50 kDa for *C*-terminal FTO constructs, although not in the case of ChFTO. These bands at increased mass are present to similar extents in each of the lanes, and for both wildtype and I367A sequences. It is likely that these bands at increased mass may be due to self-interaction or oligomerisation of the proteins.

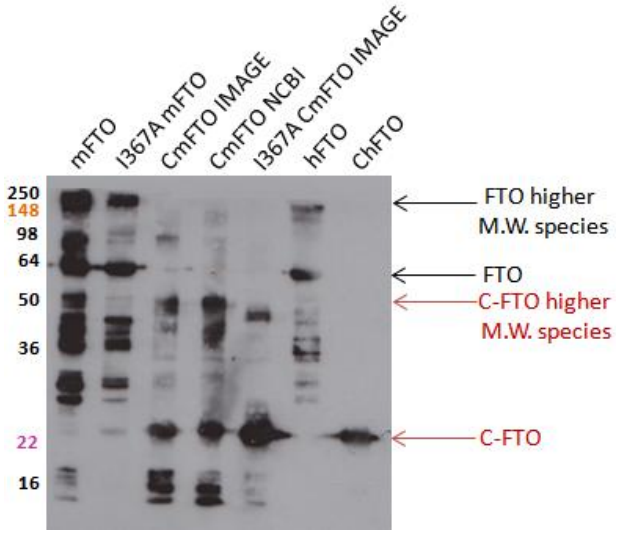


Figure 2-21: Western blot detection of recombinant FTO proteins expressed in *E. coli*. 0.5 µg protein loaded per lane. Detection using rabbit polyclonal anti-FTO (1 in 2000); anti rabbit HRP conjugate (1 in 1000). 1 minute exposure.

Formation of multimers is usually prevented in the reducing conditions employed for SDS-PAGE, (which involve addition of the detergent SDS, and the reducing agent β-mercaptoethanol, and denaturation by heating to 100 °C), although very strong interactions could result in the appearance of bands at higher mass, and protein-protein disulfide bonds may not be fully ablated by the addition of β-mercaptoethanol. As western blotting is a sensitive technique it is able to detect the low levels of higher FTO mass species present.

2.3.3 Analytical gel filtration

Analytical gel filtration is a chromatography technique that enables separation of molecules on the basis of size, and is based on the same technique used previously for

protein purification[271]. The use of a column with a reduced volume (30 ml) and a high specification sepharose resin allows analytical measurements to be made. This therefore allows analysis of the molecular masses of the proteins studied, and so the potential oligomerisation state in solution. Different sepharose matrices with different extents of crosslinking may be used for the study of proteins of different molecular masses. For these studies columns containing the GE Healthcare Superdex™ S75 (optimum separation range of 3-70 kDa, used to study C-FTO constructs) and S200 (separation range of 10-600 kDa, used for full length FTO constructs) media were used[283].

Elution of protein from the column is typically monitored as a chromatogram showing absorbance at 280 nm (A_{280}), which allows determination of the elution volume corresponding to peak maxima for different species. From this an estimation of the protein's molecular weight may be obtained through comparison to calibration curves prepared for each separation media, which are created by analysis of protein standards. Protein standards selected from across the molecular mass ranges to be studied (Chapter 7, Table 8-7) were mixed and run on columns to create calibration curves (Figure 2-22) based on the elution data obtained for the S75 and S200 columns.

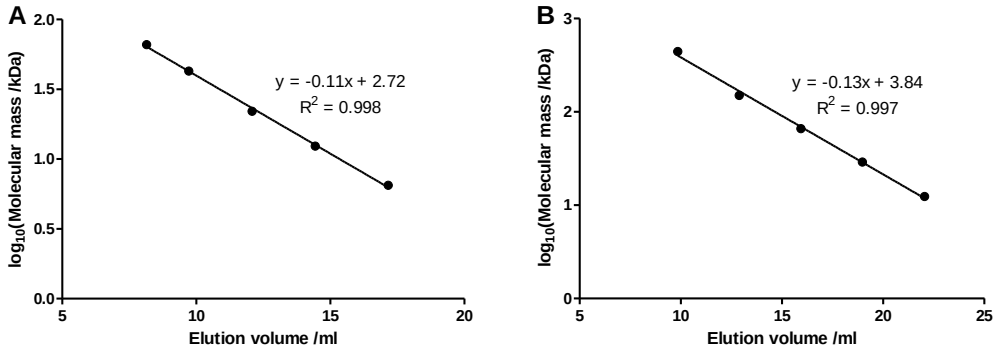


Figure 2-22: Calibration curve of molecular mass vs. elution volume for molecular mass protein standards eluted from an (A) S75 30 ml, (B) S200 30 ml analytical gel filtration column.

Analytical gel filtration analysis was carried out for FTO constructs, and confirmed the previous observations from protein purification that multiple peaks are present for wildtype proteins (Figure 2-23). The two resolved peaks visible on the observed elution profile suggest that CmFTO exists as two distinct species in solution, possibly through oligomerisation (Figure 2-23B). To verify that these two species are in fact both

CmFTO, and not due to contaminating species of different molecular masses, samples of eluted protein were precipitated using a trichloroacetic acid protocol and run on an SDS PAGE gel (Figure 2-24), which showed the presence of CmFTO protein in both peaks.

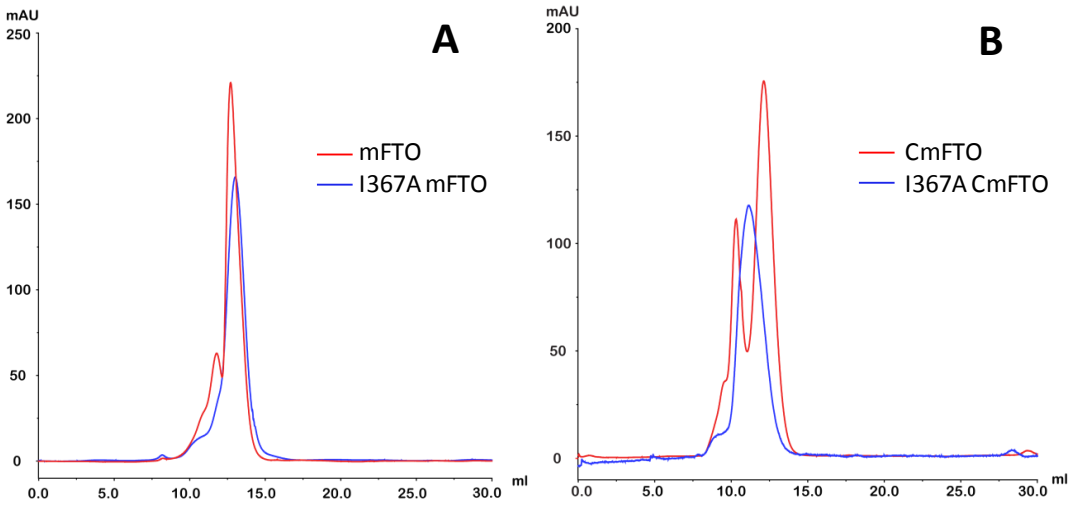


Figure 2-23: Analytical gel filtration chromatograms showing the presence of two FTO species in solution. (A) S200 30 ml: Full length mFTO (red) and I367A mFTO (blue), (B) S75 30 ml: CmFTO (red) and I367A CmFTO (blue).

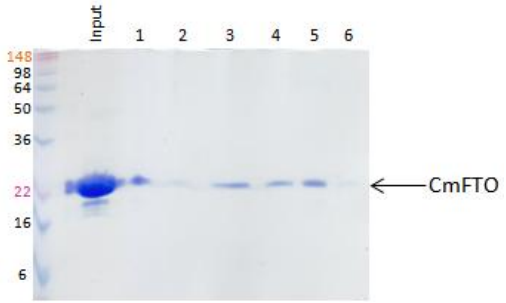


Figure 2-24: SDS-PAGE gel showing protein content of fractions from analytical gel filtration. Input: Protein loaded onto column; 1-6: Fractions of eluted protein.

Use of the previously derived calibration curves (Figure 2-22) allowed calculation of the molecular masses of the protein species corresponding to each of the observed peaks, with data shown in Table 2-4. The calculated values for molecular masses of the CmFTO species appear to correlate well with the known molecular mass of the CmFTO monomer. However the calculated mass is not particularly accurate for the full length mFTO constructs, which were run on the S200 column, with higher masses observed than were expected. This may be because the full length FTO runs more rapidly through the column than expected. Comparison of wildtype proteins to the I367A variants

revealed that whereas wildtype mFTO and CmFTO exist in both monomeric and dimeric forms, both I367A variants exist, at least predominantly, as monomers. These results therefore suggest that I367 is involved in dimerisation of mFTO, and alteration of this residue to alanine ablates dimerisation. Additionally, as CmFTO dimerises, and the I367A mutants are unable to dimerise, it can be concluded that the dimerisation of FTO appears to be partially, if not wholly, mediated by the *C*-terminal domain.

		Shoulder		**Peak 1**		**Peak 2**	
		Molecular mass /kDa	*Possible identity*	*Molecular mass /kDa*	*Possible identity*	*Molecular mass /kDa*	*Possible identity*
S75 30 ml	CmFTO	47.49	impurity	36.62	dimer	22.92	monomer
	I367A CmFTO	51.31	impurity	-	-	29.50	monomer
	ChFTO	64.19	impurity	40.39	dimer	24.47	monomer
S200 30 ml	mFTO	304.42	impurity	229.40	tetramer	175.37	trimer
	I367A mFTO	331.97	impurity	-	-	160.82	trimer
	hFTO	316.88	impurity	251.71	tetramer	199.94	trimer

Table 2-4: Summary of analytical gel filtration data for FTO proteins.

As explained in Section 2.2.5, two different mouse FTO sequences are published (I.M.A.G.E. and NCBI), with three amino acid differences. As the alteration in amino acid sequence at the single I367 position appears to be sufficient to prevent, or at least significantly reduce dimerisation in the I.M.A.G.E. sequences, the impact of these three amino acid differences in the NCBI construct was investigated. CmFTO NCBI analytical gel filtration shows multiple species as for the I.M.A.G.E. sequence (Figure 2-25), although with a slightly different mass profile (Table 2-5). Therefore it is likely that these three residue differences do not affect the oligomerisation status of FTO.

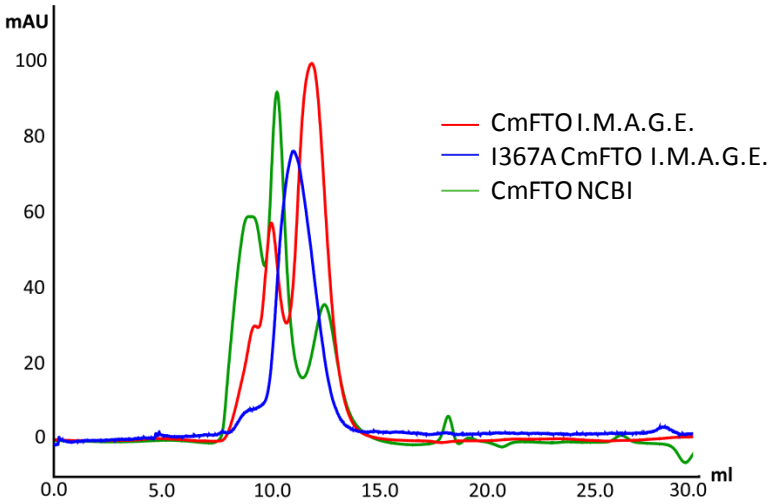


Figure 2-25: Analytical gel filtration chromatogram from S75 30 ml column showing NCBI vs I.M.A.G.E. CmFTO sequences CmFTO NCBI (green), CmFTO I.M.A.G.E. (red), CmFTO I.M.A.G.E. I367A (blue).

Protein	Molecular mass /kDa		
	Peak 1	**Peak 2**	**Peak 3**
CmFTO NCBI	52.11	38.56	22.40
CmFTO IMAGE	47.49	36.62	22.92
I367A CmFTO IMAGE	51.31	-	29.50

Table 2-5: Summary of analytical gel filtration data for CmFTO I.M.A.G.E. and NCBI proteins.

Analytical gel filtration was also attempted for hFTO constructs, with data included in Table 2-4. However these species did not run well on the columns, with unclear elution profiles containing multiple peaks at higher mass (Figure 2-26), complicating peak assignment.

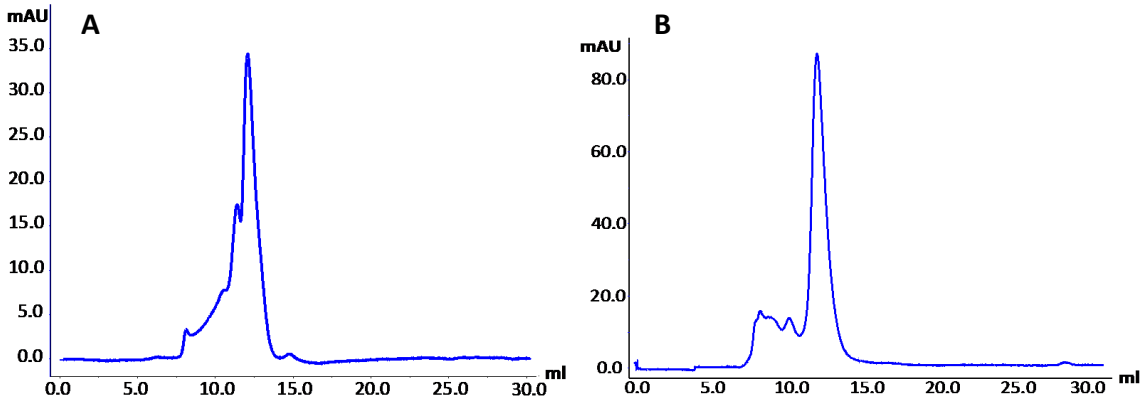


Figure 2-26: Analytical gel filtration chromatograms for (A) hFTO (S200 30 ml) and (B) ChFTO (S75 30 ml).

2.3.4 Mass Spectrometry

Soft ionisation nano-electrospray ionisation (ESI) mass spectrometry (MS) allows investigation of binding of substrates and inhibitors to proteins in a desolvated state, or investigation of protein oligomerisation. Proteins are sprayed into the spectrometer from a neutral, non denaturing buffer, typically ammonium acetate. Here they are protonated to different extents creating a range of multiple positively charged ions, so peaks for a given species occur at several m/z values in a spectrum. To combine these data the spectra may be transformed, allowing calculation of molecular masses.

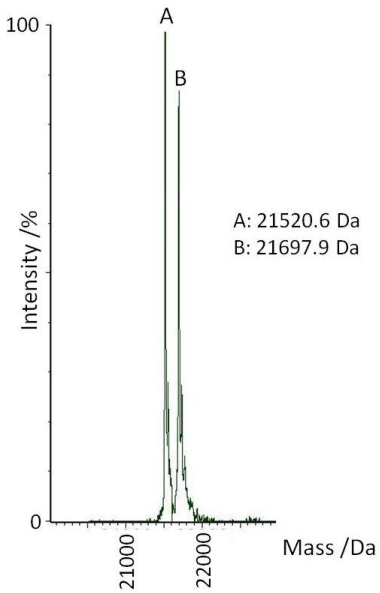


Figure 2-27: Soft Ionisation Mass Spectrometry of CmFTO. A: Unmodified CmFTO; B: +178 Da, CmFTO + *N*-gluconoyl. Cone voltage 80 V.

To further characterise the CmFTO protein construct, non-denaturing ESI-MS of CmFTO was carried out with Dr. Jasmin Mecinović (Figure 2-27). The calculated mass of CmFTO (excluding initiating methionine residue) is 21522.3 Da, which is present as peak "A" in Figure 2-27. However the transformed mass spectrum shows two major peaks corresponding to two protein species, one at a mass of +178 Da relative to the expected peak (labelled "B"). This additional peak exists even at high cone voltage (200 V, data not shown), therefore is probably a covalent modification.

It is likely that this species corresponds to phosphoglycosylation (which would give a peak at +258 Da), then subsequent dephosphorylation of this group (+178 Da), a

well-documented modification of *N*-terminally His-tagged proteins[284-285]. This protein modification has also been previously observed for proteins purified in this laboratory, including synthetic ankyrin repeat domains[286].

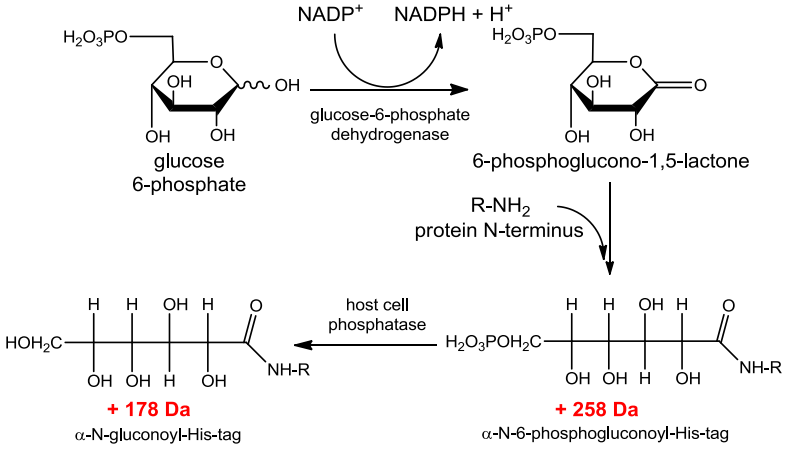


Figure 2-28: Outline of chemical route to *N*-terminal His-tag modification. Adapted from[284].

Unfortunately this glycosylation adds unwanted complication to mass spectra and can only be removed by cleavage of the His-tag, which was not possible in this construct. An attempt was made to produce a construct containing a thrombin cleavable His-tag which it was hoped would be less susceptible to thrombin cleavage, but unfortunately fragmented protein was obtained, as previously for the full length mFTOa.

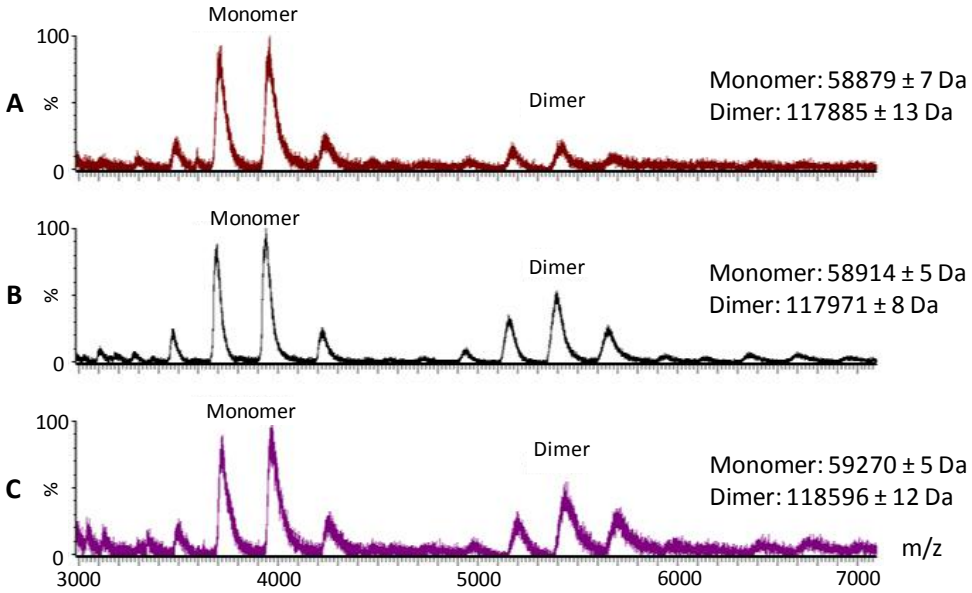


Figure 2-29: Non-denaturing electrospray ionisation mass spectrometry analyses carried out on a Waters Synapt™ HDMS™, showing the oligomeric composition of (A) I367A mFTO, (B) mFTO and (C) hFTO. Predominant peaks for the monomer correspond to +17, +16, +15 and +15 charge states; predominant peaks for the dimer correspond to +24, +23, +22 and +21 charge states.

The soft ionisation technique may also be used to study the oligomerisation state of proteins, by identifying different species present in mass spectra. This work was again carried out with Dr. Jasmin Mecinović. The spectrum of full length mFTO (Figure 2-29B) shows two sets of peaks, corresponding to monomer and dimer as labelled, with the clusters of peaks observed due to the formation of different charge states in the mass spectrometer.

Comparison of this to spectra obtained for I367A mFTO and hFTO (Figure 2-29A and C) show that each of the proteins exist as a mixture of monomeric and dimeric forms. However, the proportion of dimeric I367A mFTO is significantly reduced relative to that in the wildtype mFTO, providing a further piece of evidence to suggest that the I367 residue plays an important role in FTO dimerisation. Unfortunately, obtained spectra were less clean for C-FTO proteins, due to the *N*-terminal glycosylation previously described, but CmFTO also exhibited clear monomer and dimer peaks (Figure 2-30).

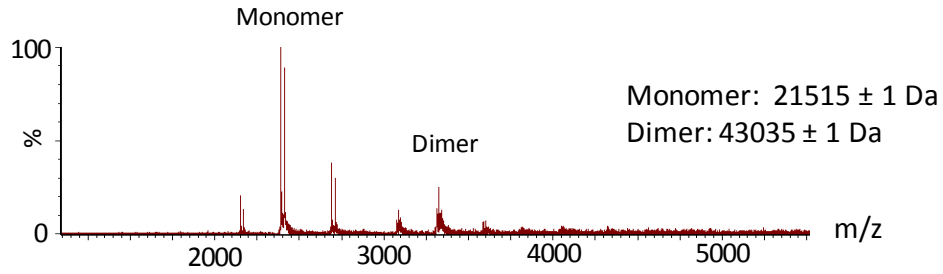


Figure 2-30: Non-denaturing electrospray ionisation mass spectrometry analyses carried out on a Waters Synapt™ HDMS™, showing the oligomeric composition of CmFTO. The three predominant peaks for the monomer correspond to +10, +9 and +8 charge states; predominant peaks for the dimer correspond to +14, +13 and +12 charge states.

Some reports suggest that data obtained using ESI mass spectrometry may not fully correlate with data from other techniques, as in the mass spectrometer the protein is present in a "non-solvated" gas-like state[287]. Additionally, although the detection of higher order species using mass spectrometry is clearly apparent here, there is some concern that low levels of dimeric species can be common low level contributions in ESI mass spectra[288], and observation of multimers may be dependent on the ESI sampling conditions used[289-290].

2.4 2OG Turnover

2.4.1 2OG turnover assay method

Activity of Fe(II) and 2OG-dependent oxygenases can be determined by an assay measuring turnover of 2OG, one of the required co-substrates. A [1-^{14}C]-labelled 2OG substrate is used in addition to unlabelled 2OG, and as these are turned over by the enzyme a corresponding proportion of $^{14}CO_2$ gas is released (Figure 2-31). This is captured by basic hyamine hydroxide (present in a small tube contained within the reaction vessel), as a hydrogen carbonate adduct. To allow time for the CO_2 to be fully captured the enzymatic reaction is quenched by addition of excess methanol (denaturing the protein present) and incubated on ice for 20 min. The β-emission of the ^{14}C labelled ionic adduct is measured by liquid scintillation counting[291].

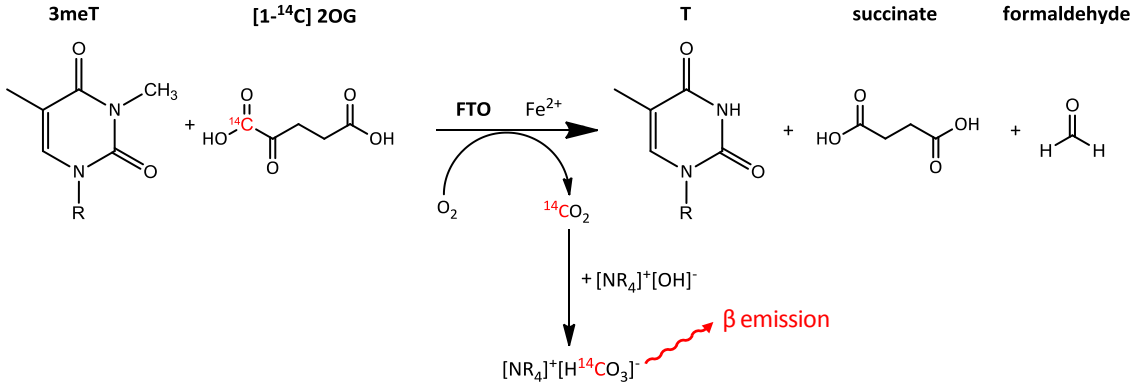


Figure 2-31: Demethylation of 3meT by FTO coupled to turnover of [1-^{14}C]-2OG, with release of $^{14}CO_2$, captured as a hydrogen carbonate adduct by hyamine hydroxide, then detected by β-emission.

A complication of this assay is that most 2OG dependent oxygenases are able to turn over 2OG even in the absence of substrate, a reaction which is uncoupled to substrate hydroxylation[13]. In this situation it is believed that the postulated Fe(IV) ferryl intermediate may become "trapped" in this oxidation state, so deactivating the enzyme from catalysing further hydroxylations. Therefore ascorbate is added to reactions to potentially reduce any Fe(IV) back to the catalytically active Fe(II) state, through its own oxidation to dehydroascorbic acid[292-293].

2OG turnover of mFTO had been previously studied by Dr Kirsty Hewitson. This assay was used for initial identification of the FTO substrate, through observation of elevated

2OG turnover in the presence of methylated oligonucleotides. Subsequent studies identified inhibition of mFTO by common generic 2OG inhibitors, such as NOG, Co(II) and 2,4-PDCA[165].

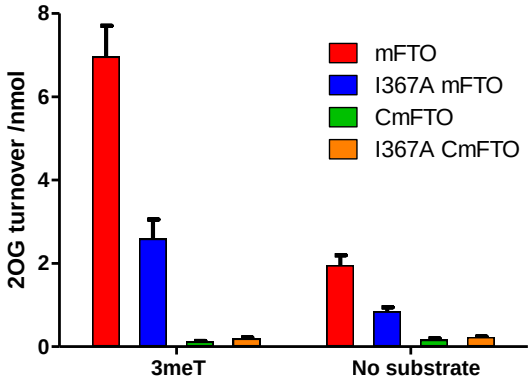


Figure 2-32: I367A mFTO (blue) shows reduced catalytic activity in a 2OG turnover assay with a 3meT-containing 18mer oligonucleotide substrate, compared to full length mFTO (red). Neither CmFTO (green) nor I367A CmFTO (orange) show significant 2OG turnover. Reaction incubation time 10 min. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

To investigate the effect of the I367A substitution on the activity of mFTO the 2OG turnover capacity was investigated. In the presence of a 3-methylthymine containing 18mer oligonucleotide substrate the catalytic activity of mFTO I367A was only ~40% that of wildtype mFTO (Figure 2-32). The turnover of the *C*-terminal domain proteins was also investigated and, as anticipated because they lack the DSBH core, neither CmFTO nor CmFTO I367A were capable of 2OG turnover (Figure 2-32).

As introduction of the I367A substitution prevents FTO dimerisation, and the full length I367A protein has reduced activity, FTO dimerisation may be required for effective catalytic activity and 2OG turnover. Addition of *C*-terminal domain proteins to the full length mFTO protein may therefore affect 2OG turnover or activity of the full length protein, perhaps by preferentially binding to the *C*-terminal domain of the full length protein and therefore blocking dimerisation of pairs of full length proteins. To determine this, the 2OG turnover efficiency of mFTO in the presence of purified CmFTO proteins was investigated. mFTO was mixed with either CmFTO, I367A CmFTO or buffer in a 1:1 ratio and the mixtures preincubated on ice for 30 minutes prior to initiation of the assay by addition of the other assay components. The assay was

then incubated at 37 °C for 10 min and quenched on ice for 20 min. As the 2OG turnover capability of the CmFTO proteins was previously determined to be almost insignificant, the turnover detected may be assumed to be entirely attributable to the full length protein. However, no significant difference was observed in the 2OG turnover by mFTO on addition of CmFTO proteins either in the presence or absence of oligonucleotide substrate (Figure 2-33), therefore suggesting that interaction with the *C*-terminal proteins does not significantly alter the activity.

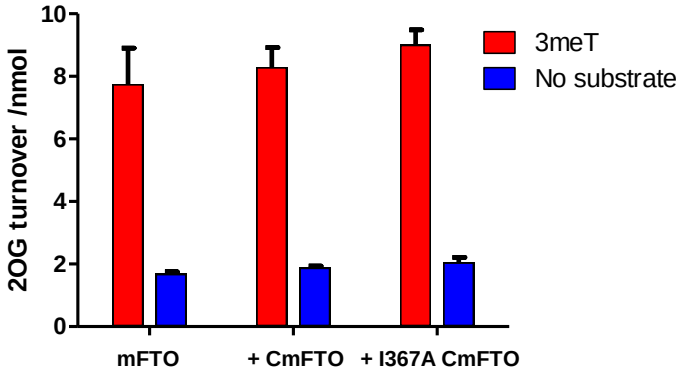


Figure 2-33: The 2OG turnover of mFTO either in the presence or absence of oligonucleotide substrate is not affected by addition of CmFTO proteins. Preincubation at 4 °C, 30 min; reaction incubation at 37 °C, 10 min. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

2.4.2 hFTO

Despite the high sequence identity between the human and murine FTO homologues, when these studies were initiated no significant activity of hFTO above background had been observed in this laboratory, using a LC-MS based assay. Therefore initial activity studies on the purified hFTO were initiated.

2OG turnover assays were carried out to compare the turnover of 2OG by the two full length hFTO constructs (hFTOa protein was provided by Dr Celia Webby). Deletion of the fusion peptide from hFTOa to form hFTO may affect activity, as in the hFTO construct the His-tag is fused directly to the *N*-terminus of FTO, without an intermediate linker region. However results for the two protein constructs (Figure 2-34) showed that both have similar 2OG turnover with 18mer oligonucleotide substrates containing a single methylated base. Uncoupled turnover observed in the absence of prime substrate (common to most 2OG-dependent oxygenases) is also high (~40 % of

that observed in the presence of the likely oligonucleotide substrate) relative to that observed in our lab for other 2OG-dependent oxygenases. This is the first time that elevated activity in the presence of oligonucleotides has been observed with human FTO in this laboratory. It was also subsequently shown by others that hFTO is capable of demethylating 3meT and 3meU lesions in DNA and RNA respectively[239].

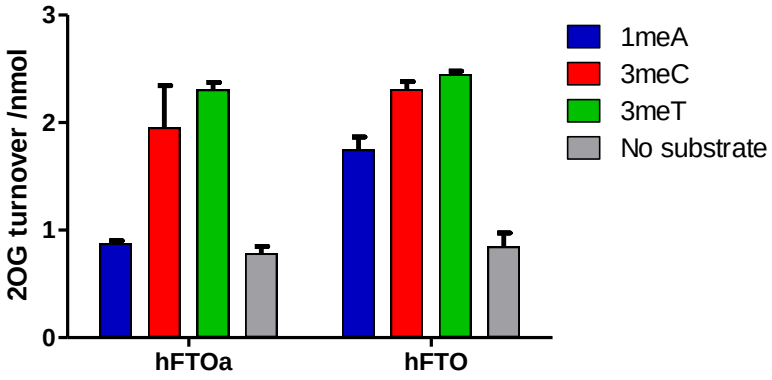


Figure 2-34: 2OG turnover activity of hFTO constructs in the presence or absence of 18mer oligonucleotides containing a single methylated base at position 3. Reaction incubation time 10 min. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

2.5 Circular Dichroism

The secondary structure of the purified mFTO proteins was investigated using Circular Dichroism (CD) spectroscopy. The spectra obtained (Figure 2-35) suggest that the proteins all have well folded structures. Any minor differences observed in curve amplitudes are likely to be due to concentration effects.

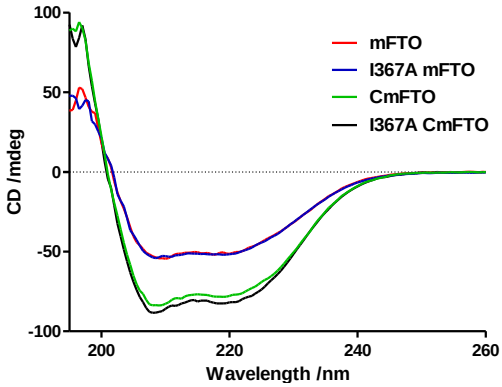


Figure 2-35: Circular dichroism spectra showing the secondary structures of 0.25 mg/ml mFTO (red), I367A mFTO (blue), CmFTO (green) and I367A CmFTO (black), at 4 °C. Data points represent the mean of triplicate experiments.

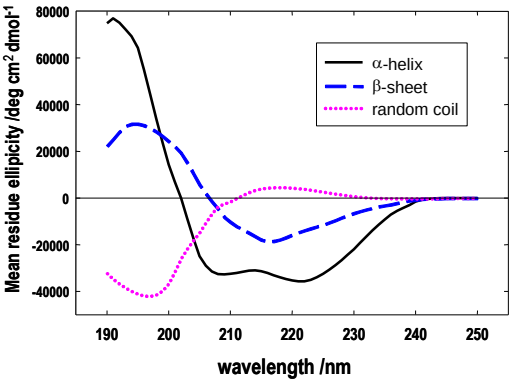


Figure 2-36: Fasman standard circular dichroism curves of polylysine in different structural forms. Taken from[294].

By comparison to standard curves (Figure 2-36) and deconvolution of the spectra obtained it is possible to determine the proportion of each structural element (i.e. α-helix/β-strand) present in the proteins[294-295]. In contrast to the full-length proteins, which contain the DSBH "core" domain of the 2OG-dependent oxygenases (comprising 8 β-strands with surrounding loops and α-helices), the *C*-terminal domain proteins are predominantly α-helical, as implied by previous structural predictions.

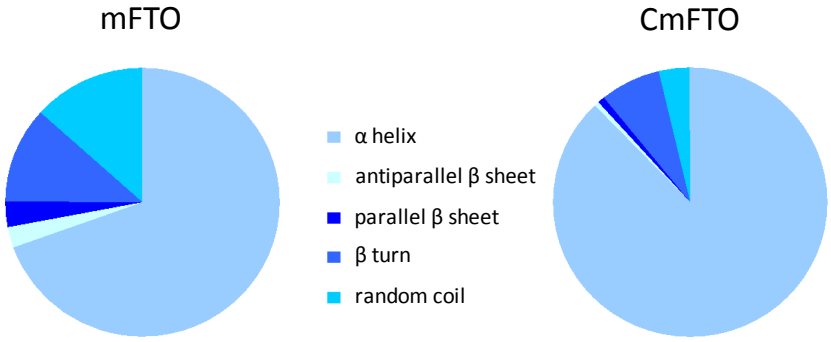


Figure 2-37: Secondary structure predictions showing relative proportions of secondary structure elements present in CmFTO (left) and mFTO (right), obtained through spectral fitting and deconvolution of circular dichroism spectra (Figure 2-35) using the program CDNN (Chirascan). A greater proportion of α-helix is identified in the *C*-terminus of mFTO (CmFTO) than for the full length mFTO. Proportions obtained for the analogous proteins for the I367A variants are very similar (data not shown).

Deconvolution of the spectra obtained assigns 91% of the secondary structure of CmFTO as helical, compared to 69 % for the full length mFTO (Figure 2-37), with a very similar breakdown of secondary structure elements for the I367A proteins, due to their highly similar spectra. A similar spectral difference was observed for hFTO and ChFTO (Figure 2-38). This difference in CD may be observed more clearly in CD

difference spectra (Figure 2-39), which show the residual CD between two spectra. This shows the clear difference in CD between *C*-terminal and full length constructs, but the high similarity between WT and I367A proteins.

The I367A substitution does not appear to alter the overall secondary structure fold of mFTO proteins (Figure 2-35), despite preventing dimerisation. This suggests that the I367A proteins are correctly folded, therefore the reduced activity and lack of dimerisation is not due to grossly altered or incorrect folding, but that isoleucine 367 is involved in dimer formation.

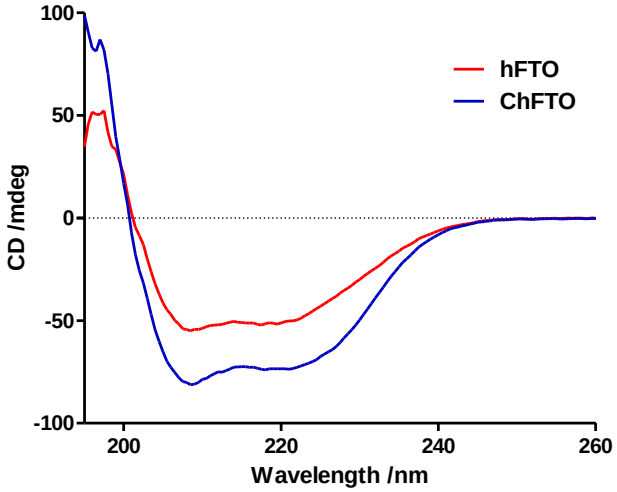


Figure 2-38: Circular dichroism spectra showing the secondary structures of hFTO (red), and ChFTO (blue) at 0.25 mg/ml, 4 °C. Data points represent the mean of triplicate experiments.

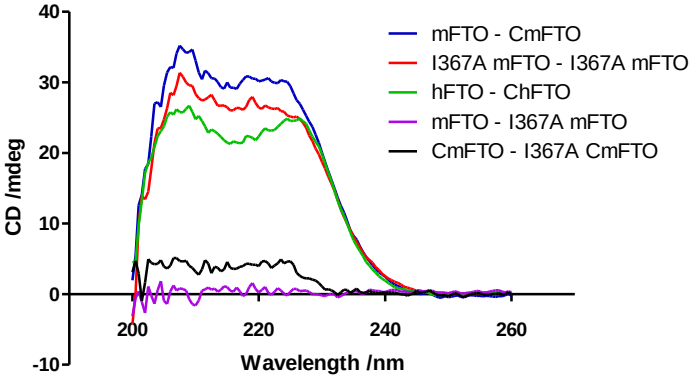


Figure 2-39: Difference circular dichroism spectra, identifying the increased proportion of α-helical content in the *C*-terminal domain relative to the wildtype FTO proteins, and indicating overall structural similarity between the wildtype and I367A variants: mFTO - CmFTO (blue), I367A mFTO – I367A CmFTO (red), hFTO – ChFTO (green), mFTO – I367A mFTO (purple), CmFTO – I367A CmFTO (black).

2.6 Oligonucleotide binding by Mass Spectrometry

To investigate whether oligonucleotides were capable of binding to the *C*-terminal domain of mFTO, non-denaturing MS was carried out on a Waters Synapt™ HDMS™, by Dr Frank Sobott. This technique also uses ESI and may provide additional information compared to standard mass spectrometers, due to the possibility of differentiation between species of different sizes and shapes as well as mass. Protein was mixed in a 1:1 ratio with oligonucleotides, and incubated for 30 min at room temperature, before spraying in the mass spectrometer.

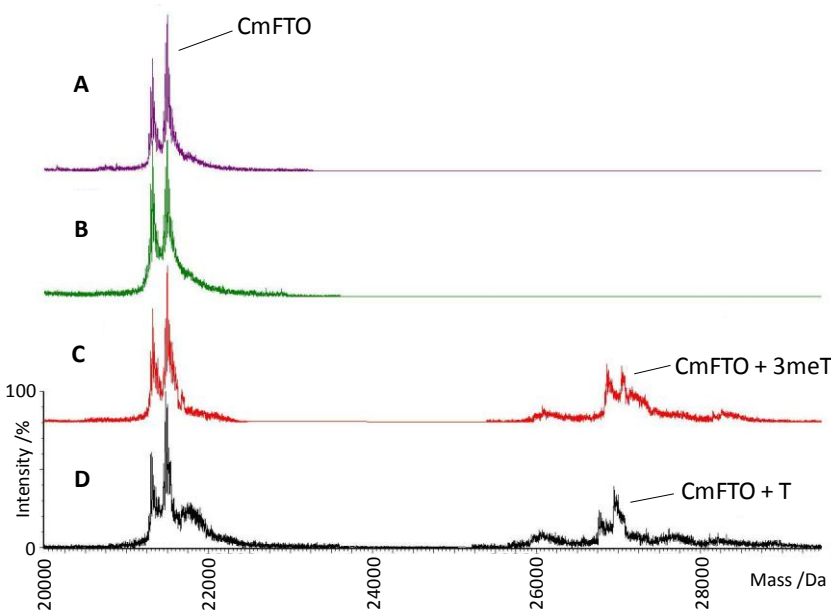


Figure 2-40: Soft ionisation mass spectrum for CmFTO binding to (A) 1meA, (B) 3meC, (C) 3meT, (D) non-methylated T [T] 18mer oligonucleotides.

Surprisingly, no binding of 1meA and 3meC to CmFTO was observed, however a control oligonucleotide without a methylated base showed a small degree of binding (~20%) Binding of the 3meT oligonucleotide was clearly visible (~30%), suggesting a preference for a sequence containing this modified base (Figure 2-40). Differentiation between size and shape as well as mass was attempted using SYNAPT mass spectrometry, which suggests that in the presence of the oligonucleotide containing 3meT the CmFTO/3meT complex structure may become more compact than the protein alone (data not shown), although no definitive conclusions could be drawn from this.

2.7 Differential Scanning Fluorimetry

2.7.1 Differential Scanning Fluorimetry method

Differential Scanning Fluorimetry (DSF) may be used to obtain protein melting curves (Figure 2-41), to determine the melting temperature (T_m) of the proteins investigated. Prior to investigation, proteins were buffer exchanged into 50 mM HEPES pH 7.5, 50 mM NaCl. Tris buffer is not favourable for use in this method due to its relatively high temperature coefficient of -0.026 pH units per °C[296] which may produce artificial data. In contrast, the temperature coefficient of HEPES is -0.014 pH units per °C[297]. Therefore the reduction in pH that will be experienced by the protein as the temperature is raised from 25 °C to 95 °C will be reduced (pH 7.5 to pH 5.7 for tris, or to pH 6.5 for HEPES).

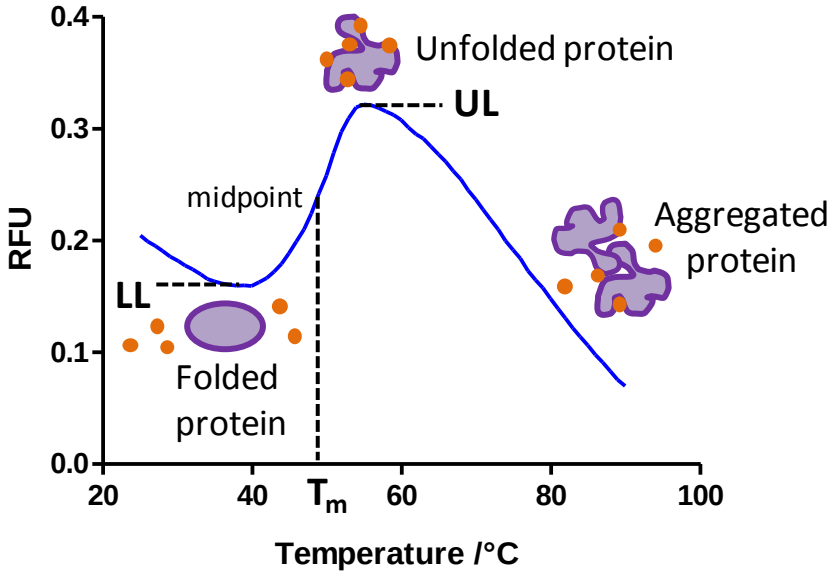


Figure 2-41: Typical recording of fluorescence intensity versus temperature for protein unfolding in the presence of SYPRO orange (represented as orange dots). As the temperature of the system is increased the proteins begin to denature through unfolding of their well defined tertiary structure. This exposes the previously buried internal hydrophobic residues to the aqueous solvent, and promotes binding of the SYPRO orange fluorophore added to the solution, with a concomitant ncrease in fluorescence. The fluorescence reaches a peak with maximal protein unfolding, but fluorescence is subsequently reduced with further temperature increase, as the unfolded protein chains aggregate in solution, masking the hydrophobic residues and preventing binding of SYPRO orange. Boltzmann curves can be fitted to the obtained sigmoidal curves with the midpoint representing the melting temperature (T_m) of the protein. LL: Lower limit/minimum intensity, UL: Upper limit/maximum intensity (Equation 8-1).

This technique may be extended to determine buffer, pH, metal and small molecule conditions which stabilise a given protein. Additives to the protein may alter the protein's T_m, causing a "T_m shift". Positive T_m shifts suggest protein stabilisation in the presence of the additive, whereas negative T_m values indicate protein destabilisation. The resulting information may be used to improve protein storage conditions and enhance crystallisation attempts[298].

2.7.2 DSF optimisation

Initially, the suitability of the proteins in this technique was assessed. A range of protein concentrations was tested to ensure that the proteins underwent a standard two-state transition, and to determine an appropriate concentration for addition of compounds. Due to protein availability, the full length protein investigated was hFTO (Figure 2-42), whereas CmFTO was used for studies on the *C*-terminal domain (Figure 2-43). Samples were run in triplicate, with T_m values determined for each replicate. Both proteins exhibited normal two state transition curves across the concentration range tested however curves for CmFTO at lower concentrations (1 and 2 μM) were of low intensity, making accurate T_m determination difficult. The importance of maintaining a constant enzyme concentration is apparent from the range of T_ms determined for CmFTO, with a range of ~2 °C observed. To minimise protein used, while obtaining well shaped and reproducible curves, a 4 μM protein concentration will be used for future experiments.

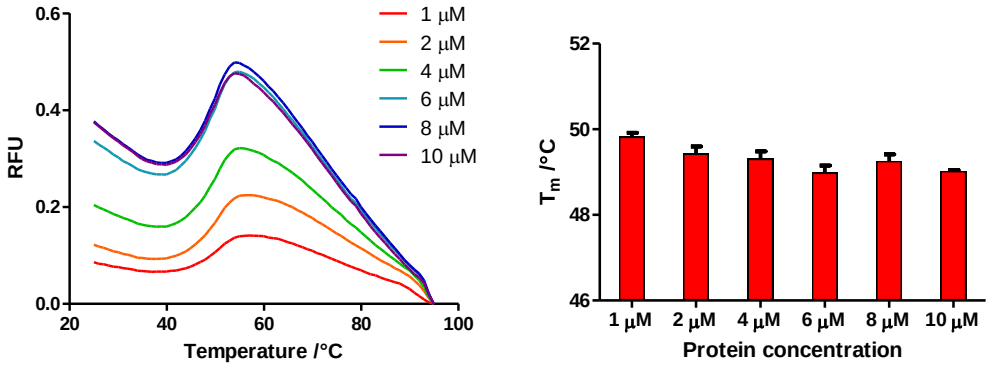


Figure 2-42: Differential scanning fluorimetry of hFTO at a range of concentrations; represented as melting curves and derived T_m values. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

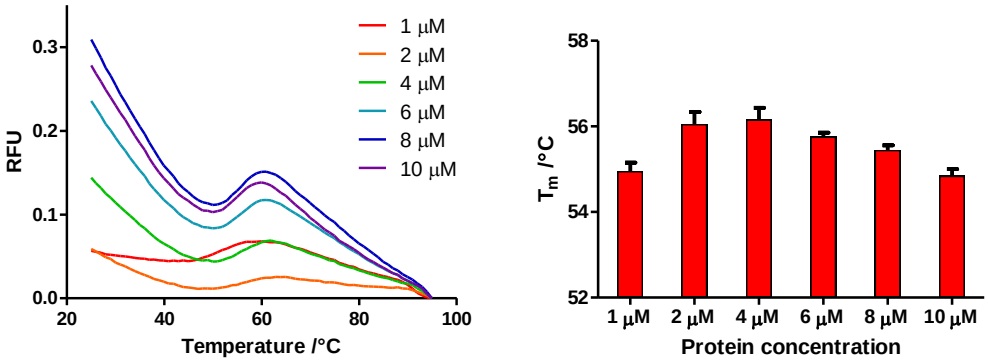


Figure 2-43: Differential scanning fluorimetry of CmFTO at a range of concentrations; represented as melting curves and derived T_m values. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

The T_m shifts observed on addition of metal (II) ions was then investigated. The greatest T_m for hFTO is observed with Ni(II), which is the metal used in the crystallisation of hFTO[264]. Co(II) appears to also give an increased T_m with hFTO, but due to the low magnitude of the observed peak it is difficult to fit a Boltzmann curve, giving a large variation. Addition of Co(II) to CmFTO produced an anomalous melting curve, so no T_m could be obtained, suggesting that significant protein destabilisation may occur. Interestingly, although Ni(II) stabilises the full length FTO, it gives a significantly reduced T_m with CmFTO, suggesting that the stabilisation produced by addition of the metal is mediated by binding at the metal centre in the *N*-terminal domain. Zn(II) causes a slight increase in T_m for both hFTO and CmFTO, suggesting that this may cause stabilisation. As no large T_m shifts were observed for CmFTO it suggests that there is no metal binding site within this domain, in agreement with structural predictions.

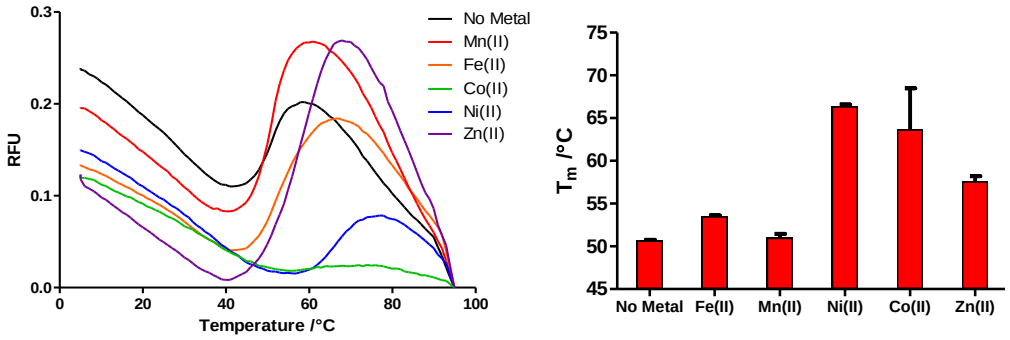


Figure 2-44: Differential scanning fluorimetry of hFTO, with addition of 50 μM metal (II) as indicated. Data are represented as melting curves and derived T_m values. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

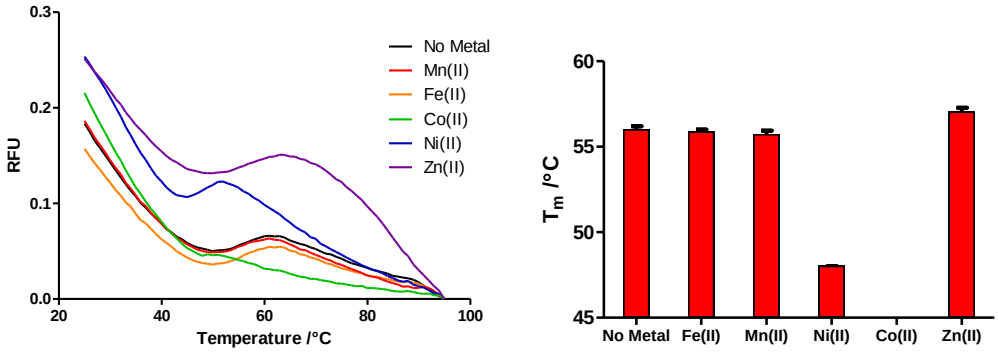


Figure 2-45: Differential scanning fluorimetry of CmFTO, with addition of 50 µM metal (II) as indicated. Data are represented as melting curves and derived T_m values. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

2.7.3 CmFTO compound screening

In addition to this preliminary characterisation, a further DSF screen to study the *C*-terminal domain of FTO was carried out at the Structural Genomics Consortium (SGC) Oxford, with the help of Chitra Bhatia. As the CmFTO construct does not contain the 2OG oxygenase DSBH core, and has little known homology to other proteins, a wide range of compounds were tested with this protein, to identify any binding partners that may reveal more about the role of the domain. As crystallisation studies of the protein were planned, an initial screen was carried out to identify buffer, salt and pH conditions for optimal protein stability. The standard buffer used for the initial assay (used as the reference condition for T_m calculation) was 50 mM HEPES pH 7.5, 150 mM NaCl, as this is routinely used in SGC screens[298].

Results shown in Figure 2-46A suggest that CmFTO is most stable in the pH range 6.5-8.3. Variation of the NaCl concentration revealed that the previously used protein storage conditions (50 mM NaCl) were found to be optimal, with destabilisation relative to this at different NaCl concentrations.

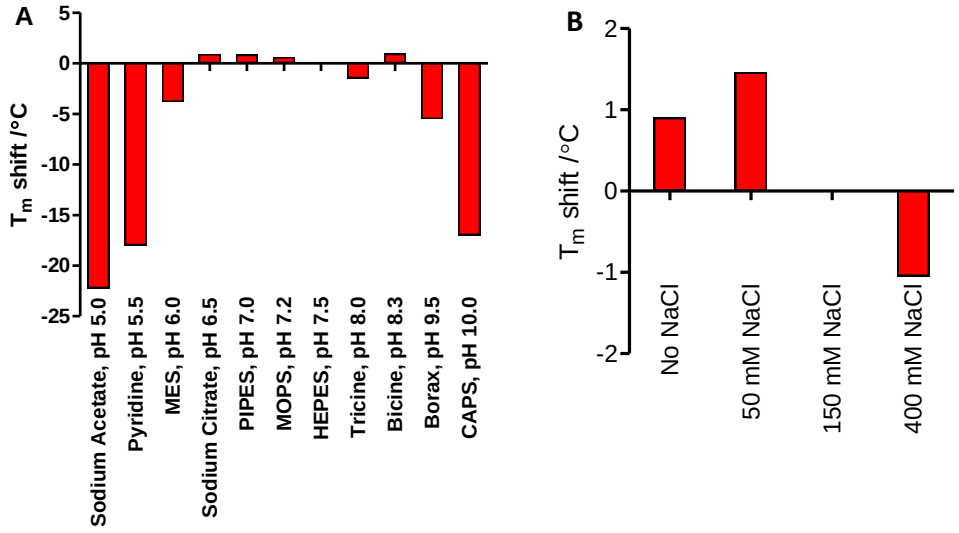


Figure 2-46: DSF screen of CmFTO: (A) Variation of pH and buffer conditions; all buffers at 50 mM, with addition of 150 mM NaCl; (B) 50 mM HEPES pH 7.5, with addition of NaCl to concentration indicated. T_m shift results referenced to 50 mM HEPES pH 7.5, 150 mM NaCl[298].

In total CmFTO was screened against >300 further compounds and conditions from the SGC collection, including sugars, nucleotides, amino acids, 2OG analogues, DNA oligonucleotides and peptides, and other common protein cofactors (e.g. NADH).

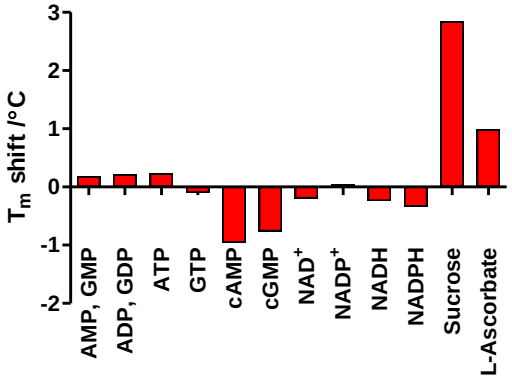


Figure 2-47: DSF screen of CmFTO, with addition of small molecules at 5 mM, except sucrose (500 mM) and L-ascorbate (100 mM). Referenced to 50 mM HEPES, pH 7.5, 150 mM NaCl.

The presence of individual nucleosides (e.g. guanosine) led to abnormal transitions (not shown), which could not be fitted to the standard two state model in order to obtain T_m values. No significant effects were observed on addition of nucleotide phosphates, or other compounds containing base structures, having homology to the substrate of FTO (Figure 2-47). Addition of individual amino acids created minor T_m differences. Stabilisation was observed on addition of sugars, exemplified by sucrose (Figure 2-47), although this is common to most proteins screened (Chitra Bhatia, personal

communication), possibly because the standard screens used contain sugars present at a much higher concentration (two orders of magnitude greater, 500 mM vs. 5 mM) than the other compounds in the screen. The relationship between increased compound concentration and T_m is logarithmic[298-299], therefore this significantly higher compound concentration could cause the large observed T_m shift.

A further compound tested was ascorbate, which is added to assays to reduce Fe(IV) that may be formed during assay cycles uncoupled to prime substrate[292]. As observed in Figure 2-47, addition of 1 mM L-ascorbate causes a T_m shift of ~+1 °C.

Other screens containing peptides and 2OG mimics showed no significant effects on protein stability (T_m within 1°C of control), therefore identification of compounds that could significantly stabilise the protein for crystallisation was not possible. A selection of short (6-30mer) unmodified oligonucleotides was also investigated to determine whether any binding may be observed. However, the sequences investigated yielded a T_m shift of -1 – 0 °C, suggesting slight protein destabilisation.

2.7.4 hFTO oligonucleotide binding

Initial studies of binding of oligonucleotides to hFTO (Figure 2-48) showed that a large T_m shift was obtained when a non-methylated oligonucleotide was incubated with 2OG and Fe(II), suggesting that oligonucleotides are bound most successfully in the presence of the two co-substrates.

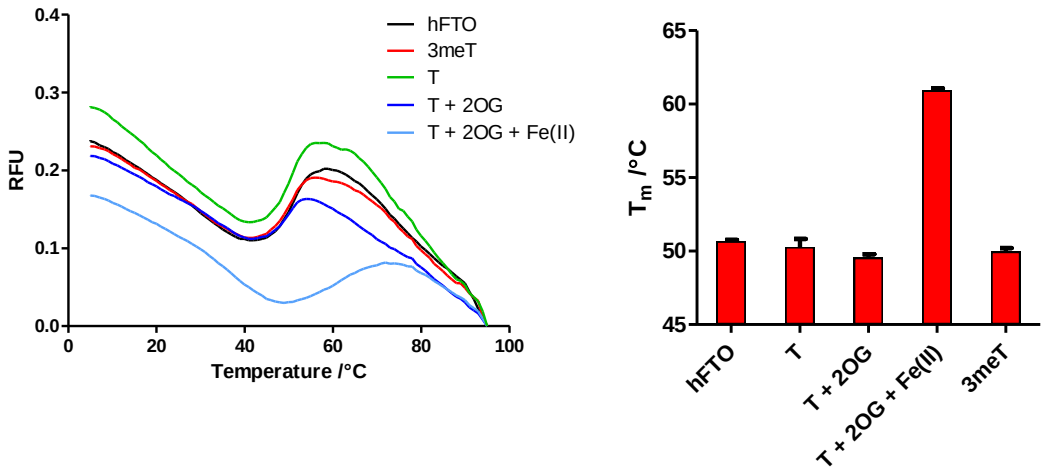


Figure 2-48: DSF of hFTO with 18mer oligonucleotides containing 3meT: (3meT residue) or T (unmodified T residue). Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

2.8 CmFTO crystallisation attempts

Given the possibility of the CmFTO construct comprising a discrete protein domain, crystallisation trials of the protein were attempted. Prior to this, dynamic light scattering was carried out to confirm protein monodispersity. CmFTO crystallisation was attempted at the Crystallisation Facility, Department of Biochemistry. Protein was buffer exchanged into 50 mM tris pH 7.5 as the presence of relatively high salt concentrations can lead to salt crystallisation rather than protein crystallisation. A wide range of conditions were probed using several commercially available sparse matrix screens, with trays stored both at 4°C and 20°C and monitored periodically for signs of crystal development. Unfortunately no conditions were found to enable protein crystallisation therefore alternate biophysical investigations of CmFTO were pursued.

2.9 FTO crystal structure

While these studies were underway, a crystal structure of hFTO was published[264] (Figure 2-49), providing a useful complement to the findings presented here. A catalytically active ΔN31 hFTO construct was used for crystallisation; removal of the *N*-terminal residues has also been employed in crystallisation of other ABH family members[192-194]. The crystal structure was obtained in complex with a 3meT mononucleotide and NOG, a 2OG mimic used as a generic inhibitor of 2OG oxygenases.

hFTO forms a crystallographic trimer, mediated by W278. The structure also identified separate *N*- (residues 32-326; mFTO residues 32-323) and *C*-terminal (residues 327-498; mFTO residues 324-495) domains, as discussed in this work. The structure of the *N*-terminal domain of hFTO has the greatest similarity to AlkB. The *C*-terminal domain of hFTO is helical, as predicted, with formation of a three-helix bundle, one end of which forms extensive interactions with the *N*-terminal domain. Alteration of residues involved in this interaction surface led to decreased demethylation activity, suggesting that the presence of the *C*-terminus stabilises the DSBH domain.

The authors were able to prepare the *N*- and *C*-terminal domains of FTO as separate constructs, but both of these separately expressed domains were catalytically inactive. When the two domain constructs were coexpressed in *E. coli* a stable complex was obtained, and partial rescue of demethylation activity was observed.

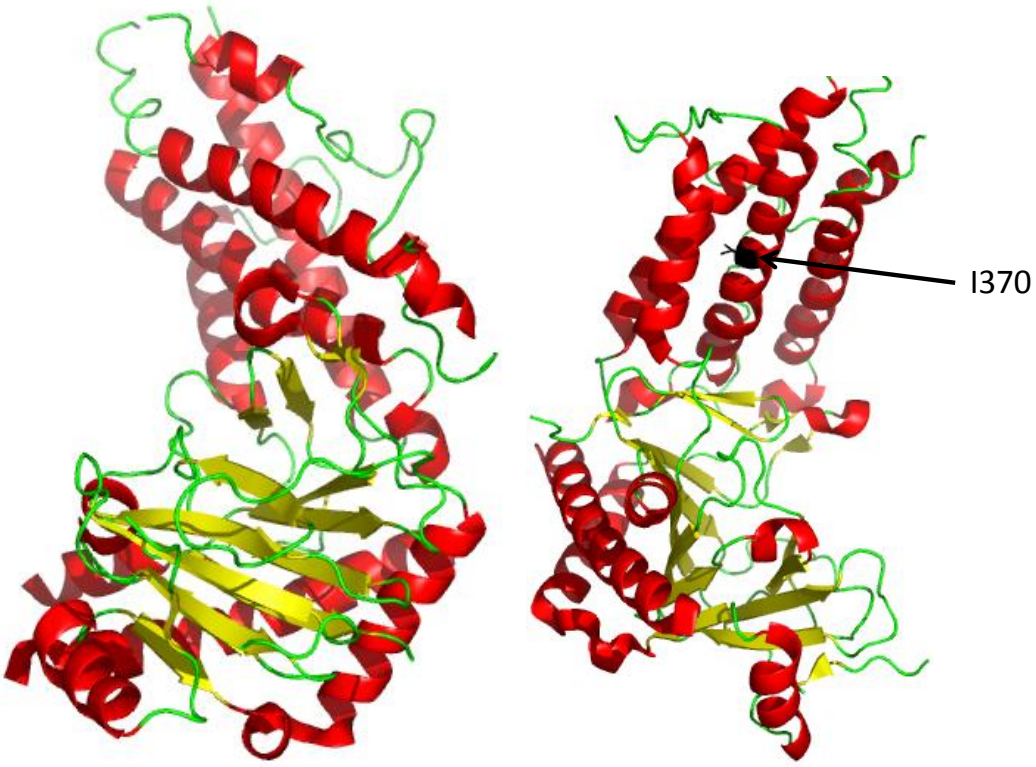


Figure 2-49: Two views of crystal structure of ΔN31 FTO. (A) Positioned to view *N*-terminal DSBH (bottom), and separate *C*-terminal domain (top). (B) Rotated to show I370 residue (highlighted in black). PDB ID 3LFM[264].

However, the work presented in this paper also raises several questions when considered with the work presented here. Gel filtration analyses suggested that dimerisation was not observed. However the ΔN31 hFTO construct was prepared in this laboratory by Wei Shen Aik, and the gel filtration purification step gave a UV trace extending across a wide elution volume range, as seen for the other FTO constructs investigated, suggesting that oligomerisation again occurs. Unfortunately crystallisation attempts for the ΔN31 hFTO protein have not yet yielded crystals.

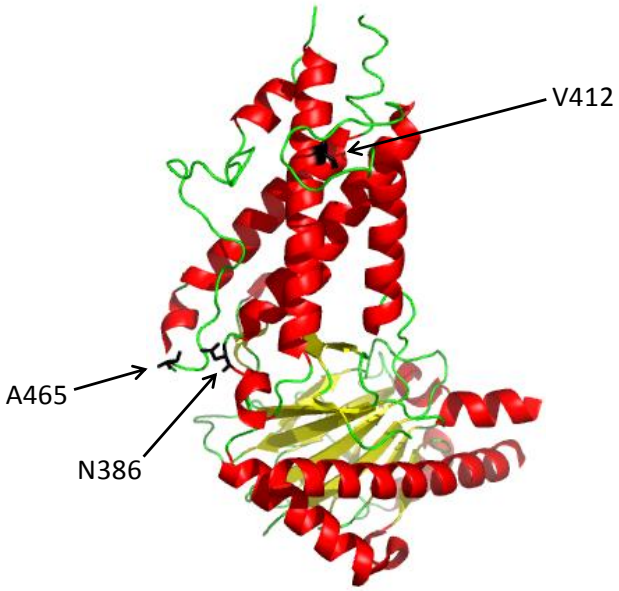


Figure 2-50: *C*-Terminus of hFTO with residues different between I.M.A.G.E. and NCBI mouse sequences highlighted in black. PDB ID 3LFM[264].

The I370 residue, which is the equivalent of the murine I367 residue, is buried in the helical *C*-terminal domain (labelled in Figure 2-49), therefore the authors suggest that alteration of this residue may not have a major effect on potential protein dimerisation. However, replacement of I367 with a different residue may alter the structure of the *C*-terminal domain, and therefore destabilise the catalytic *N*-terminal domain. Additionally, this crystal structure allows analysis of the position of the three residues different between the I.M.A.G.E. and NCBI mFTO sequences (Figure 2-50). Two of the residues (A465 and N386) are found near the interface between the two domains of the protein, but appear to project out into the solvent. The third residue, V412, is found in one of the helices forming the helix bundle, but here projects out towards a loop region, rather than being significantly involved in formation of the helix bundle. This therefore suggests that substitutions of these residues with relatively similar species is unlikely to have a large structural effect.

Following publication of this crystal structure, the PDB coordinates of the *C*-terminal domain were submitted in a DALI search, which allows comparison of protein structures in three dimensions. The structural search mode may reveal biologically interesting similarities that are not detectable by comparing sequences alone[300]. This

identified that the *C*-terminal domain of FTO has a significant structural similarity to the PHAT domain of the *Drosophila melanogaster* protein SMAUG. This protein plays a role in translational regulation in early Drosophila embryogenesis by repression of maternal nanos mRNA translation, by recognition of translation control element hairpins[301-302]. The structure of this multidomain protein was solved in 2003, and the region with similarity to the *C*-terminal domain of FTO forms a cylindrical 5 helix bundle[303]. This domain allows RNA recognition via an interaction surface formed from many basic residues, forming a not previously observed motif. This therefore suggests by structural similarity that the *C*-terminal domain, as postulated, may be responsible for nucleic acid binding.

2.10 In vivo *studies of I367F mFTO*

In parallel to the work described here, *in vivo* studies of the I367F mFTO mutation were carried out by researchers at MRC Harwell, and the Department of Physiology, Anatomy and Genetics, Oxford.

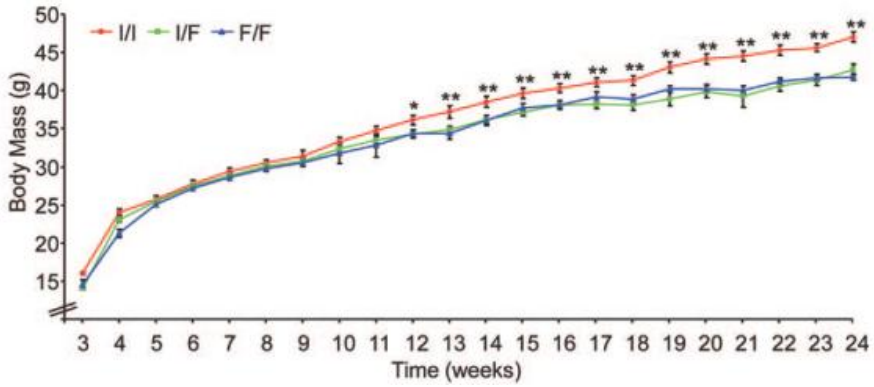


Figure 2-51: Body weights of male heterozygous (I/F; n = 21) and homozygous (F/F; n = 28) I367F FTO mice, and wildtype littermates (I/I; n = 15). Statistical analysis was performed using Student's t-test. * P<0.05; ** P<0.01 for differences between I367F FTO heterozygous or homozygous mice and wildtype littermates. Figure prepared by Dr Chris Church[242].

Mice heterozygous and homozygous for I367F mFTO were developed and studied by Dr Chris Church. In general, the same phenotype was observed for mice either homozygous or heterozygous for the I367F mutation. A maturity onset reduction in weight was observed from 12 weeks for both I367F mFTO heterozygotes and homozygotes (Figure 2-51), but with no difference in food intake, or physical activity.

Dual Energy X-ray absorptiometry (DEXA) analysis of 24 week old I367F mFTO mice showed reduced total fat mass per mouse compared to wildtype counterparts.

18 week I367F mFTO mice showed increased O_2 consumption and CO_2 production, causing an increased respiratory exchange ratio (RER), suggesting that the mutation causes an increase in whole body metabolism and a switch to greater carbohydrate metabolism. There was no difference in a glucose tolerance test at either 12 or 16 weeks. However, markers of carbohydrate metabolism, including serum levels of glucagon, glucose and lipids, were increased at 24 weeks.

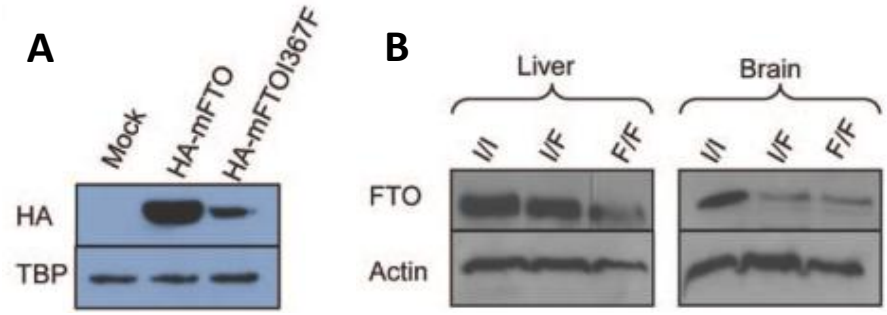


Figure 2-52: (A) Cos7 cells expressing HA-mFTO or HA-mFTO I367F lysed with RIPA buffer and separated by SDS-PAGE. Immunoblotted with an anti-HA antibody and anti-TATA-binding protein (TBP) antibody as a loading control for nuclear proteins. (B) Protein from liver and brain of wildtype (I/I), heterozygous (I/F) and homozygous (F/F) animals were separated by SDS PAGE, followed by an immunoblot against a polyclonal rabbit anti FTO antibody with an anti actin antibody as a loading control. Figure prepared by James McTaggart[242].

To further investigate this, microarray based gene expression profiling was carried out by Dr Sheena Lee, to identify genes with altered expression in homozygous I367F mice, as compared to wildtype mice. This did not identify any changes in metabolic genes, but some genes involved in carbohydrate catabolism were slightly upregulated in the skeletal muscle of I367F mice. Additionally, inflammation genes were down-regulated in abdominal white adipose tissue of I367F mice, consistent with their decreased mass, but genes involved in fat synthesis were upregulated, maybe as a secondary adaptation to compensate for the decreased fat levels.

Immunoblotting of cell lysates by James McTaggart showed that the I367F substitution results in reduced mFTO protein levels, both for transient transfection of HA-tagged I367F mFTO in Cos7 cells, and in the brain and liver of heterozygous and homozygous

I367F mFTO mice (Figure 2-52). This reduced expression is likely to impair mFTO function in I367F animals, but as some I367F mFTO protein appears to be expressed and correctly targeted to the nucleus a total knockout is not observed.

2.11 Conclusions

These studies have demonstrated identification and purification of CmFTO, a novel domain at the *C*-terminus of FTO, which was shown to have predominantly helical secondary structure by circular dichroism. *In vitro* wildtype mFTO dimerises, possibly via its novel helical *C*-terminal domain, and can exist in both monomeric and dimeric forms, as observed by analytical gel filtration and mass spectrometry. It is likely that the *C*-terminus of hFTO enables dimerisation in a similar manner. Attempts to purify a series of *C*-terminally truncated constructs were unsuccessful, reinforcing the important role of this protein domain. A difference in FTO sequences between inbred mouse strains was identified, and the corresponding sequences were cloned, but *in vitro* expression of proteins was poor. Differential scanning fluorimetry of CmFTO was performed in an attempt to identify any potential stabilising additives, thereby giving an indication of the role of the protein domain, or aid crystallisation attempts. Unfortunately no significant alterations to the T_m of CmFTO were observed, but hFTO was seen to be stabilised on incubation with Fe(II), 2OG and an unmodified oligonucleotide, suggesting that oligonucleotides are bound most strongly in the presence of the co-substrates.

Substitution of the I367 residue in the *C*-terminus of FTO for phenylalanine resulted in insoluble protein when expressed in *E. coli* and a substantially reduced level of expression in mammalian cells. However alanine substitution at residue 367 resulted in soluble protein. I367A mFTO has a similar secondary structure to mFTO, although the fraction of dimeric protein was significantly reduced and its catalytic activity was decreased. Therefore it is possible the I367A (and presumably I367F) substitution disrupts dimerisation and results in a reduction of catalytic activity. However it could

also be possible that, in addition to the presumed lower activity of I367F mFTO, reduced stability of I367F mFTO *in vivo* may lead to aggregation and/or proteolytic degradation, contributing to the mouse phenotype.

The I367F point mutation in the mFTO gene increases energy expenditure leading to thinner mice, without changing physical activity. This substitution occurs in a structurally novel domain and modifies FTO function probably by altering its dimerisation state[242]. This may involve a dominant negative effect from the I367F protein, or a hypomorphic allele, where the altered gene product has a reduced level of activity. Heterozygous FTO knockout mice resemble wildtype mice, suggesting that the presumed reduction in mFTO is not sufficient to elicit a phenotype (haploinsufficiency)[241]. However, because heterozygous and homozygous I367F FTO mice exhibited a similar phenotype, the I367F mFTO protein appears to exert a dominant negative effect on mFTO function, possibly by disrupting the wildtype mFTO subunit by formation of a heterodimeric protein complex (i.e. an antimorphic allele). Gene expression profiling revealed increased expression of some fat and carbohydrate metabolism genes, and an improved inflammatory profile, in white adipose tissue of mutant mice.

These data provide direct functional evidence that FTO is a causal gene underlying obesity. This mouse model suggests that a search for human coding mutations in FTO may be informative and that inhibition of FTO activity, either directly at the metal centre in the *N*-terminal DSBH or through disruption of the helical *C*-terminal domain identified here, may be possible for the treatment of morbid obesity.

Chapter 3 - Cellular Studies of FTO

3.1 Previous FTO cell-based studies

Despite significant interest in the link between FTO and obesity, little definitive data have been published describing the role of FTO in the cell. FTO appears to be regulated by the transcription factor CUTL1 via a sequence in the first intron of FTO[251], however FTO may also act as a transcriptional coactivator, facilitating expression from both methylated and non-methylated genes[304]. A role for FTO has been suggested in the control of insulin secretion in pancreatic β-cells, possibly related to the obesity T2D phenotype link[305]. Gene expression analysis has suggested that genes involved in fatty acid catabolism and synthesis are upregulated in mice with a loss-of -function mutation in the FTO sequence, whereas inflammation genes are downregulated[242].

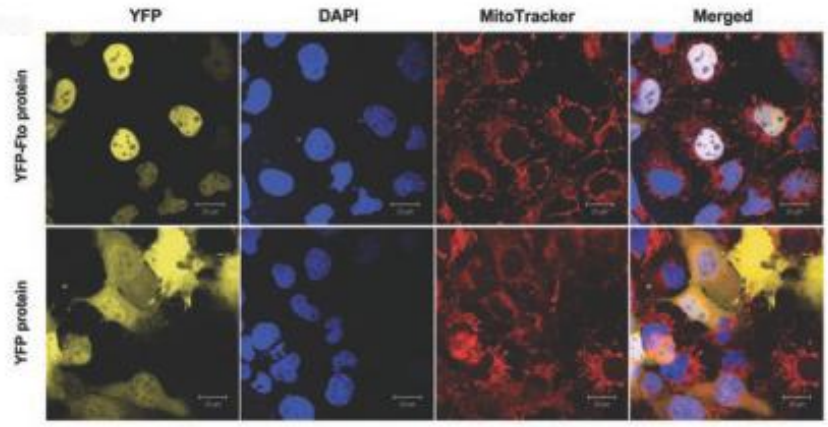


Figure 3-1: Subcellular localisation of YFP-mFTO in COS-7 cells. Confocal fluorescence images of COS-7 cells expressing YFP-mFTO or YFP show YFP-mFTO localising to the nucleus. Nuclei were visualised by DAPI staining and mitochondria with MitoTracker (Invitrogen). Colocalisation of YFP (yellow) and DAPI (blue) in the merged images produces a white signal. Figure adapted from[165].

To aid understanding of FTO function, studies of the subcellular localisation of FTO in mammalian cells were carried out in the Department of Physiology, Anatomy and Genetics, Oxford by Dr Christophe Girard and James McTaggart. mFTO *N*-terminally tagged with yellow fluorescent protein (YFP) was transfected into COS-7 (African green monkey kidney) cells. Confocal microscopy revealed colocalisation of the yellow YFP signal with 4',6-diamidino-2-phenylindole (DAPI), a DNA intercalating dye,

suggesting nuclear localisation of FTO (Figure 3-1)[165]. A similar study using YFP-I367F mFTO transiently transfected in COS-7 cells showed that the I367F mFTO was stable in mammalian cells, and localises to the nucleus (Figure 3-2)[242].

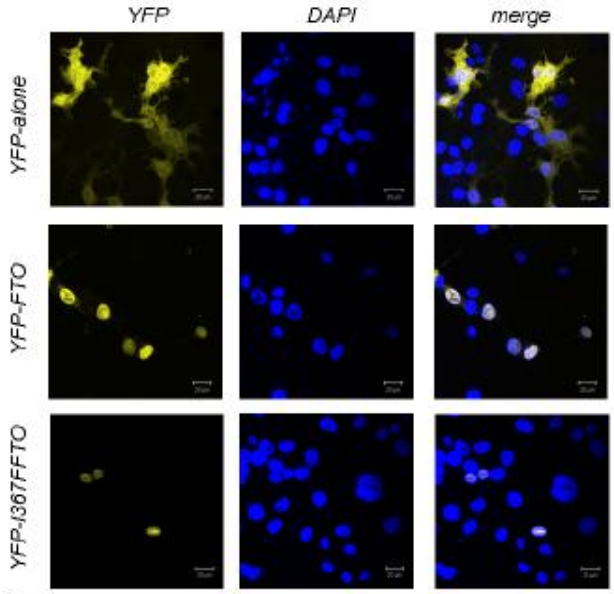


Figure 3-2: Confocal fluorescence images of COS-7 cells expressing YFP, YFP-mFTO or YFP-I367F mFTO, as indicated. Figure adapted from[242].

As a complementary line of investigation to previous *in vitro* analyses and whole animal mouse studies, a number of human FTO (hFTO) constructs appropriate for mammalian protein expression were developed, and expressed in human cell culture lines by transient transfection. In order to understand more about the *in vivo* role of the hFTO protein, cell lysis and immunoprecipitation methods were developed and used for investigation of interacting proteins by immunoprecipitation and mass spectrometry based methods.

3.2 Construct production

3.2.1 Cloning of FTO into mammalian expression vectors

To enable a range of experimental approaches, the hFTO sequence was cloned into several vectors for protein expression in mammalian cell lines. In each of these vectors expression is under the control of the cytomegalovirus (CMV) promoter, therefore overexpression of the gene of interest is induced. The vectors produce an untagged protein construct (pcDNA3), a construct with an *N*-terminal 3xFLAG tag (p3xFLAG-

CMV10) or constructs with either an *N*-terminal or *C*-terminal enhanced green fluorescent protein (EGFP) fusion (pEGFP-C1 and pEGFP-N1 respectively). All primers used for cloning are detailed in Appendix B, and sequences were fully verified by DNA sequencing.

The DNA sequence of hFTO was amplified by PCR from the pET-28a(+) hFTO bacterial expression vector. Three sets of primers were designed in order to allow ligation of the resulting sequence into the different vectors. In each case the primers contained flanking KpnI and BamHI restriction sites, introduced as overhangs. Cloning into both pcDNA3 and pEGFP-C1 was possible using the same pair of primers (hFTO FwdA + RevA). However, appropriate in-frame cloning of hFTO into p3xFLAG-CMV10 required addition of one base into the primer sequence, to ensure that the hFTO sequence was in frame with the *N*-terminal FLAG tag sequence (primers hFTO FwdB + RevA). No stop codon was included in the primers used for cloning into pEGFP-N1 (primers hFTO FwdA + RevB) to allow read through from hFTO into the *C*-terminal EGFP tag. PCR amplification products were run on a 1 % agarose gel, which in each case showed products at the expected size of ~1.5 kbp (Figure 3-3).

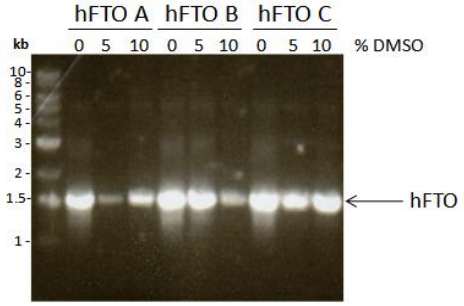


Figure 3-3: 1 % agarose gel showing hFTO PCR amplification products.

The PCR products were digested with the restriction enzymes KpnI and BamHI, and ligated with the four vectors, which had previously been treated with KpnI and BamHI and calf intestinal phosphatase. Test restriction digests of the extracted plasmids visualised by agarose gel electrophoresis identified clones containing an insert of the desired size (Figure 3-4). Correct plasmid constructs were purified on a large scale using a Qiagen Maxiprep kit to produce transfection grade DNA.

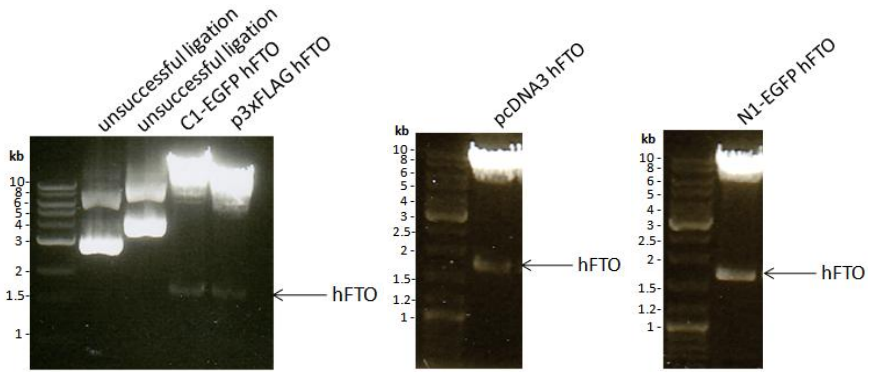


Figure 3-4: 1 % agarose gels of digested hFTO ligation products for verification of the presence of the hFTO insert.

3.2.2 N-*Terminal domain*

As FTO is known to comprise two separate domains, constructs were designed to investigate these individually. The *N*-terminal fragments encompass residues 1-331, containing the full DSBH region, and are referred to as NhFTO. Site directed mutagenesis was used to simultaneously insert a stop codon into the full length sequence at the 332 position, and delete the FTO sequence corresponding to residues 333-505 via a loop deletion reaction (as previously described in Section 2.2.3). This deletion approach prevented accidental readthrough of the stop codon and expression of the remaining FTO protein sequence, ensuring that only the NhFTO domain was expressed.

3.2.3 C-*Terminal domain*

To form the complementary constructs, the *C*-terminal domain was also cloned from the full length sequence using a site directed mutagenesis deletion strategy. Initially the full *N*-terminal region was deleted by this method, with a loop deletion removing residues 1-331, giving a protein composed of residues 332-505. This construct is analogous to the ChFTO construct prepared *in vitro* by recombinant expression.

The assigned nuclear localisation signal (NLS) for FTO is found in residues 2-18 of the hFTO sequence[166], so deletion of this sequence is likely to affect the localisation of the truncated protein. To attempt to localise the *C*-terminal domain protein in the nucleus, where it would likely be able to interact with its native partners, an alternate construct was produced from the full length sequence by deletion of residues 21-331. This

construct contains the hFTO-NLS in addition to the *C*-terminal domain, and was produced by site-directed loop deletion mutagenesis. Figure 3-5 gives a summary of the constructs produced.

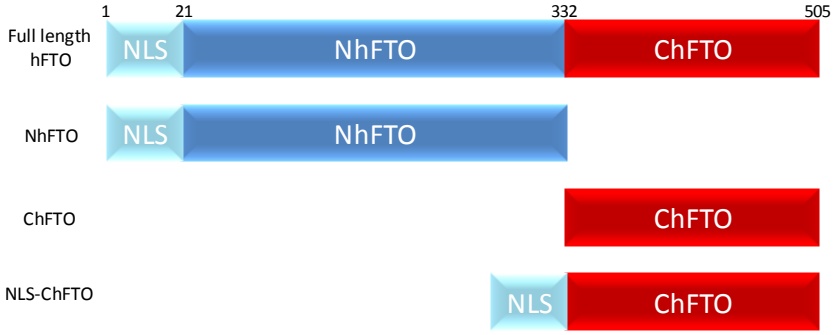


Figure 3-5: Summary of truncated hFTO constructs produced.

3.2.4 hFTO point mutants

For further investigation of the role of FTO, point mutants were also investigated. To study the effect of the I367F point mutation previously studied in mice[242], (discussed in Chapter 2), the corresponding I370F mutation was made in each of the constructs containing residue 370 (i.e. not NhFTO constructs), using site-directed mutagenesis.

Previous *in vitro* studies have identified that the conserved 2OG and Fe(II) binding residues are important for FTO catalytic activity. Substitution of the second iron binding histidine with alanine (mFTO H304A) significantly reduces 2OG turnover activity[165]. In contrast, substitution of the 2OG binding mFTO R313 residue for alanine completely ablates 2OG turnover[165]. In humans a mutation coding for substitution of this residue by glutamine (R316Q) leads to postnatal growth retardation and early lethality[248]. Protein with this alteration is still localised to the nucleus, but has no 2OG turnover *in vitro*[248]. To investigate the effect of overexpression of these inactive or reduced activity mutants in cells, the corresponding point mutations H307A and R316A were produced in the C1 EGFP, N1 EGFP and 3xFLAG hFTO constructs, using site-directed mutagenesis.

3.2.5 Cell lines

In vivo investigations of hFTO were carried out using the HeLa and HEK 293T *H. sapiens* epithelial cell lines. The HEK 293T cell line is derived from human embryonic kidney cells and additionally contains the SV40 Large T-antigen[306], allowing

episomal replication of plasmids containing the SV40 origin of replication. Therefore transfected plasmids can be amplified allowing extended protein expression. The immortal HeLa cell line is derived from a rapidly proliferating cervical cancer[307]. Both cell lines were used here for cell imaging and protein overexpression.

3.3 hFTO localisation studies

To identify the subcellular localisations of the constructs, and to verify that variant constructs are found in the appropriate region of the cell to allow native interactions of FTO with other species, cells expressing the constructs were studied by microscopy. Cells were grown on glass coverslips, then transfected if required and grown for 24 h post-transfection, to allow a complete cell cycle during which protein expression could occur.

3.3.1 Subcellular localisation of EGFP constructs

GFP fusion proteins present in cells can be visualised directly through the fluorescence of the GFP tag. However, the large size of the GFP protein may cause interference with the endogenous localisation of the protein of interest, either by altering its folding, or blocking of localisation target signals. To account for this it was considered important to use both the *N*- and *C*-terminally GFP tagged constructs.

Cells were stained with DAPI, to show the location of the nucleus. As DAPI (which appears blue in the images obtained) has fluorescent properties orthogonal to those of GPF the two species can be viewed in different channels without interference.

As control samples, HeLa cells were transiently transfected with either the empty C1-EGFP vector, or with a C1-EGFP NLS vector, which targets EGFP to the nucleus via addition of a nuclear localisation sequence (NLS) to the *N*-terminus of the EGFP sequence (C1-EGFP NLS kindly provided by the Heinrich Leonhardt laboratory, LMU Munich). Imaging of these cells (Figure 3-6) reveals that untargeted EGFP is found diffusely throughout the cells, whereas NLS-EGFP is fully targeted to the nucleus.

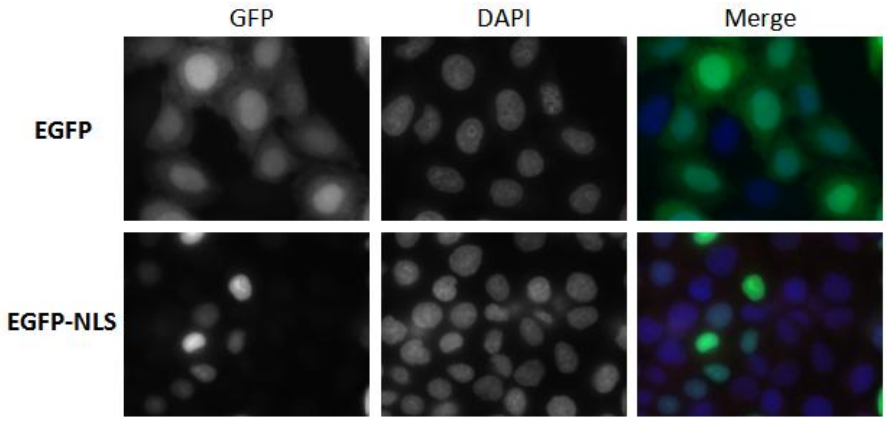


Figure 3-6: Localisation of EGFP constructs in HeLa cells, detected by direct fluorescence of GFP.

Samples were prepared in the same manner to investigate HeLa cells transfected with full length hFTO tagged with GFP at either the *N*- or *C*-termini. Localisation images show that the GFP-tagged proteins are both correctly localised in the nucleus (Figure 3-7), as has been observed in previous studies for tagged and endogenous FTO constructs[165,246,305].

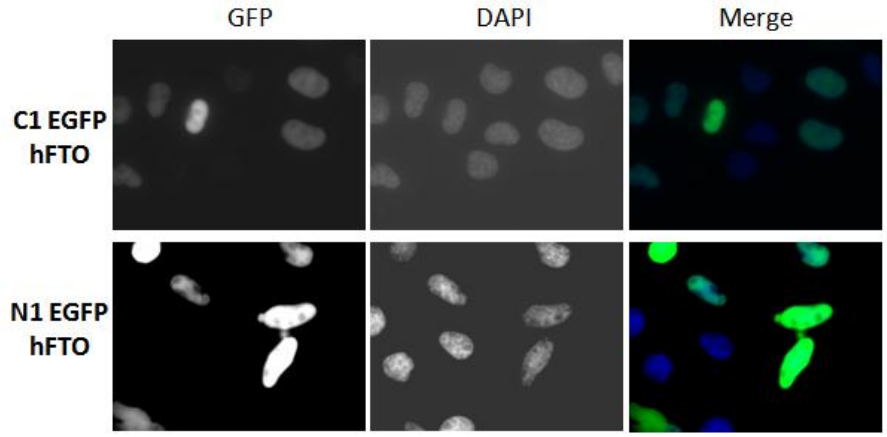


Figure 3-7: Localisation of N1 and C1 EGFP-tagged hFTO constructs in HeLa cells, detected by direct fluorescence of GFP.

Similarly to the observation for the mouse I367F protein[242], the I370F hFTO constructs seem to be transfected with a lower efficiency than for the corresponding wildtype constructs (summary of transfection efficiencies in Table 3-2), although nuclear localisation is still observed (Figure 3-8). The EGFP-C1 I370F hFTO construct is found partially in the cytoplasm in addition to the nucleus.

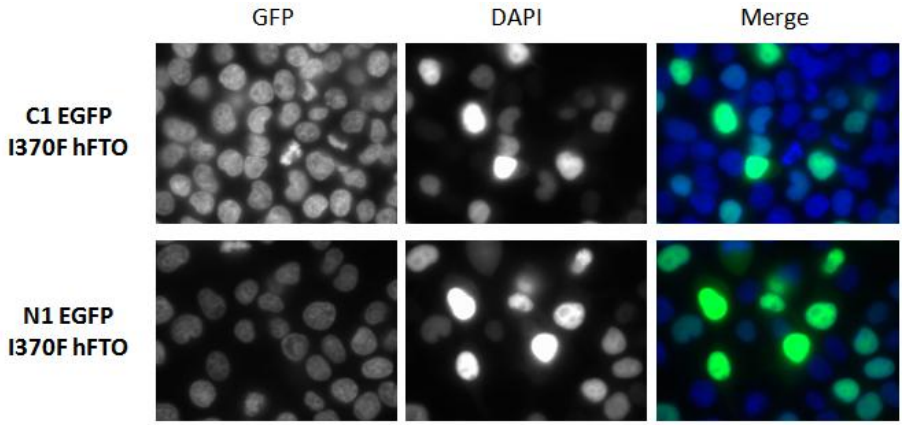


Figure 3-8: Localisation of EGFP-tagged hFTO I370F point mutant constructs in 293T cells.

3.3.2 Subcellular localisation of 3xFLAG hFTO constructs

To view the subcellular location of transfected FLAG-tagged constructs, antibody staining is required using mouse anti-FLAG antibodies. Following formaldehyde fixation, cells are permeabilised by treatment with detergent (Triton X100) to facilitate entry of large antibody molecules into the cells. The anti-mouse secondary antibody is conjugated to Cy3, a fluorescent cyanine dye. The fluorescence characteristics of Cy3 are orthogonal to those of both GFP and DAPI, and Cy3 appears red in microscopy images. Any background staining, and therefore the specificity of the secondary antibody for the primary antibody, can be determined by staining cells with only the secondary antibody. No signal was observed in the Cy3 channel for cells stained using only secondary antibody showing that, under the exposure conditions used, no background staining by the anti-mouse antibody was observed, and that there is no cross-channel overlap (Figure 3-9, row 3).

FLAG-hFTO showed complete localisation in the nucleus (Figure 3-9), as previously observed for the GFP constructs. This therefore suggests that the addition of the tags has not affected folding or subcellular targeting of the wildtype proteins, so native interactions and reactions are likely to be maintained under the overexpression conditions. As was observed with C1-EGFP constructs, the FLAG I370F hFTO construct is found in both the nucleus and the cytoplasm, suggesting that the presence of

an *N*-terminal tag adjacent to the hFTO NLS in addition to the introduction of the I370F substitution in the protein sequence reduces the specificity of the nuclear localisation.

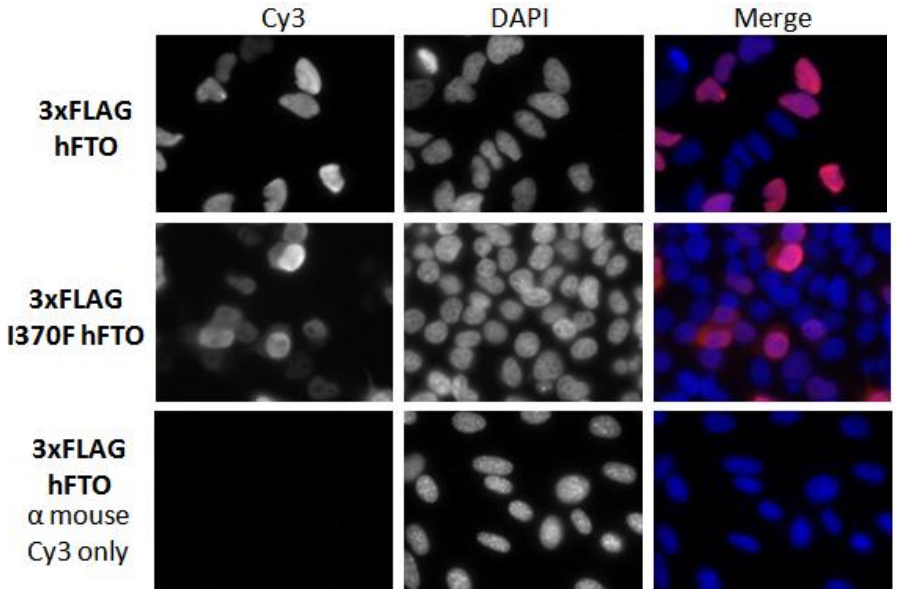


Figure 3-9: Localisation of 3xFLAG-tagged hFTO constructs in 293T cells detected using anti-FLAG antibody. Control probed using only secondary anti-mouse Cy3 antibody.

3.3.3 Localisation of active site FTO mutants

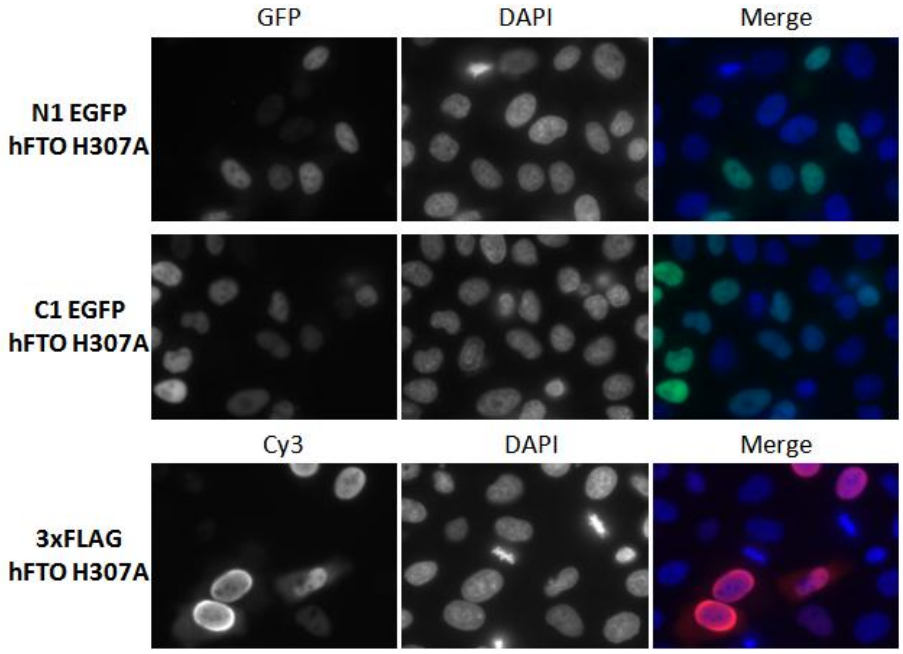


Figure 3-10: Localisation of H307A hFTO constructs in HeLa cells detected either by direct visualisation of GFP or by anti-FLAG antibody staining.

GFP tagged H307A hFTO was found only in the cell nucleus (Figure 3-10). Although most cells transfected with 3xFLAG H307A hFTO have signal confined to the nucleus,

in a few cells (~ 25 %) FLAG-tagged material is also found through the cytoplasm, although this may be an artefact of the induced protein overexpression. In common with the previously reported R316Q mutant[248], the R316A constructs were all found in the nucleus (Figure 3-11). This therefore suggests that despite these substitutions the proteins are correctly targeted to the correct location within the cell, suggesting that the reduced activity is likely mediated by prevention of co-substrate binding.

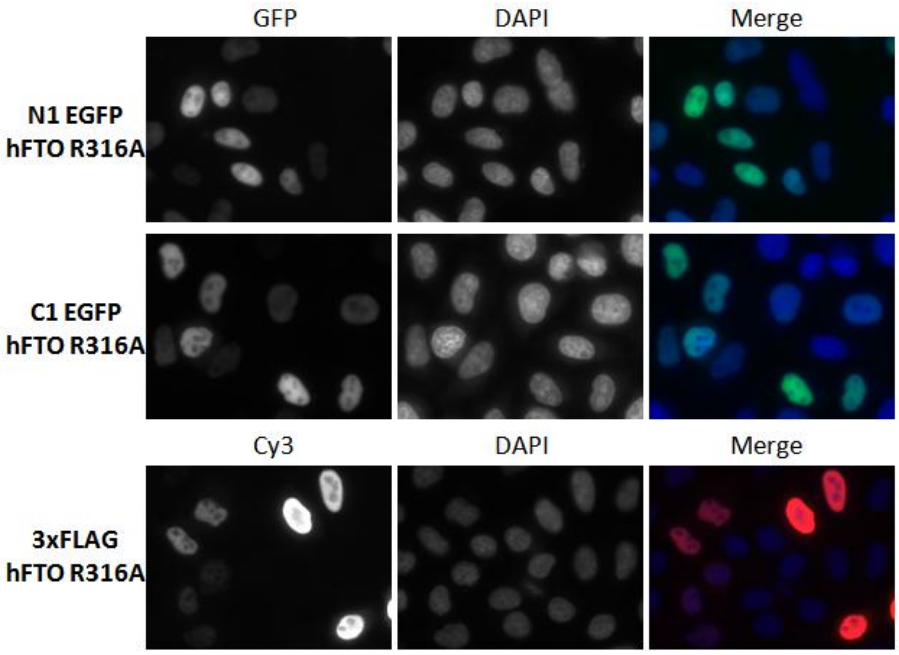


Figure 3-11: Localisation of R316A hFTO constructs in HeLa cells detected either by direct visualisation of GFP or by anti-FLAG antibody staining.

3.3.4 Detection of endogenous FTO

Direct detection of the endogenous FTO protein is also possible using a primary antibody specific to the protein of interest. Here a Eurogentec FTO antibody raised to recombinant mFTO protein was used. As for the detection of FLAG hFTO described previously, a secondary (anti-rabbit) antibody conjugated to Cy3 was used for visualisation by fluorescence microscopy. A control sample of cells stained with only the α-rabbit Cy3 coupled secondary antibody revealed no background staining (Figure 3-12, row 2). Staining 293T cells using the Eurogentec FTO antibody at a dilution of 1 in 1000 was successful, and the recognised protein (hFTO) can be seen to be expressed in all cells, localised entirely in the nucleus (Figure 3-12).

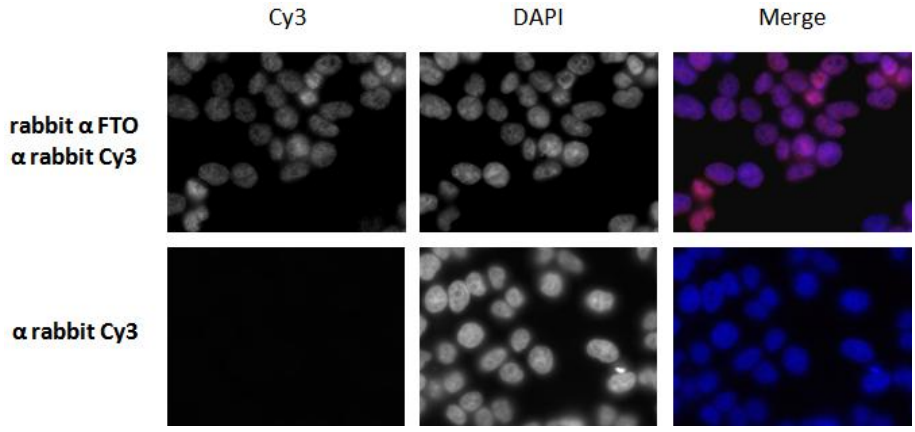


Figure 3-12: Localisation of endogenous hFTO in 293T cells detected using anti-FTO antibody staining. Control samples (row 2) probed only with anti-rabbit secondary antibody.

Probing cells using an anti-FTO antibody also allows detection of protein overexpressed from the pcDNA3 vector, proteins expressed from which are untagged. Again, all cells were observed to express hFTO, although due to the overexpression of protein caused by transfection a wide range of intensities were observed. This allows comparison between the levels of protein produced endogenously and on overexpression.

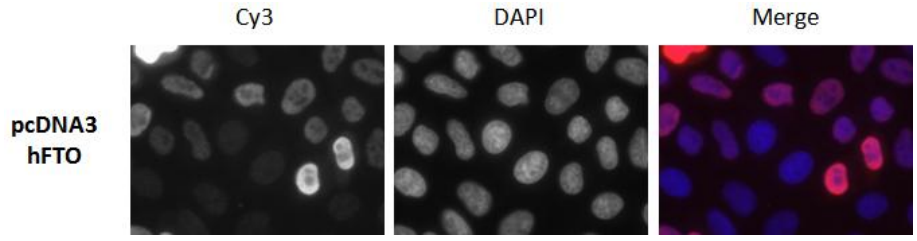


Figure 3-13: Localisation of pcDNA3 hFTO in HeLa cells detected using anti-FTO antibody staining.

3.3.5 C-*NLS and* N-*terminal FTO constructs*

To enable analysis of the different domains of FTO, the localisation of the different constructs was investigated. As these constructs involve large deletions of protein sequence it is possible that they may be incorrectly folded in the cell, possibly causing aggregation or a loss of nuclear targeting.

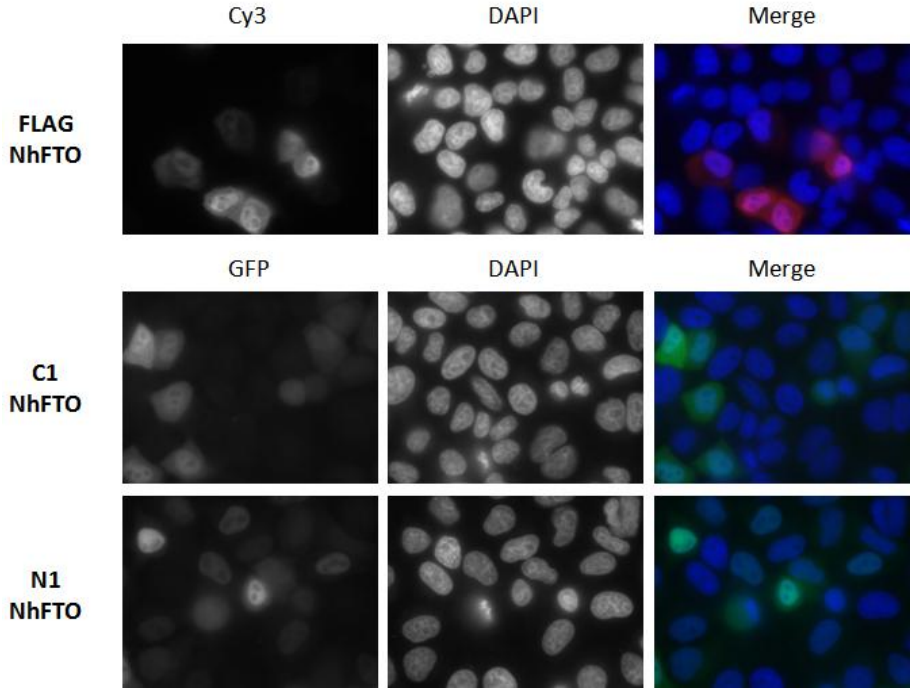


Figure 3-14: Localisation of NhFTO constructs in HeLa cells detected either using anti-FLAG antibody staining, or by direct visualisation of GFP.

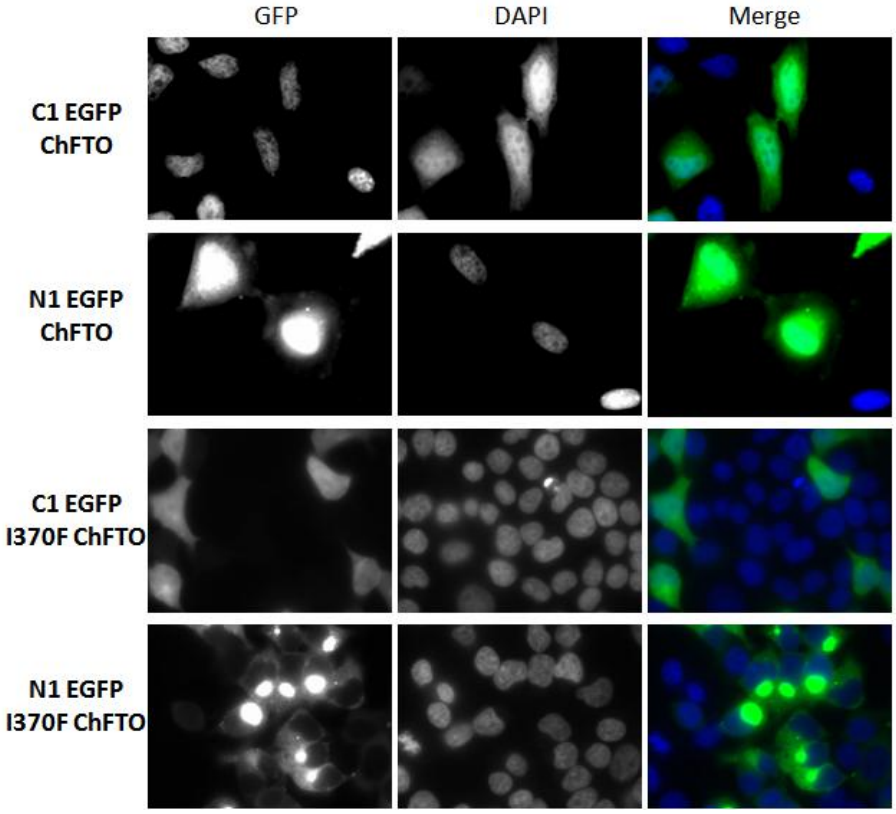


Figure 3-15: Localisation of N1 and C1 EGFP-tagged ChFTO constructs in HeLa cells, detected by direct visualisation of GFP fluorescence.

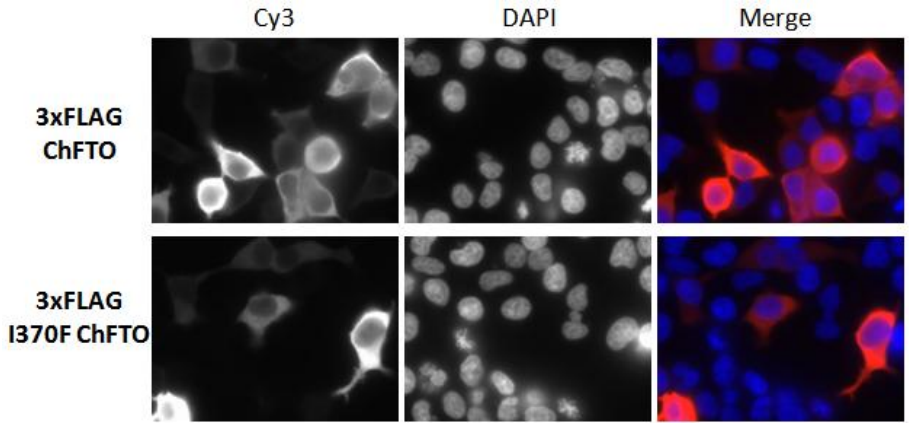


Figure 3-16: Localisation of FLAG-tagged I370F ChFTO constructs in HeLa cells, detected by anti-FLAG antibody staining.

Expression of tagged constructs encoding the separate *N*-terminal domain (NhFTO) in HeLa cells showed that in each case the protein is present both in the nuclei and, to a lesser extent, dispersed throughout the cytoplasm (Figure 3-14).

Transfection of the corresponding tagged ChFTO constructs showed the presence of GFP throughout the cell (Figure 3-15), therefore not all of the protein has been correctly targeted to the nucleus, likely due to deletion of the *N*-terminal NLS. Both I370F N1-EGFP hFTO, and particularly ChFTO, may aggregate on overexpression, as seen by very bright signals throughout the cytoplasm, suggesting protein accumulation.

FLAG-tagged ChFTO constructs (Figure 3-16) are also found throughout the cells, but in this case relatively little FLAG staining is visible in the nuclei, suggesting that these constructs are either excluded from the nuclei, or are not correctly targeted. Therefore it was hoped that inclusion of the NLS sequence at the *N*-terminus of the ChFTO sequence (NLS-ChFTO constructs) would allow correct subcellular targeting, as is seen for the full length protein.

Transfection of each of the NLS-ChFTO constructs in HeLa cells (Figure 3-17) showed nuclear targeting. However, additional disperse expression is also apparent through the cytoplasm, although generally at a lower level than is observed in the nucleus. It appears that the inclusion of the NLS in these constructs is an improvement on the initially designed constructs, but it has not produced complete targeting to the nucleus, as is seen with the full length constructs. This suggests that there may be additional factors

involved in the nuclear targeting. It is possible that the addition of the NLS to the *N*-terminus of ChFTO prevents correct folding of either, or both, of the portions of the fused proteins, thereby preventing correct localisation.

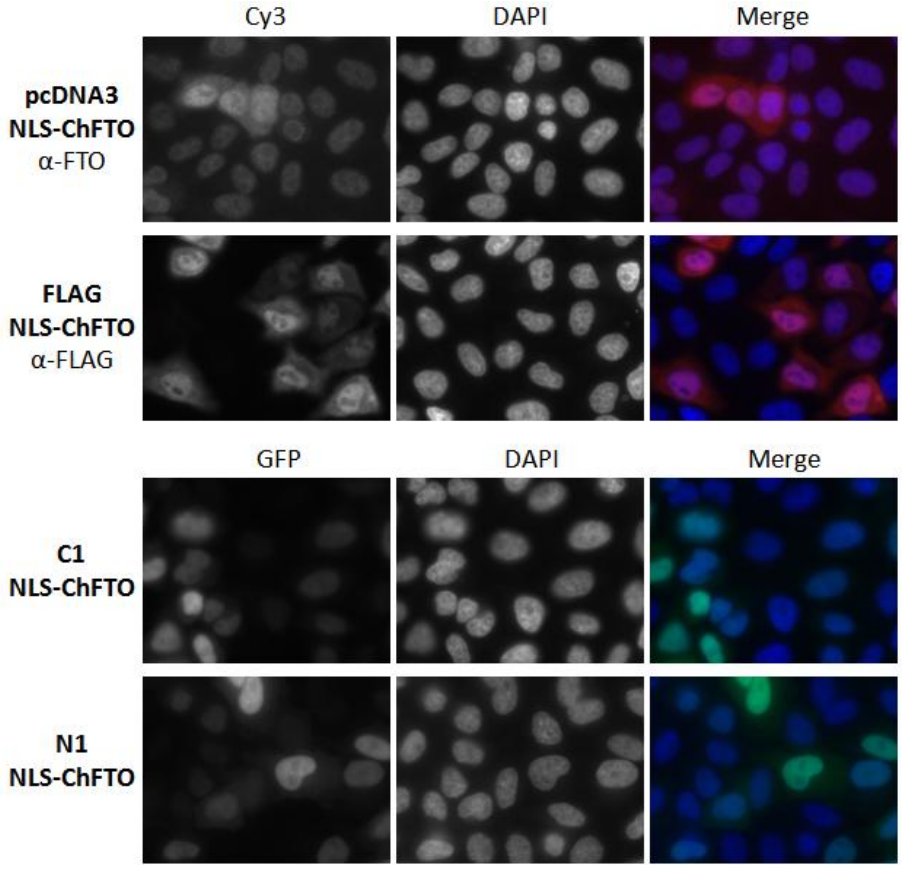


Figure 3-17: Localisation of NLS-ChFTO constructs in HeLa cells detected either using anti-FTO or anti-FLAG antibody staining as indicated, or by direct visualisation of GFP.

Transfection of NLS-ChFTO I370F constructs in HeLa cells showed similar results, with protein observed dispersed throughout the cell (Figure 3-18), however transfection and expression of pcDNA3 I370F NLS-ChFTO was very low (Table 3-2), and led to nuclear aggregation of the hFTO protein.

Based on the current literature understanding of the cellular localisation of FTO, it would be assumed that all constructs containing residues 2-18, the putative NLS[166], will localise solely in the nucleus. However it appears here that some of the constructs containing this NLS sequence are not fully confined in the nucleus. Therefore the NLS appears to be insufficient in directing the protein to the nucleus, suggesting that this region may not be the only factor in causing localisation. The constructs with the

greatest expression away from the nucleus are highly mutated, for example I370F ChFTO, which contains both a point mutation and a truncation. In particular, fusion of the NLS to the *C*-terminal domain of FTO may produce an unstable protein not capable of correct folding. These alterations may therefore cause too much of an change to the three-dimensional structure of the proteins, meaning that correct processing to the nucleus may not be successful in all cases.

As a summary, Table 3-1 details the localisation pattern for each of the hFTO constructs studied here, and Table 3-2 shows transfection efficiencies for each protein construct, obtained by counting the proportion of cells transfected (~5 images in each case).

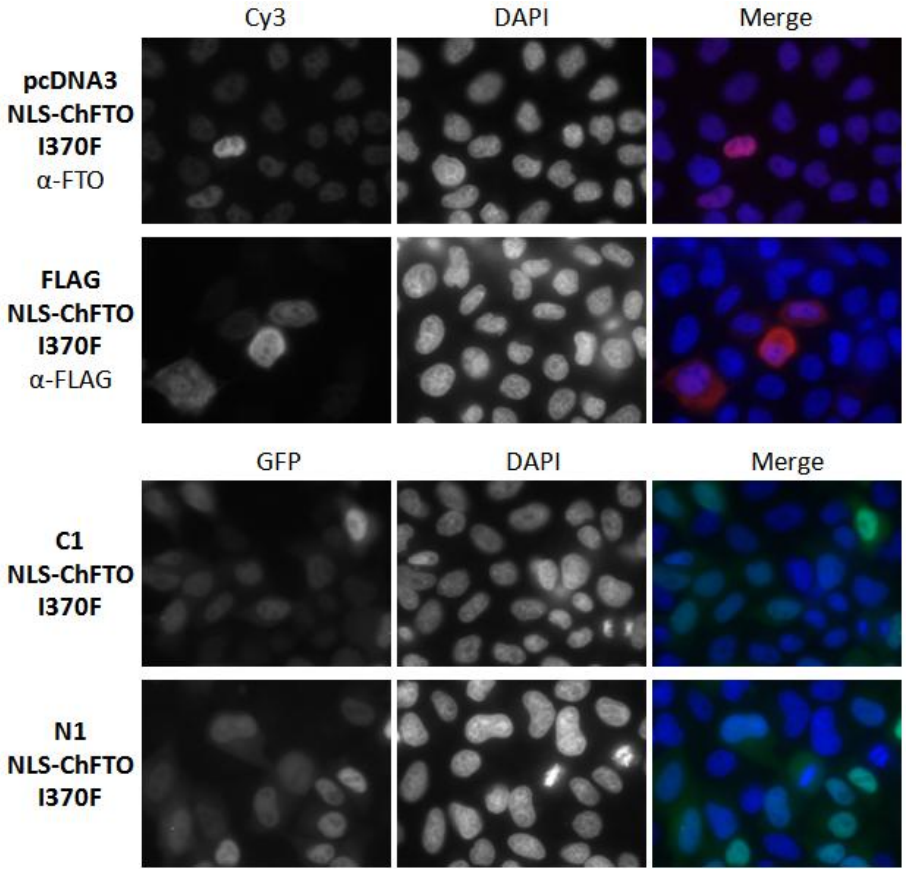


Figure 3-18: Localisation of NLS-ChFTO I370F constructs in HeLa cells detected either using anti-FTO or anti-FLAG antibody staining as indicated, or by direct visualisation of GFP.

Construct	C1 EGFP	N1 EGFP	FLAG (Cy3)	pcDNA3
hFTO	Nuclear only	Nuclear only	Nuclear only	Nuclear only
I370F hFTO	Mainly nuclear but some cytoplasmic	Nuclear only	Mainly nuclear but some cytoplasmic	-
H307A hFTO	Nuclear only	Nuclear only	Mainly nuclear but some cytoplasmic	-
R316A hFTO	Nuclear only	Nuclear only	Nuclear only	-
NhFTO	Nuclear and cytoplasmic	Nuclear and cytoplasmic	Nuclear and cytoplasmic	-
ChFTO	Nuclear and cytoplasmic	Nuclear and cytoplasmic but aggregation	Cytoplasmic only	-
NLS-ChFTO	Mainly nuclear but some cytoplasmic	Mainly nuclear but some cytoplasmic	Nuclear and cytoplasmic	Nuclear and cytoplasmic
I370F ChFTO	Nuclear and cytoplasmic	Cytoplasmic only but aggregation	Cytoplasmic only	-
I370F NLS-ChFTO	Nuclear and cytoplasmic	Mainly nuclear but some cytoplasmic	Nuclear and cytoplasmic	Nuclear aggregates

Table 3-1: Summary of subcellular localisation of hFTO constructs. Cell line used: Green: HeLa; Red: 293T; Blue: both HeLa and 293T.

Construct	% cells transfected			
	C1 EGFP	N1 EGFP	FLAG (Cy3)	pcDNA3
hFTO	60/34	73	56	14
I370F hFTO	39	49	32	-
H307A hFTO	60	69	48	-
R316A hFTO	56	53	44	-
NhFTO	31	40	29	-
ChFTO	65	73	31	-
NLS-ChFTO	43	38	23	30
I370F ChFTO	22	38	18	
I370F NLS-ChFTO	26	32	13	5

Table 3-2: Summary of approximate transfection efficiencies of hFTO constructs. Cell line used: Green: HeLa; Red 293T.

3.4 Cell lysis optimisation

The first step in investigating cellular interacting partners for FTO is to lyse the cells. As hFTO is a nuclear protein, it is important to optimise protein extraction from cells to enrich for nuclear proteins. In proteomic analyses the identification of proteins which are true FTO interactors would be simplified if the proteins present in the cell lysate were derived predominantly from the nucleus. As each protein has unique interactions with other cellular components, lysis conditions need to be optimised for individual proteins.

Cell lysis conditions can be optimised through variation of salt, detergent and buffer. Furthermore, addition of nucleases or sonication of samples may be employed to shear nucleic acids and aid soluble protein extraction, particularly for proteins interacting with DNA. In other cases, mechanical cell lysis may be employed, utilising dounce homogenisers or glass beads. Generally, lysis and subcellular fractionation methods involve initial lysis of the cell membrane through resuspension of cells in a lysis buffer followed by incubation on ice. Intact nuclei obtained by gentle pelleting may be lysed by addition of further buffer, and pelleting at high speed. The initial supernatant contains the cytoplasmic protein fraction, with the final supernatant containing the nuclear component. Any material in the pellet is classed as insoluble under the lysis conditions employed.

3.4.1 Initial lysis attempts

A selection of cell lysis methods was tested here to investigate optimal release of hFTO from nuclei, in a "nuclear" fraction. Lysis buffers of different stringencies were selected for the general lysis method described above, using two salt and detergent concentrations (Method A: 300 mM NaCl, 0.5 % NP40; Method B: 500 mM NaCl, 1 % NP40). A paper by Gilljam et al. reporting the interaction of ABH2 with PCNA used a variation of this method but with a lower salt concentration and the inclusion of glycerol, DNase and RNase[168], therefore this was tested (referred to as the "Gilljam" method), as it had been successfully used for a related nuclear protein. An alternate nuclear extraction protocol (developed by the Klose laboratory, Biochemistry Department, University of Oxford) was also tested, which involved manual cell lysis using a dounce homogeniser. This method additionally involved an initial "cell swelling" step, with the pellet initially resuspended in hypotonic buffer. To test these lysis methods, 293T cells transfected with hFTO constructs were harvested 24 h post-transfection. Following cell lysis and subcellular fractionation, the hFTO present in each fraction was detected by western blot (Figure 3-19).

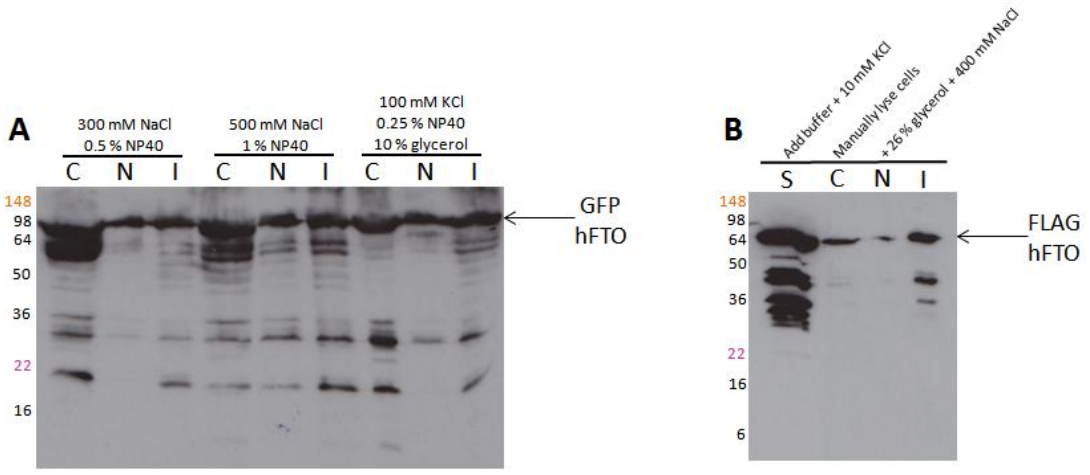


Figure 3-19: (A) 9.5 x 10^5 293T cells transfected with N1-EGFP hFTO. Cells were harvested, frozen at -80 °C then lysed using three different methods, as indicated. C: cytoplasmic fraction; N: nuclear fraction; I: insoluble fraction. 30 s exposure. (B) 9.5 x 10^5 293T cells transfected with 3xFLAG hFTO. Harvested cells were lysed using a dounce homogenisation method. S: cell swelling; C: cytoplasmic fraction; N: nuclear fraction; I: insoluble fraction. 1 min exposure.

Western blot detection with either the anti-GFP (Figure 3-19A) or anti-FLAG (Figure 3-19B) antibodies was successful, with distinct bands observed at the expected masses of the fusion proteins. Protein degradation is observed for each of the samples, with many bands visible at reduced masses, and particularly at the expected mass of the discrete GFP protein (~30 kDa, Figure 3-19A). Degradation of the purified recombinant FTO proteins has also been observed previously (Chapter 2), and has also been observed by collaborators working on both mouse and human FTO (personal communication, Dr Chris Church, James McTaggart). Degradation bands are therefore likely to be a feature of western blots detecting FTO proteins.

Although immunofluorescence imaging of full length hFTO constructs indicates that all observed hFTO is present in the nuclei (Figure 3-6, Figure 3-9), a large proportion of the protein extracted was found in the first protein fraction (constituting either the "cytoplasmic" or "cell swelling" fractions), irrespective of the method or buffer conditions used. Typically the transfected cells were harvested then the cell pellets immediately frozen at -80 °C. These cell pellets were then lysed immediately after defrosting of the pellet. However, it is possible that the early release of FTO from the cells may be due to the process of freezing and thawing the cells, possibly compromising the integrity of the cell membranes. Furthermore, the mechanical

scraping method used to economically harvest cells from the surface of the plate could cause damage to the cell membrane, causing lysis, therefore a potentially gentler pipette resuspension method was attempted. Two batches of 293T cells were freshly grown and harvested either using the standard mechanical scraping method, or using a pipette to gently resuspend the cells in phosphate buffered saline (PBS). Following pelleting of the cells, the dounce homogenisation lysis method was immediately attempted (Figure 3-20). In both of these cases a similar pattern of protein extraction was observed to that from the previously frozen cell pellets, suggesting that the freeze-thaw cycle does not have a detrimental effect on cell lysis. An increased proportion of FTO protein was present in the "cell swelling" fraction of the pipette resuspension method, suggesting this is less successful as a gentle cell harvesting method.

Comparing the proportions of protein present in each fraction for the lysis conditions tested, the greatest proportion of protein present in the nuclear fraction appears to occur using the lysis conditions previously reported for ABH2 immunoprecipitation (Figure 3-19, lanes 6-9).

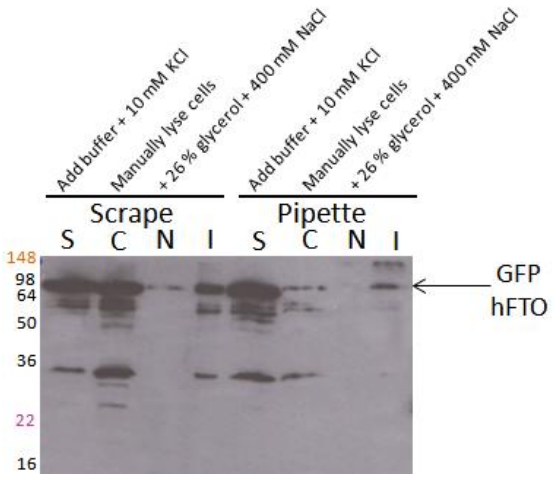


Figure 3-20: 9.5 x 10^5 293T cells transfected with N1-GFP hFTO and lysed using a dounce homogenisation method immediately following cell harvesting. Harvesting was carried out either using a rubber scraper or by pipette resuspension. S: Cell swelling; C: cytoplasmic fraction; N: nuclear fraction; I: insoluble fraction. 30 min exposure.

3.4.2 HeLa cell lysis attempts

To further test the lysis conditions, additional cell lysis attempts were carried out for HeLa cells transfected with 3xFLAG hFTO. Similar methods were attempted in this

test, but rather than using the most stringent buffer condition tested previously (500 mM NaCl, 1 % NP40) a lower stringency method was used to attempt to prevent immediate release of FTO from the cells.

Samples obtained were probed using three different primary antibodies. Probing with anti-FLAG (Figure 3-21A), allowed detection of hFTO in the lysates. β-actin was also detected, as this is highly abundant in the cytoplasm[308], therefore enables tracking of "cytoplasmic" proteins (Figure 3-21B). Finally, detection of Histone H3 was attempted (Figure 3-21C). Due to their basic nature, successful and efficient extraction of soluble histones generally requires treatment with acid or high salt[309]. However, the detection of histones as a nuclear marker is still commonly used, and can be a useful indication of the effectiveness of the nuclear extraction method.

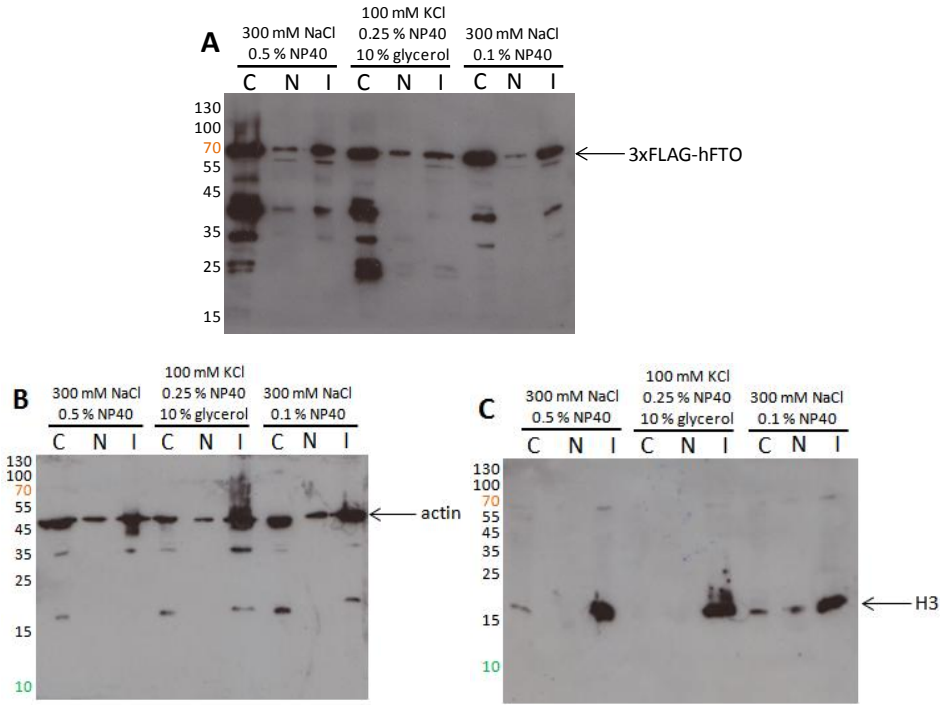


Figure 3-21: Test of HeLa cell lysis 5.5 x 10^5 HeLa cells transfected with 3xFLAG-hFTO. Cells were harvested and lysed using three different methods, as indicated. Western blot detection carried out by probing with either α-FLAG, α-actin or α-Histone H3 antibodies. C: cytoplasmic fraction; N: nuclear fraction; I: insoluble fraction. 30 min exposures.

No significant differences are observed between lysis of 293T and HeLa cells and, as observed previously (Section 3.4.1), the majority of the hFTO produced by HeLa cells is present in the initial "cytoplasmic" lysate fraction (Figure 3-21A), with the greatest

proportion of hFTO in the "nuclear" fraction observed with the Gilljam method (Figure 3-21, lanes 4-6). In each case β-actin is present in all fractions (Figure 3-21B), although predominantly in the insoluble fraction rather than the expected "cytoplasmic" fraction. As anticipated, Histone H3 is predominantly found in the insoluble fractions for each of the extraction conditions tested, although a small amount of H3 is found in the cytoplasmic and nuclear fractions for the mildest lysis condition tested. Unfortunately it appears that the methods tested here have not been efficient in either fractionating subcellular components, as observed by blotting with anti-β-actin and anti-H3 antibodies, or able to produce fractions enriched for FTO.

3.5 Optimisation of immunoprecipitation conditions

Immunoprecipitation (IP) involves precipitation of an analyte out of solution using an antibody specifically binding the analyte. For successful precipitation of the antibody-analyte complex a solid support is normally used, typically agarose beads (which may be separated from the mixture by centrifugation), or magnetic beads, separated by application of a magnetic field to the solution. These beads may be either prebound to antibodies, or may be conjugated to protein A/G to which the antibody used can bind during the course of the IP reaction.

Successful IP requires a strong, robust antibody-analyte interaction which can survive the IP reaction conditions. Commercially available antibodies to well characterised fusion tags (such as GFP and FLAG) have been developed to withstand such conditions, but problems may arise with non-standard antibodies, so it is necessary to carry out tests to determine whether the interaction is sufficient. Additionally, successful IP requires specific recognition of the protein in its native, folded environment. Although detection of proteins may be successful in western blots where the proteins are denatured and unfolded, the detection of folded proteins, which have conformational epitopes, may not be possible[310].

In these studies the aim of IP reactions was the co-immunoprecipitation of FTO with any interaction partners. Although FTO is a nucleic acid demethylase, interactions with

other proteins *in vivo* may assist this role. FTO may exist as part of a complex, for example enabling location of the sites of 3meT or 3meU modifications. However, the FTO gene[228] and protein[243] have well established links to obesity, therefore it is possible that the protein may play additional roles within the cell. Identification of interaction partners would assist identification of the involvement of FTO in other cellular processes, which may explain the role of FTO in obesity. Previous microarray gene expression analysis showed upregulation of some carbohydrate catabolism and fat synthesis genes, and downregulation of inflammation genes in homozygous I367F mice[242]. The mechanism through which this is achieved on alteration of the FTO sequence is currently unknown, but analysis of species interacting with FTO may shed light on how this occurs. A related application of IP is Chromatin Immunoprecipitation (ChIP), in which the protein of interest is crosslinked to interacting nucleic acids. The sequence of the interacting DNA/RNA may be determined, allowing identification of motifs with which the protein may interact or, for the case of transcription factors, particular promoters in the genome, altering gene expression.

3.5.1 Anti-GFP immunoprecipitation test

Initially, immunoprecipitation was attempted using GFP-tagged proteins. As GFP is a large protein, the presence of the additional domain attached to the terminus of the protein of interest may prevent correct folding, or may shield interactions with normal interaction partners at a particular region of the protein surface. It is therefore important to study constructs which code for GFP at either end of the protein sequence, to maximise the possibility that binding sites for all interaction partners are available.

To determine whether the optimal cell lysis conditions determined previously are compatible with the protein-antibody interaction, and whether hFTO is efficiently immunoprecipitated, a small scale test was carried out. 293T cells transiently tranfected with N1-EGFP hFTO were lysed using the previously described Gilljam protocol, because this method appeared to yield the greatest proportion of hFTO in the nuclear fraction. The nuclear fraction obtained from this method was used as the IP input, and

was added to GFP-Trap beads (Chromotek). Following incubation of the bead-lysate mixture, the beads were washed extensively, then bound proteins were eluted by boiling the beads in SDS-PAGE loading buffer. The IP reaction is followed throughout by removal of protein samples at each stage, which are analysed by western blotting (Figure 3-22A). The advantage of using the commercial GFP-Trap beads is that they are based on camel antibody technology[311] therefore bands corresponding to the heavy or light chain antibody protein do not appear on blots, thereby simplifying analysis.

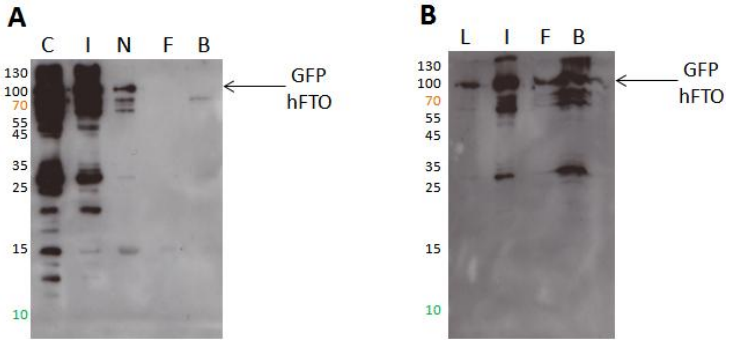


Figure 3-22: Small scale test of cell lysis and immunoprecipitation conditions. 9.5 x 10^5 293T cells transfected with N1-GFP hFTO C: cytoplasmic fraction; I: Insoluble fraction; N: Nuclear fraction; F: Bead flow through; B: Beads; L: Whole cell lysate. (A) Gilljam lysis method; Nuclear fraction loaded onto beads. (B) Cell lysis buffer 300 mM NaCl, 0.5 % NP40, 10 mM Tris pH 7.5; Whole cell lysate loaded onto beads. 30 min exposures.

Unfortunately, this immunoprecipitation reaction was not very successful. Although a good amount of hFTO protein can be seen to be produced by the cells (Figure 3-22), the proportion of hFTO present in the cytoplasmic and insoluble fractions is significantly greater than that in the nuclear fraction obtained from the cell lysis. Therefore, only a small amount of GFP tagged protein was available in the nuclear fraction for immunoprecipitation, and this does not appear to have been successfully immunoprecipitated as there is no protein at the mass expected for hFTO-GFP in the "bead" lane of Figure 3-22A. The faint band at ~80 kDa in this lane may be due to degradation of immunoprecipitated hFTO-GFP.

To attempt to improve the pull down efficiency, another immunoprecipitation reaction was carried out, using a whole cell lysate. In this situation more protein should be available to be bound by the beads in the immunoprecipitation reaction. Figure 3-22B shows that under these conditions, although there is still a significant proportion of

protein in the insoluble fraction, the anti-GFP immunoprecipitation has been successful with a good amount of GFP-tagged protein bound to the beads. Therefore this whole cell lysate method can be used for further immunoprecipitation studies, to ensure that plenty of protein is obtained for subsequent analyses.

3.5.2 Anti-FLAG immunoprecipitation test

For complementary experiments, FLAG-tagged hFTO protein was immunoprecipitated using anti-FLAG beads. The 3xFLAG peptide (~2.5 kDa) is significantly smaller than the GFP protein (~27 kDa), so the smaller tag may interact with fewer proteins in solution, simplifying proteomic analysis, and is less likely to obscure interactions of FTO with other proteins.

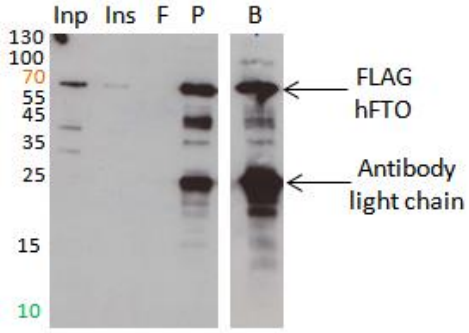


Figure 3-23: Test of immunoprecipitation of hFTO using anti-FLAG beads. 1.5 x 10^7 293T cells transfected with 3xFLAG-hFTO. Inp: IP input; Ins: Insoluble fraction; F: bead flow through; P: protein eluted from beads; B: anti-FLAG beads. 1 min exposure.

To test the anti-FLAG immunoprecipitation reaction 293T cells transiently transfected with 3xFLAG-hFTO were lysed, and the whole cell lysate incubated with anti-FLAG beads. The commercial anti-FLAG beads (Sigma-Aldrich, A2220) used for this IP reaction have a conjugated mouse monoclonal antibody, and in western blots of bead samples bands at ~25 kDa are visible, corresponding to the light chain of the antibody[312], although the heavy chain of the anti-FLAG antibody does not appear to be visible on blots, as no additional bands are observed at the expected mass (~50 kDa).

In this attempt elution of the protein from the beads was tested, by incubation with 1 % SDS (v/v), 100 mM $NaHCO_3$ for 1 h at 65 °C. Treatment with detergent denatures the immobilised anti-FLAG antibody, destroying the interaction with the bound 3xFLAG fusion peptide. As previously, protein samples were analysed by western blot (Figure

3-23). This IP reaction is very efficient, with no FLAG-tagged protein visible in the flow through lane of the western blot. It appears that approximately half of the FLAG-hFTO present has been eluted from the beads, while proportionally less of the antibody light chain has been eluted. However incubation at an elevated temperature may have caused some degradation of the FLAG-tagged hFTO, as the bands at ~40 kDa have increased intensity.

3.5.3 Formaldehyde crosslinking anti-FLAG immunoprecipitation

A further area of interest for immunoprecipitation studies, particularly for a protein known to interact with DNA, is chromatin immunoprecipitation (ChIP). These experiments involve crosslinking of the protein of interest to interacting DNA by addition of formaldehyde to cells prior to harvesting. The covalent bridging formaldehyde links formed allow immunoprecipitation of any bound species along with the targeted protein. Subsequent reversal of the cross links releases the interaction partners for analysis. In addition to the interest in the possible nucleic acid sequences with which FTO may interact it is also possible that the interactions of FTO with other proteins may be mediated by a common interaction with a nucleic acid species. There is precedent for such an interaction in the family of zinc finger proteins[313], which can bind both DNA and protein. Additionally, it has been postulated that the adaptive response protein AidB, the specific function of which is currently unknown, may have indirect interactions with other protein species, through simultaneous binding to RNA[150].

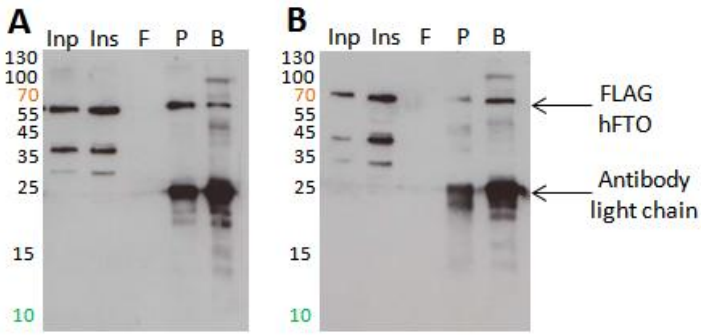


Figure 3-24: Test of anti-FLAG immunoprecipitation. 1.5x10^7 293T cells transfected with 3xFLAG-hFTO. (A) Cells crosslinked with 1 % (v/v) HCHO before cell harvesting; (B) cells crosslinked with 1 % (v/v) HCHO before cell harvesting then crosslinks reversed following IP reaction. Inp: IP input; Ins: Insoluble fraction; F: Bead flow through; P: Protein eluted from beads; B: anti-FLAG beads. 1 min exposures.

Prior to harvesting cells, crosslinking was initiated by addition of concentrated formaldehyde, then was quenched by addition of 125 mM glycine, and the IP reaction carried out as previously (Figure 3-24). Following elution of protein from the beads, the supernatant from one of the IP reactions was incubated overnight at 65 °C to reverse the formaldehyde crosslinks (Figure 3-24B). Despite the likely modification of the anti-FLAG epitopes by reaction with formaldehyde the immunoprecipitation is still successful, suggesting that the anti-FLAG antibody is robust and capable of recognising the FLAG peptide even when modified. Compared to the previous FLAG IP test (Figure 3-23), more FLAG-hFTO is present in the insoluble protein fraction after treatment with formaldehyde, probably due to protein aggregation on crosslinking, so less protein remains for binding to the beads. Following the overnight reversal of crosslinking, less protein eluted from the beads is visible on the blot (Figure 3-24B), possibly due to degradation or aggregation of some of the protein during the extended incubation at 65 °C, therefore this was not pursued further.

In addition to western blot analysis, small aliquots of eluted protein from the IP reactions were run on an agarose gel for analysis of any nucleic acid present, which may have been immunoprecipitated alongside hFTO, but no distinct DNA signal was apparent. However, an important step in ChIP experiments is shearing of DNA, usually achieved by sonication, into short fragments appropriate for further analysis. If hFTO is crosslinked to large fragments of genomic DNA, or other high molecular weight species, the resulting complex may be insoluble, and unavailable for involvement in the immunoprecipitation reaction. Sonication should shear the DNA and potentially reduce the proportion of insoluble protein present.

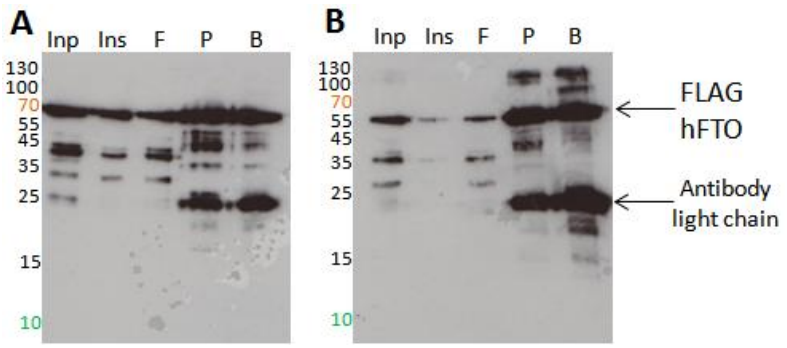


Figure 3-25: Test of anti-FLAG immunoprecipitation. 1.5x10^7 293T cells transfected with 3xFLAG-hFTO. (A) Cell lysis including sonication, standard IP conditions; (B) cells crosslinked with HCHO before harvesting. Inp: IP input; Ins: Insoluble fraction; F: Bead flow through; P: Protein eluted from beads B: Beads. 1 min exposures.

Further batches of 293T cells lysed by sonication on ice were then analysed as previously. The anti-FLAG IP reaction was again successful, with large quantities of FLAG-hFTO bound to the beads (Figure 3-25). Interestingly, species at increased mass are present in the eluted protein and bead lanes for the cells crosslinked with formaldehyde, whereas these are not seen when not treated with formaldehyde (Figure 3-25B). Additionally, the proportion of insoluble protein in the crosslinked samples is decreased. This may therefore suggest that the higher mass bands present correspond to FLAG-hFTO crosslinked to other species, and that sonication has reduced the proportion of these species in the insoluble fraction.

3.6 FTO antibody investigation

Although many robust and well-characterised commercial antibodies to common protein and peptide tags exist, there are many advantages to using high quality IP grade antibodies that recognise the native untagged protein, therefore antibodies to FTO were investigated. Antibodies recognising FTO would allow IP reactions to be carried out on the native protein without the need for a tag, simplifying interpretation of results as the species detected are likely to interact directly with the protein of interest rather than the tag. Tagging of proteins can cause localisation in different cellular locations, and may obscure native interactions or introduce new and unnatural interactions, complicating analysis, and requiring the use of many controls to discount false positives. If tags are not required, analysis of endogenous protein is possible. Analysis of tagged proteins at

endogenous levels requires establishment of a cell line stably expressing the protein of interest fused to an appropriate tag, although in some cases this may not be possible as the protein insertion may be toxic to cells.

3.6.1 Initial antibody testing

Initially, commercially available anti-FTO antibodies were tested for their FTO immunoreactivity alongside a rabbit anti-recombinant mFTO antibody prepared by Eurogentec (described in Section 2.3.2, referred to as "Eurogentec"). Anti-FTO antibodies from Abcam (mouse monoclonal) and Novus Biologicals (rabbit polyclonal), referred to as "Abcam" and "Novus" respectively, were also studied (full antibody details in Chapter 7, Table 8-9).

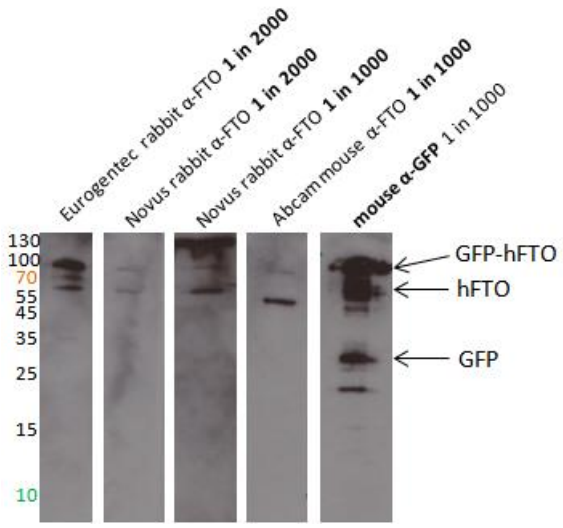


Figure 3-26: Detection of FTO from cell lyate of 9.5 x 10^5 293T cells transiently transfected with N1-EGFP hFTO using antibodies and dilutions indicated. 30 min exposure.

Initially, detection of N1-EGFP hFTO in a 293T cell lysate by western blot was carried out, using the antibody dilutions recommended by suppliers (Figure 3-26). One blot sample was probed using a commercial α-GFP antibody, allowing verification of the presence of hFTO-GFP in the lysate. Overexpressed GFP-tagged hFTO (~90 kDa) and endogenous hFTO (59 kDa) appears to be detected by the anti-FTO antibodies. Interestingly, both the Novus and Abcam antibodies appear to detect endogenous hFTO more effectively than GFP-hFTO, with more intense bands at ~60 kDa than ~90 kDa.

Detection of the tagged N1-EGFP hFTO species may be affected by the presence of a large tag at the *C*-terminus, reducing access to the recognised epitopes.

3.6.2 Abcam anti-FTO

The Abcam antibody produces a band at a lower mass than was observed for detection by the Eurogentec and Novus antibodies (~55 kDa rather than ~60 kDa). To further investigate recognition by the Abcam antibody, the detection of different tagged FTO constructs was attempted; either *N*- or *C*-terminal GFP, or *N*-terminal 3xFLAG. In each case, bands were identified at the same mass (Figure 3-27), with no significant bands at higher masses.

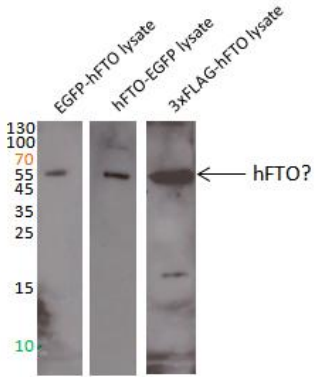


Figure 3-27: FTO constructs overexpressed in 293T cells by transient transfection, with detection using Abcam α-FTO, 1 in 1000. 30 min exposure.

As both of these further species are tagged at the *N*-terminus, the tags may prevent effective protein detection by the antibody. However as the epitope of the monoclonal antibody is in the *C*-terminus of the protein it would seem unlikely, although not impossible, that this region is shielded by an *N*-terminal GFP/FLAG fusion. The western blot of the FLAG-hFTO lysate has an additional band at ~18 kDa, suggesting either the detection of a specific degradation fraction, or cross-reactivity with a different species in the cell lysate.

Recombinant FTO proteins are successfully detected by the Eurogentec antibody (Section 2.3.2), but unfortunately no signal was observed on attempted detection of recombinant proteins (hFTO, ChFTO, mFTO and CmFTO) with the Abcam antibody. It therefore seems unlikely that the band recognised by the Abcam antibody at ~55 kDa in cell lysates is FTO, suggesting that the antibody may be cross-reactive, detecting a

different component of the cell lysate. Therefore the Abcam antibody does not appear suitable for hFTO detection, and was not used further.

3.6.3 Novus anti-FTO

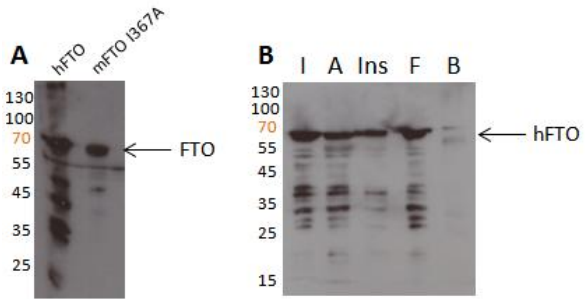


Figure 3-28: (A) Detection of recombinant FTO proteins (1 µg loaded per lane) using Novus α-FTO (1 in 1000), 10 s exposure. (B) Test IP of hFTO using 20 µg Novus α-FTO. 1.5x10^7 293T cells. I: soluble cell lysate, IP input; A: IP input + Novus α-FTO; Ins: insoluble cell lysate; F: IP flowthrough; B: beads. Blot detected using Eurogentec α-FTO. 30 min exposure.

The Novus anti-FTO antibody was also investigated further, with a blot of recombinant proteins (Figure 3-28A) showing good detection of the full length proteins. Immunoprecipitation of FTO by the Novus antibody was also attempted, using whole cell lysates of 293T cells transfected with pcDNA3 hFTO. The antibody-lysate mixture was incubated for 1 hour at 4 °C, then protein G beads were added for a further 1 hour incubation. Using this method, the antibody may bind to the protein before the antibody binds to the beads, enabling separation from the remaining unbound lysate. Although hFTO is well expressed, very little hFTO is present in the bead fraction, as most protein appears to be present in the flow through fraction having not bound to the beads (Figure 3-28B). It is possible that this is due to the small amount of antibody used for the IP reaction, but as the IP by the Novus antibody does not appear to be particularly efficient very large quantities of the antibody would be required for IP, which would be prohibitively expensive, therefore this approach was not pursued.

3.6.4 Eurogentec anti-FTO

The most promising antibody for testing in immunoprecipitation experiments was the Eurogentec rabbit polyclonal antibody raised to recombinant FTO protein, therefore a large scale IP reaction was attempted. Three batches of 293T cells were prepared, one

set of which was not transfected, one set was treated with tranfection reagents without addition of plasmid (mock transfected), and the final set was transfected with pcDNA3 hFTO. The IP was carried out similarly to the reaction for the Novus antibody described previously, but due to the greater availability of the Eurogentec antibody and to try to bind the maximum quantity of hFTO, 1 mg antibody was added to the IP reactions. The mouse anti-rabbit antibody was then bound to protein G beads. As a control, the lysate from mock transfected cells was used in a negative IP reaction, where no antibody was added to the IP mixture. This reaction would give a good indication of any false positive samples which could be present in proteomic data, due to non-specific binding to the agarose beads.

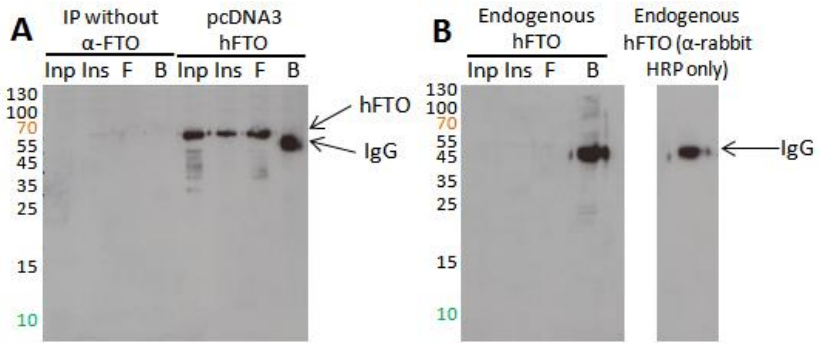


Figure 3-29: 2 x 10^8 293T cells (A) mock transfected control, with transfected pcDNA3 hFTO, or (B) non transfected (endogenous levels of hFTO). Western blot testing protein expression and IP efficiency. Antibody stain control – secondary anti rabbit HRP conjugate antibody probe only. Inp: IP input, Ins: insoluble, F: Bead flow through, B: Beads. 1 min exposures.

Only a very small quantity of endogenous protein was observed by western blot (very faint bands in Figure 3-29A and B lanes 1-3, more visible on longer exposure of films). Although a good amount of hFTO has been produced from the pcDNA3 construct very little, if any, hFTO appears to be present in the bead lane of the positive IP reaction, as the band observed is at a lower mass than the predominant bands in the previous three lanes. This band corresponds to the rabbit anti-FTO antibody heavy chain, and can be visualised by staining with only the secondary anti-rabbit HRP-coupled antibody (Figure 3-29B). This suggests that the IP reaction has not been very efficient, but due to the proximity of the bands corresponding to hFTO (~60 kDa) and the heavy chain of the

anti-FTO antibody (~55 kDa) it is difficult to determine whether any hFTO has been pulled down.

To investigate the products of the IP reaction further, approximately one third of the protein eluted from the beads in the IP reaction was analysed on a 4-12 % gradient NuPAGE gel, allowing better resolution of the proteins at higher mass. The lane corresponding to overexpression of hFTO in Figure 3-30 shows the presence of a small additional band at ~60 kDa, compared to lane 4, which shows staining of the blot with only the secondary antibody. Although this shows that hFTO was pulled down, the quantity is extremely low, given that this comprises one third of a reaction in which overexpression of hFTO was achieved, and a large quantity of anti-hFTO antibody was used. As the IP efficiency was very low, the samples could not be analysed by mass spectrometry to determine interacting species.

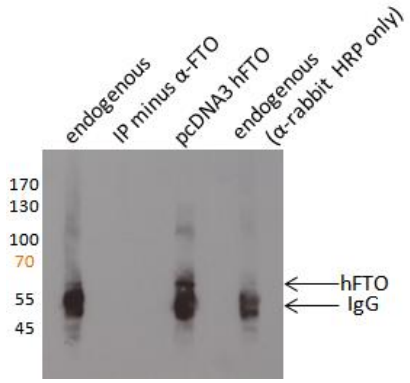


Figure 3-30: 2 x 10^8 293T cells either transfected with pcDNA3 hFTO, or not transfected (producing endogenous levels of hFTO). Western blot testing presence of hFTO in IP bead samples. Lane 4 Antibody stain control – secondary anti rabbit HRP conjugate antibody probe only. 1 min exposure.

Due to the poor IP reactivity of the antibodies tested, it was not considered possible to use these for successful ChIP assays. Instead, tagged protein would need to be used, so the tag could be exploited to 'pull-down' FTO and its interacting partners. However, ChIP experiments should be carried out using endogenous, rather than grossly overexpressed, levels of protein. If these studies were to be carried out in the future it would be necessary to establish and maintain cell lines stably expressing tagged hFTO species at a level close to that for endogenous expression, a line of investigation which

would require several months of work, and was not pursued here, as the identification of potential protein interactors was prioritised.

3.7 Expression of N- *and* C*-terminal FTO constructs*

As FTO is known to have two distinct domains[264] it is likely that different species interact with the two domains, due to their dissimilar three dimensional structures, and likely different roles. Investigation of the domains was therefore carried out in intial IP experiments. GFP-tagged constructs producing each of these domains were separately transfected into 293T cells, and small scale IP reactions were carried out using anti-GFP beads. Western blot analysis (Figure 3-31) showed poor production of both constructs (particularly in the case of NhFTO) when compared to production of full length wildtype proteins carried out under the same conditions (Figure 3-22B). Poor expression of the *N*-terminal domain of FTO is not unprecedented, as a similar observation was made previously for *in vitro* expression of recombinant mFTO proteins truncated at the *C*-terminus (Section 2.2.2), with very low levels of insoluble truncated protein produced. Additionally, significant degradation to GFP is observed for these proteins, as the bands corresponding to GFP are much more intense on blots than those corresponding to the full length proteins.

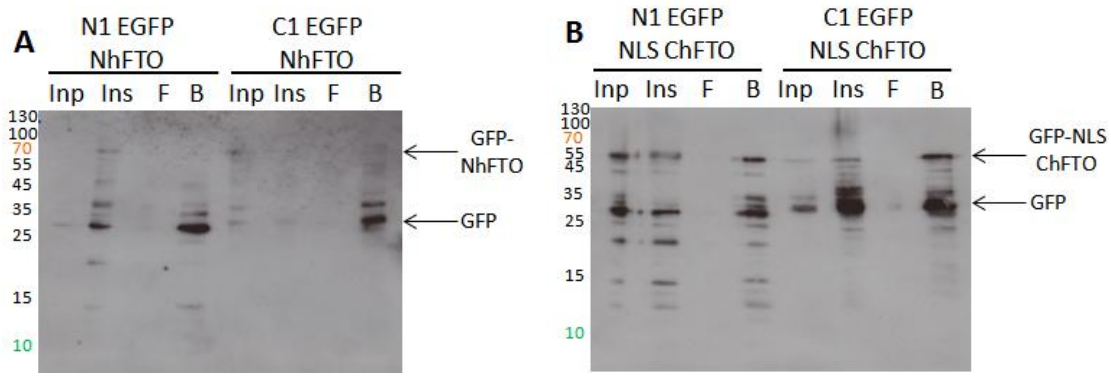


Figure 3-31: Small scale test of cell lysis and IP conditions for (A) NhFTO and (B) NLS-ChFTO constructs. 9.5 x 10^5 293T cells were transfected with the constructs indicated. Inp: IP input; Ins: insoluble fraction; F: Bead flow through; B: Beads. 30 min exposure.

As the subcellular localisation of these proteins is not distinct (Section 3.3.5) and their production levels are poor, it does not seem that progression to large scale production and analysis of these proteins would be worthwhile. However, these constructs should

be valuable for later determination of the domain of FTO with which any protein or nucleic acid partners may interact.

3.8 hFTO dimerisation in cells

Dimerisation of mFTO and hFTO was previously identified *in vitro* for purified recombinant proteins (see Section 2.3). However, this may be an artefact due to use of highly concentrated proteins, and may only occur in the absence of other factors present in cells. Therefore, a dual immunoprecipitation approach was taken to determine whether dimerisation may be observed in cell lysate samples. 293T cells were co-transfected with equal quantities of 3xFLAG-hFTO and empty/hFTO EGFP constructs. By use of two orthogonal affinity tags, hFTO can be immunoprecipitated using one tag, followed by probing of the immunoprecipitated proteins for the presence of hFTO tagged with the alternate affinity tag. If hFTO with both of the tags used is detected in the immunoprecipitated samples, it can be assumed that hFTO units interacts with other hFTO units in the cell, therefore oligomerisation may be occurring. Samples were analysed for hFTO tagged with the GFP protein at either the *N*- or *C*-termini to try to prevent potential blocking of any binding or interaction between the hFTO species by the large GFP protein. Lysates of cells transfected with non-fused GFP protein were studied to rule out a possible interaction of the GFP protein itself with the FLAG-hFTO protein. The immunoprecipitation reactions were carried out using either anti-GFP or anti-FLAG beads. Samples from both IP reactions were analysed by western blot and probed with GFP and FLAG antibodies.

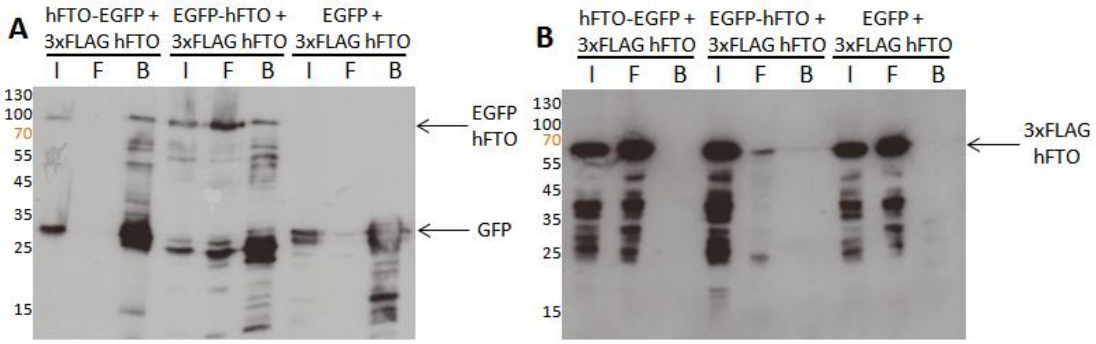


Figure 3-32: 5.5 x 10^6 293T cells co-transfected with 3xFLAG hFTO, with either N1-EGFP hFTO, C1-EGFP hFTO or empty EGFP. Cells lysed, then α-GFP immunoprecipitation using GFPTrap beads. Western blots probed using either (A) α-GFP, or (B) α-FLAG antibodies. I: Soluble cell lysate, used as IP input; F: Bead flow through; B: Beads. 1 min exposures.

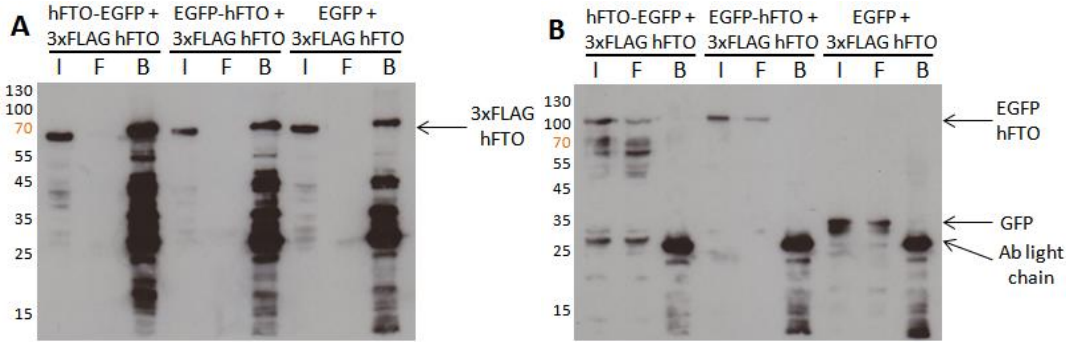


Figure 3-33: 5.5 x 10^6 293T cells cotransfected with 3xFLAG hFTO, with either N1-EGFP hFTO, C1-EGFP hFTO or empty EGFP. Cells lysed, then α-FLAG immunoprecipitation using α-FLAG beads. Western blots probed using either (A) α-FLAG, or (B) α-GFP antibodies. I: Soluble cell lysate, used as IP input; F: Bead flow through; B: Beads. 1 min exposures.

In each case, both the FLAG and GFP constructs have been well produced (lanes labelled "I" in Figure 3-32 and Figure 3-33), and the appropriately tagged proteins have bound to the beads used for the IP reaction. However, no bands are visible at the expected mass of the hFTO species with the alternate tag to that used for the IP reaction (lanes labelled "B" in Figure 3-32B and Figure 3-33B). In some cases it appears that "bleeding" from lanes containing large quantities of protein into the neighbouring lane may have occurred, for example in lane 6 of Figure 3-32B, where it appears that a band may be present at ~60 kDa. This therefore makes it difficult to determine whether a true band exists. However, no bands were visible when increased quantities of bead samples were run on fresh blots (blots not shown). This therefore suggests that, under the conditions tested, no interaction occurs between overexpressed FLAG-hFTO and GFP-

hFTO, and potentially that FTO does not dimerise in cells. To verify this more completely untagged or endogenous proteins would need to be investigated.

3.9 Large scale immunoprecipitation of hFTO and proteomic analysis

3.9.1 Anti-GFP immunoprecipitation

Following the described establishment of lysis and IP conditions (Sections 3.4 and 3.5), proteins coimmunoprecipitating with hFTO may be identified by mass spectrometry. Initially, sufficient bait protein must be produced in cells. Therefore, large scale cell growths were transfected with EGFP-NLS, C1 EGFP-hFTO or N1 hFTO-EGFP, and an IP reaction was carried out using whole cell lysates. To determine the level of expression of the protein constructs, and the IP efficiency, western blots were run using 1 % of the total samples obtained (Figure 3-34). These showed that high levels of each of the proteins were synthesised by the cells, with a good proportion of the protein bound to the beads in each case.

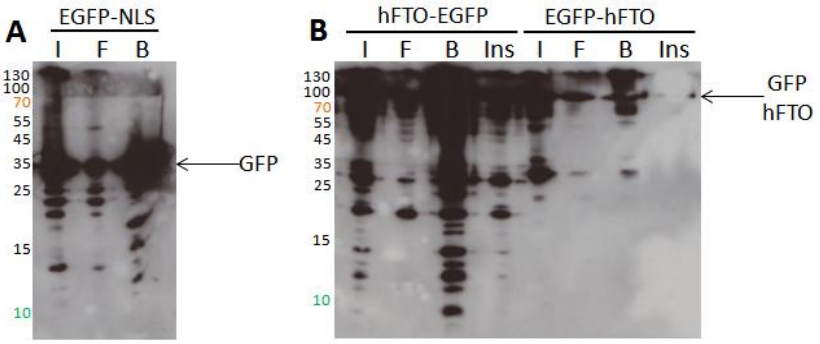


Figure 3-34: Western blots testing expression and immunoprecipitation of large scale GFP and hFTO constructs. 2 x 10^8 cells transfected with either (A) EGFP-NLS, or (B) C1-GFP hFTO or N1-GFP hFTO, and immunoprecipitated using GFPTrap beads. 1 µl each sample I: IP input; F: Bead flow through; B: Beads; Ins: Insoluble fraction. 30 min exposures.

As the IP reaction was successful, the immunoprecipitated "bead" samples were separated by electrophoresis, using a precast NuPAGE gradient gel, which allows greater resolution of protein bands across the whole mass spectrum (Figure 3-35A). Staining of the gel was carried out using a Colloidal Coomassie stain (Invitrogen), due to the higher sensitivity of this detection method, which means that proteins at low concentrations may be visualised.

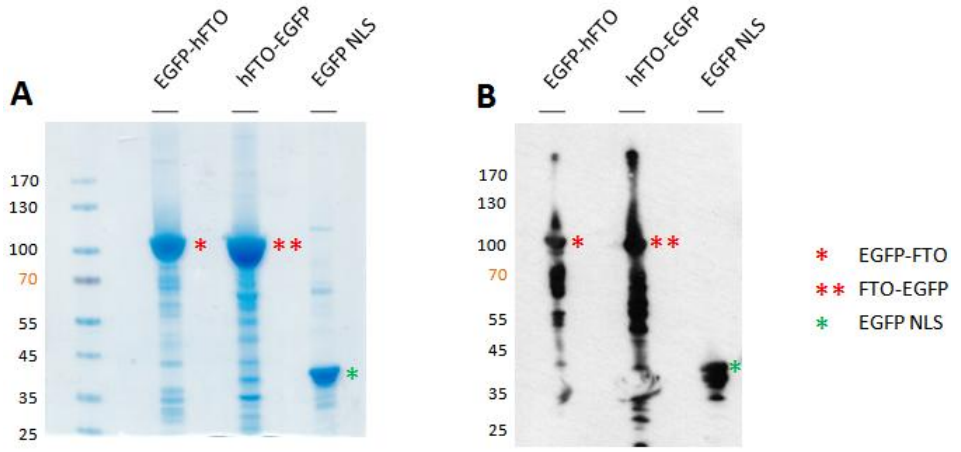


Figure 3-35: (A) 4-12 % gradient NuPAGE gel stained with Colloidal Coomassie. 2 x 10^8 cells transfected with either EGFP-NLS, C1-GFP hFTO, or N1-GFP hFTO, and immunoprecipitated using GFPTrap beads. Proteins eluted from beads using SDS-PAGE loading buffer, 35 µl each sample loaded per lane. (B) Western blot transferred from 4-12 % NuPAGE gel. 2 x 10^8 cells transfected with either EGFP-NLS, C1-GFP hFTO, or N1-GFP hFTO, and immunoprecipitated using GFP-Trap beads. Proteins eluted from beads using SDS-PAGE loading buffer, 1 µl each sample loaded per lane. Probed using mouse anti-GFP and anti-mouse HRP conjugate. 1 min exposure.

As anticipated from western blot analysis, significant bands were observed at the expected masses of the hFTO-EGFP species (~90 kDa) and EGFP-NLS (~35 kDa), suggesting that large quantities of the overexpressed target proteins had been produced. In addition to these bands, many bands at lower mass were visible. An initial indication of possible protein interacting species may be given by the observation of significant bands visible in the "experimental" lane(s), but not in the control sample. Comparison of lanes 1 and 2 (EGFP-hFTO and hFTO-EGFP) with lane 3 (EGFP-NLS) in Figure 3-35A reveals that the majority of bands in the hFTO sample lanes are not observed in the GFP control lane. However, as has been observed in all western blots of hFTO, multiple bands corresponding to hFTO or tagged species are identified, suggesting that a significant amount of degradation of hFTO occurs. Protease inhibitors were added upon cell lysis, although not at subsequent stages through the IP process, but their addition would not preclude protein degradation by other mechanisms, such as hydrolysis and oxidation. Furthermore oxidative self-cleavage of the protein backbone in the presence of iron and ascorbate may occur, and has been observed for several other 2OG oxygenases[314-315]. Additionally, FTO is known to be rapidly turned over in cells,

because it is degraded via the ubiquitin/proteasome pathway[305]. To attempt to determine which, if any, of the bands in the gel can be attributed to the degradation of FTO, the same samples were run on a further NuPAGE gel using the same conditions as previously, to allow direct comparison of the bands present. Figure 3-35B shows that bands across the mass range studied are identified by the anti-GFP antibody, therefore it is likely that the majority of the stained bands in the FTO sample lanes correspond to protein degradation rather than interacting species.

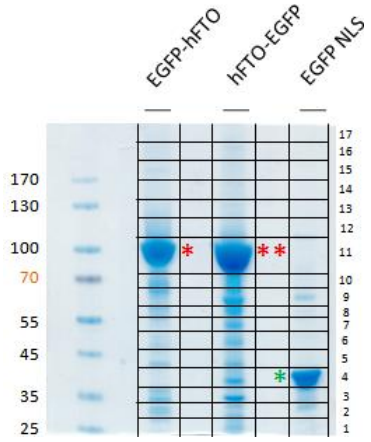


Figure 3-36: 4-12 % gradient NuPAGE gel stained with Colloidal Coomassie; samples as in Figure 3-35. Grid and numerical labelling demarcates dissection of gel pieces for trypsin digestion.

To try to identify bands on the gel, the gel lanes were excised, cut into 17 equivalent portions, (outlined in Figure 3-36) then subjected to in-gel trypsinolysis prior to mass spectrometry[316]. LC/MS/MS analysis of the tryptic peptides was carried out by Dr Holger Kramer (Department of Physiology, Anatomy and Genetics, University of Oxford). The mass profiles obtained for each lane were merged, then searched using the MASCOT search engine, which uses mass spectrometry peptide data to identify protein matches from human genome primary sequence databases. Common contaminant proteins such as trypsin, keratin, heat shock proteins and immunoglobulins were excluded from the data obtained, and comparison of results to those obtained from the control GFP-NLS IP allowed deletion of false positive proteins also present in the GFP

control. The remaining proteins (Tables 3-4 and 3-5) are those that are likely to interact with FTO.

Accession number	Protein	Score
Q9C0B1	Alpha-ketoglutarate-dependent dioxygenase FTO	14645
P10759	AMP deaminase 1	105
Q9ULA0	Aspartyl aminopeptidase	86
Q8NBS9	Thioredoxin domain-containing protein 5	82
Q3SZ18	Hypoxanthine-guanine phosphoribosyltransferase	75
P70670	Nascent polypeptide-associated complex subunit alpha, muscle-specific form	66
Q9EQG6	Kinase D-interacting substrate of 220 kDa	56
P31943	Heterogeneous nuclear ribonucleoprotein H	55
Q52KB6	C2 domain-containing protein 3	53
Q4ZG55	GREB1	53
A3KMX7	Putative malate dehydrogenase 1B	50
Q3ZCF3	S-phase kinase-associated protein 1	47
Q5RBA5	Sequestosome-1	46
O43150	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 2	45
Q6B4U9	Peroxiredoxin-1	44
P62258	14-3-3 protein epsilon	44
Q9D2G2	Dihydrolipoyllysine-residue succinyltransferase component of 2OG dehydrogenase complex,	42
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2	40
Q4R6F8	T-complex protein 1 subunit beta	36
P70378	Fibroblast growth factor 11	36
Q5RAP1	T-complex protein 1 subunit theta	36
O95980	Reversion-inducing cysteine-rich protein with Kazal motifs	35
Q5JTC6	Protein FAM123B	34
B3VSB7	Hippocalcin-like protein 1	34
P39689	Cyclin-dependent kinase inhibitor 1	34
Q9Z351	Potassium voltage-gated channel subfamily KQT member 2	34
Q9H2L5	Ras association domain-containing protein 4	33
Q9BXT5	Testis-expressed sequence 15 protein	33
P97820	Mitogen-activated protein kinase 4	33
Q6Q899	Probable ATP-dependent RNA helicase DDX58	32
P51572	B-cell receptor-associated protein 31	32
A7MB11	Transforming growth factor-beta receptor-associated protein 1	32
Q5TJG6	Bromodomain-containing protein 2	32
Q95132	Intercellular adhesion molecule 1	32
P32362	Uroporphyrinogen decarboxylase	32
Q6P8K8	Carboxypeptidase A4	32
Q5FVM4	Non-POU domain-containing octamer-binding protein	30
Q63035	NACHT, LRR and PYD domains-containing protein 6	30
Q13232	Nucleoside diphosphate kinase 3	30
P79457	Histone demethylase UTY	30
Q9H013	Disintegrin and metalloproteinase domain-containing protein 19	30
Q9C0D5	TANC1	29
Q570Y9	DEP domain-containing mTOR-interacting protein	29
P36542	ATP synthase subunit gamma, mitochondrial	28
Q8BMG7	Rab3 GTPase-activating protein non-catalytic subunit	28

Accession number	Protein	Score
Q5M7W4	Transmembrane channel-like protein 5	28
Q9R0Q2	Leukotriene B4 receptor 1	28
Q8CG80	SH2 domain-containing adapter protein F	27
O02789	DNA nucleotidylexotransferase	27
Q86W25	NACHT, LRR and PYD domains-containing protein 13	27
P51828	Adenylate cyclase type 7	26
Q8QZV7	Cell cycle regulator Mat89Bb homolog	26
Q13112	Chromatin assembly factor 1 subunit B	25
Q9R0I8	Phosphatidylinositol-5-phosphate 4-kinase type-2 alpha	25
Q9Z0T0	Thiopurine S-methyltransferase	25
Q52LR7	Enhancer of polycomb homolog 2	25
A5D7U0	Cytoskeleton-associated protein 2	25
Q6PAL8	DENN domain-containing protein 5A	25
Q9Y543	Transcription factor HES-2	24
Q9D338	39S ribosomal protein L19	24
P32766	Cystatin-8	24
A5D7M3	Growth arrest-specific protein 8	23
Q05920	Pyruvate carboxylase, mitochondrial	23
Q99490	Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2	23
O18737	Calcium-binding and coiled-coil domain-containing protein 2	22
Q6AYP4	FAM164C	22
P08680	5-aminolevulinate synthase, erythroid-specific, mitochondrial	21
P09871	Complement C1s subcomponent	21
Q8BWS5	G protein-regulated inducer of neurite outgrowth 3	21
Q5RB23	Tyrosine-protein kinase JAK2	21
Q91ZN5	Adenosine 3'-phospho 5'-phosphosulfate transporter 1	20

Table 3-3: IP proteins from GFP-hFTO (C1 EGFP) overexpression and anti-GFP pulldown. Blue shading: FTO; Green shading: proteins identified in several IP reactions; Red text, nuclear/nucleic acid associated proteins; Yellow shading: ATP related proteins.

Accession number	Protein	Score
Q9C0B1	Alpha-ketoglutarate-dependent dioxygenase FTO	13708
Q8SQH5	ADP/ATP translocase 2	114
P02722	ADP/ATP translocase 1	112
Q3SZ18	Hypoxanthine-guanine phosphoribosyltransferase	102
Q6S8J3	POTE ankyrin domain family member F	101
P10759	AMP deaminase 1	94
Q5VTE0	Putative elongation factor 1-alpha-like 3	82
P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2OG dehydrogenase complex	81
Q6B4U9	Peroxiredoxin-1	77
P0C274	Ubiquitin	75
P32007	ADP/ATP translocase 3	70
P14136	Glial fibrillary acidic protein	68
P07814	Bifunctional aminoacyl-tRNA synthetase	64
Q3ZCF3	S-phase kinase-associated protein 1	63
Q9UBN7	Histone deacetylase 6	63
Q99497	Protein DJ-1	61
P49327	Fatty acid synthase	59
Q8NBS9	Thioredoxin domain-containing protein 5	57
Q3V132	ADP/ATP translocase 4	57
Q5RBA5	Sequestosome-1	56
Q99714	3-hydroxyacyl-CoA dehydrogenase type-2	55
Q4ZG55	GREB1	55
Q2TA29	Ras-related protein Rab-11A	55
Q9D883	Splicing factor U2AF 35 kDa subunit	53
P27708	CAD protein	53
P02786	Transferrin receptor protein 1	53
Q4R924	S-adenosylmethionine synthase isoform type-2	50
Q5R8R5	Glutathione S-transferase P	48

Accession number	Protein	Score
Q5THK1	Peroxiredoxin-4	48
Q9BGI2	GTP-binding protein SAR1a	45
Q3T0D7	Poly(rC)-binding protein 1	43
Q5E9A3	Nascent polypeptide-associated complex subunit alpha, muscle-specific form	42
P70670	GAP with SH3 domain, ANK repeat and PH domain-containing protein 2	42
A1A5D9	Bicaudal D-related protein 2	41
P56524	Histone deacetylase 4	38
P22102	Trifunctional purine biosynthetic protein adenosine-3	37
Q8TER0	Sushi, nidogen and EGF-like domain-containing protein 1	36
P04574	Calpain small subunit 1	35
Q9BSJ1	SPRY domain-containing protein 5	35
B3VSB7	Hippocalcin-like protein 1	35
Q5R7H5	DDB1- and CUL4-associated factor 11	34
Q5E9Z8	U6 snRNA-associated Sm-like protein LSm1	34
Q5EAD2	D-3-phosphoglycerate dehydrogenase	33
P15246	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	33
Q6Q899	Probable ATP-dependent RNA helicase DDX58	33
Q14410	Glycerol kinase 2	32
P31662	Orphan sodium- and chloride-dependent neurotransmitter transporter NTT4	32
Q9D0F6	Replication factor C subunit 5	32
Q80T14	Extracellular matrix protein FRAS1	32
Q9D2L9	Protein FAM111A	32
Q63624	Splicing factor, arginine/serine-rich 19	32
Q5VZ89	DENN domain-containing protein 4C	32
Q6P8K8	Carboxypeptidase A4	31
Q9P2N4	Disintegrin and metalloproteinase with thrombospondin motifs 9	31
Q96K78	Probable G-protein coupled receptor 128	31
P16125	L-lactate dehydrogenase B chain	30
Q96GE4	Coiled-coil domain-containing protein 45	29
O35298	Acyloxyacyl hydrolase	29
Q06194	Coagulation factor VIII	29
Q9H8Y5	Ankyrin repeat and zinc finger domain-containing protein 1	29
Q8VIK3	Histone H1oo	29
Q5E983	Elongation factor 1-beta; Short=EF-1-beta	28
P26378	ELAV-like protein 4	28
Q8TEM1	Nuclear pore membrane glycoprotein 210	28
Q9D9J8	Short palate, lung and nasal epithelium carcinoma-associated protein 3 homologue	27
P48500	Triosephosphate isomerase	27
Q08DP3	RELT-like protein 1	27
Q8VEG4	Exonuclease 3'-5' domain-containing protein 2	27
P07522	Pro-epidermal growth factor	26
Q05921	2-5A-dependent ribonuclease	26
P97372	Proteasome activator complex subunit 2	26
A2AAE1	Fragile site-associated protein homologue	25
Q14692	Ribosome biogenesis protein BMS1 homologue	25
Q008S8	Epithelial cell-transforming sequence 2 oncogene-like	25
Q5E971	Transmembrane emp24 domain-containing protein 10	25
Q5RCF3	Cullin-2	25
P43116	Prostaglandin E2 receptor EP2 subtype	24
Q9NYB0	Telomeric repeat-binding factor 2-interacting protein 1	24
Q866F4	Adenylate cyclase type 10	24
A2A9I7	Doublesex- and mab-3-related transcription factor B1	24
P10166	Protein L-Myc-1	23
Q61001	Laminin subunit alpha-5	23
Q3T0J3	39S ribosomal protein L16, mitochondrial	23
P47863	Aquaporin-4	22
P25119	Tumor necrosis factor receptor superfamily member 1B	21

Table 3-4: IP proteins from hFTO-GFP (N1 EGFP) overexpression and anti-GFP pulldown. Blue shading: FTO; Green shading: proteins identified in several IP reactions; Red text, nuclear/nucleic acid associated proteins; Yellow shading: ATP related proteins.

Some of the proteins identified from the MASCOT search of the UniProt database are likely false positives. To attempt to estimate the proportion of the proteins identified that are likely falsely present, a further MASCOT search can be run using the same search parameters as for the original search, but this time searching against a "decoy" database. This database contains randomised protein sequences, in which it is expected that no true matches will be identified. Any matches identified from the decoy database are therefore false positives, giving a good estimate of the false positives also identified in the target database. This methodology has been applied previously as a method for increasing the stringency of reported database searches[317]. Decoy searches for the GFP IP samples (Table 3-5) suggest that, using the identity threshold, ~3 % of the protein hits identified by the MASCOT search are in fact false positives, although this proportion is increased when using the more broad criteria of homology.

Peptide match %	EGFP-hFTO	hFTO-EGFP	GFP-NLS
Above identity threshold	2.83%	2.36%	3.31%
Above identity or homology threshold	7.69%	7.27%	9.73%

Table 3-5: Percentage identification rate for MASCOT decoy search analysis of IP samples for hFTO-GFP.

Tables 3-4 and 3-5 show that in both anti-GFP reactions FTO has the highest protein score, therefore the IP reactions have been successful in immunoprecipitating FTO from the cell lysates. Although a number of potentially interacting proteins were found by this analysis method, to confirm that these results are true it is beneficial to verify the identification of these immunoprecipitated proteins using a complementary IP approach with a different tag. If the same interacting proteins are identified with both tags, it is more likely that a true interaction has been identified.

3.9.2 Anti-FLAG immunoprecipitation

Complementary IP reactions were carried out with 3xFLAG-hFTO. As a control, cells were transfected with the empty 3xFLAG-CMV10 vector. FLAG-hFTO is well produced, with a good quantity binding to the beads (Figure 3-37). Other than the band

corresponding to the light chain of the anti-FLAG antibody, no further bands were visible on a western blot of protein samples from these cells (Figure 3-37) likely because the 3xFLAG peptide is too small (23 residues, ~3 kDa) to be properly produced by the cells, or visualised on the SDS-PAGE gel used.

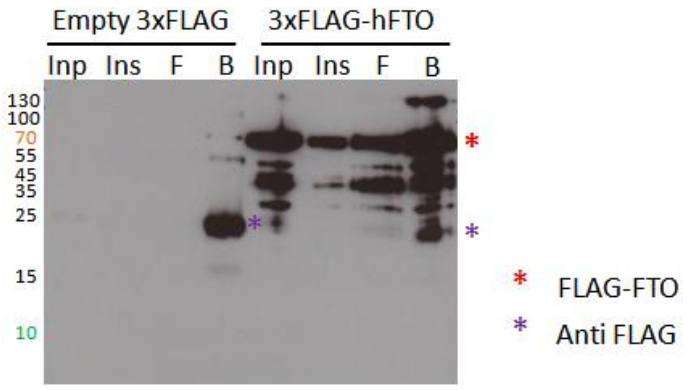


Figure 3-37: Western blot testing expression and immunoprecipitation of 3xFLAG and 3xFLAG-hFTO constructs. 2 x 10^8 cells transfected with either empty 3xFLAG vector or 3xFLAG-hFTO, and immunoprecipitated using anti-FLAG beads. 1 µl each sample loaded. Inp: IP input; Ins: Insoluble fraction; F: Bead flow through; B: Beads. 1 min exposure.

The IP samples were analysed on a NuPAGE gel (Figure 3-38), with a large band at the expected mass of FLAG-hFTO confirming that the IP was successful. The gel lanes were excised and divided into 20 pieces, and analysed by trypsinolysis and LC/MS/MS. The proteins likely interacting with hFTO are listed in Table 3.6.

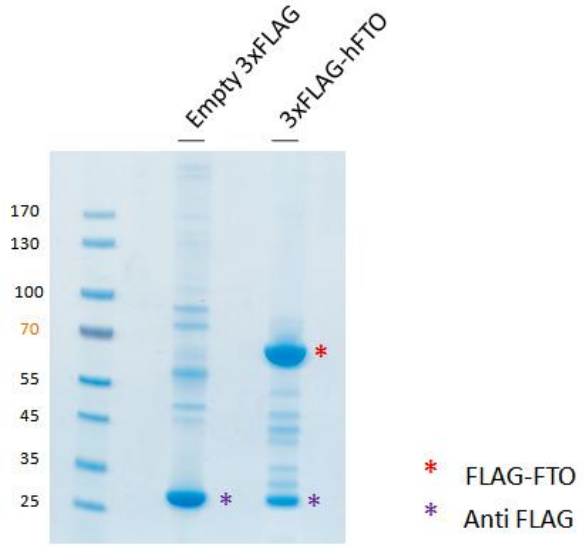


Figure 3-38: 4-12 % NuPAGE gel stained with Colloidal Coomassie. 2 x 10^8 cells transfected with either empty 3xFLAG vector or 3xFLAG-hFTO, and immunoprecipitated using anti-FLAG beads. Proteins eluted from beads using SDS-PAGE loading buffer, 35 µl each sample loaded per lane.

Accession number	Protein	Score
Q9C0B1	Alpha-ketoglutarate-dependent dioxygenase FTO	6957
P36957	Dihydrolipoyllysine-residue succinyltransferase component of 2OG dehydrogenase complex	103
P04075	Fructose-bisphosphate aldolase A	76
P05109	Protein S100-A8	48
O95980	Reversion-inducing cysteine-rich protein with Kazal motifs	46
Q13501	Sequestosome-1	42
P08195	4F2 cell-surface antigen heavy chain	40
P06702	Protein S100-A9	40
Q9H2L5	Ras association domain-containing protein 4	38
Q7L0J3	Synaptic vesicle glycoprotein 2A	37
P98160	Basement membrane-specific heparan sulfate proteoglycan core protein	37
Q96R06	Sperm-associated antigen 5	36
Q9H2J7	Orphan sodium- and chloride-dependent neurotransmitter transporter NTT73	36
Q8TER0	Sushi, nidogen and EGF-like domain-containing protein 1	35
Q0ZLH3	Pejvakin	34
Q8N8J0	Putative phosphatidylinositol 4-kinase alpha-like protein P1	34
Q13103	Secreted phosphoprotein 24	34
O43909	Exostosin-like 3	34
Q6P158	Putative ATP-dependent RNA helicase DHX57	33
Q14CB8	Rho GTPase-activating protein 19	33
Q13243	Serine/arginine-rich splicing factor 5	33
P0C0S5	Histone H2A.Z	33
P0C0S8	Histone H2A type 1	33
Q96L34	MAP/microtubule affinity-regulating kinase 4	32
Q6ZP82	Coiled-coil domain-containing protein 141	32
P33981	Dual specificity protein kinase TTK	31
Q5THJ4	Vacuolar protein sorting-associated protein 13D	31
Q9HDC9	Adipocyte plasma membrane-associated protein	31
Q8TAF3	WD repeat-containing protein 48	31
A4D0S4	Laminin subunit beta-4	31

Accession number	Protein	Score
Q9BTC0	Death-inducer obliterator 1	30
P35556	Fibrillin-2	29
Q9NXN4	Ganglioside-induced differentiation-associated protein 2	29
Q8NFR7	Coiled-coil domain-containing protein 148	29
Q9BXU1	Serine/threonine-protein kinase 31	28
Q96MZ4	Uncharacterized protein C4orf39	28
Q99665	Interleukin-12 receptor subunit beta-2	28
Q13470	Non-receptor tyrosine-protein kinase TNK1	27
Q8WXS8	A disintegrin and metalloproteinase with thrombospondin motifs 14	27
P49796	Regulator of G-protein signaling 3	27
Q5JU85	IQ motif and SEC7 domain-containing protein 2	27
Q8WWQ8	Stabilin-2	27
Q14690	Protein RRP5 homolog	27
Q8NDB2	B-cell scaffold protein with ankyrin repeats	27
Q7Z408	CUB and sushi domain-containing protein 2	26
O14980	Exportin-1	26
Q5T8P6	RNA-binding protein 26	26
Q7Z698	Sprouty-related, EVH1 domain-containing protein 2	26
Q7Z418	Potassium channel subfamily K member 18	26
P54760	Ephrin type-B receptor 4	26
P59510	A disintegrin and metalloproteinase with thrombospondin motifs 20	25
Q8NCL4	Polypeptide N-acetylgalactosaminyltransferase 6	25
Q9NZR2	Low-density lipoprotein receptor-related protein 1B	25
O94903	Proline synthase co-transcribed bacterial homolog protein	25
P21709	Ephrin type-A receptor 1	25
Q7Z402	Transmembrane channel-like protein 7	25
P62979	Ubiquitin-40S ribosomal protein S27a	25
P37059	Estradiol 17-beta-dehydrogenase 2	24
Q15075	Early endosome antigen 1	24
Q6IQ23	Pleckstrin homology domain-containing family A member 7	23
P52594	Arf-GAP domain and FG repeats-containing protein 1	23
Q6UWR7	Ectonucleotide pyrophosphatase/phosphodiesterase family member 6	23
Q96S15	WD repeat-containing protein 24	23
P52948	Nuclear pore complex protein Nup98-Nup96	23
O00451	GDNF family receptor alpha-2	23
Q6P3S6	F-box only protein 42	23
Q5SYB0	FERM and PDZ domain-containing protein 1	23
P11498	Pyruvate carboxylase, mitochondrial	23
Q9BRC7	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase delta-4	23
Q6T4P5	Lipid phosphate phosphatase-related protein type 3	23
Q9UKT7	F-box/LRR-repeat protein 3	22
A6NDE4	RNA-binding motif protein, Y chromosome, family 1 member B	22
Q15378	Putative RNA-binding motif protein, Y chromosome, family 1 member H	22
Q14186	Transcription factor Dp-11;	22
Q6ZRI8	Rho GTPase-activating protein 36	21
Q9NTZ6	RNA-binding protein 12	21
Q04837	Single-stranded DNA-binding protein, mitochondrial	20
Q8WUJ3	Protein KIAA1199	20

Table 3-6: IP proteins from FLAG-hFTO overexpression and anti-FLAG pulldown. Blue shading: FTO; Green shading: proteins identified in several IP reactions; Red text, nuclear/nucleic acid associated proteins.

3.9.3 Comparison of immunoprecipitation results

Table 3-7 shows a comparison between proteins pulled down in both GFP and FLAG hFTO IP reactions. Only 6 proteins were identified in both IP reactions sets, but not in either of the control reactions. There is therefore little crossover between different IP method results, with few common families or links between the proteins identified, and no obvious hits. Additionally many of the identified proteins have low scores, and are not found in the nucleus, where FTO is known to be localised.

Protein	Score		
	GFP-hFTO	hFTO-GFP	FLAG-hFTO
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	42	81	103
Reversion-inducing cysteine-rich protein with Kazal motifs	35	n/a	46
Sequestosome-1	46	56	42
Ras association domain-containing protein 4	33	n/a	38
Ubiquitin-40S ribosomal protein S27a	n/a	75	24
Pyruvate carboxylase, mitochondrial	23	n/a	23

Table 3-7: Proteins identified in both GFP and FLAG hFTO IP products.

The mass spectra of samples corresponding to FTO were also studied using MASCOT to determine whether any modifications to the FTO proteins themselves are observed. Possible phosphorylation sites were observed on tryptic peptides through the FTO sequence. No hydroxylation or SUMOylation of sequences was observed. It is likely that FTO is ubiquitinated, as ubiquitin is a consistent hit in the pull downs (Table 3-7), and a further hit found in each of the FTO IP products is sequestosome (also known as p62), which is a ubiquitin binding protein[318]. However using these data it is not possible to identify any possible sites of ubiquitination as the iodoacetamide used for methylation of cysteine residues has been found to be incompatible with this analysis[319]. In support of the ubiquitin interaction a previous study has also identified that hFTO is rapidly degraded *in vivo* by the ubiquitin-proteasome pathway[305].

An antibody was available for the highest scoring of the potential hits from proteomic analysis, the dihydrolipoyllysine-residue succinyltransferase component of the 2OG dehydrogenase complex (DLST)[320]. The 2OG dehydrogenase complex has multiple copies of three enzymatic components (oxoglutarate dehydrogenase (OGDH), DLST and dihydrolipoyl dehydrogenase (DLD))[321] and catalyses the overall conversion of

2OG to succinyl-CoA and CO_2, as part of the citric acid cycle[322]. DLST is therefore a mitochondrial protein, but has a link to the turnover of 2OG, one of the co-substrates of the FTO reaction.

To try to verify the presence of DLST in proteins immunoprecipitated from the hFTO or control samples a western blot was carried out (Figure 3-39), using the suggested antibody dilution range. The expected mass of DLST is 48.8 kDa, however on the blot using a dilution of 1 in 500 very faint bands can be observed at ~60, 55 and 35 kDa. The detection of multiple bands suggests that the antibody is non-specifically detecting different species in the IP samples. Unfortunately no positive control was available which was known to contain a significant quantity of DLST to determine whether detection using this antibody is successful.

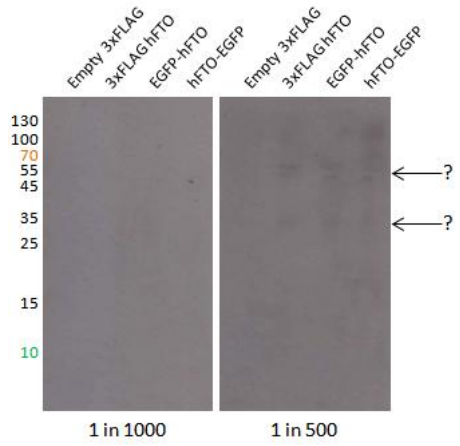


Figure 3-39: Attempt to detect DLST in immunoprecipitation samples. 5 µl IP samples run, and western blots probed with indicated dilutions of α-DLST primary antibody. 1 hour exposure.

3.10 IP and proteomic attempts in mouse muscle tissue

The ultimate goal of this work is analysis of cellular interaction partners of FTO using tissue samples from mice with different genetic backgrounds. This could give valuable whole-organism information about the interactions, and consequently about the role of FTO in mammalian obesity. In contrast to using the more artificial environment of culture of cancer cell lines, which can often have significantly different gene expression and developmental characteristics from more "normal" cells, mouse tissues would contain proteins at their normal endogenous levels. However, these experiments are significantly more challenging than the work previously discussed here. In addition to

the vastly increased time and cost of looking after mice, and the ethical considerations of using animal tissues, the quantity of tissue available particularly for the necessary extensive optimisation is very limited when compared to cell culture. Furthermore, interpretation of proteomic results from some tissues can also be complicated as each cell type produces different proteins to different extents, for example muscle cells contain large amounts of actin and myosin[323], adipose tissue contains high levels of lipids, and nervous system tissues contain significant quantities of synaptic proteins.

Despite the challenges of this approach, the decision was taken to attempt immunoprecipitations of FTO from mouse tissues based on the previously established conditions, to determine the feasibility of such an approach. These experiments were carried out at MRC Harwell, with the assistance of Dr Chris Church.

3.10.1 IP from I367F mouse tissues

As discussed in Chapter 2, a mouse with an I367F substitution in the FTO sequence was identified and studied, with mouse lines homozygous and heterozygous for the mutation developed. In addition to the phenotypic and *in vitro* analyses[242], initial immunoprecipitation reactions were carried out by Dr Chris Church. Due to the ease of protein extraction from brain tissues, these were used for IP analyses.

Whole cell lysate of brain tissue from WT and homozygous I367F mice was immunoprecipitated using the Eurogentec anti-mFTO antibody. Despite the problems previously encountered when attempting IP of hFTO with the Eurogentec antibody, mFTO appears to be bound slightly more efficiently by the antibody than hFTO (Dr Chris Church, unpublished observations). These samples were processed alongside a control negative IP reaction, using lysate from tissue from a wildtype mouse. Trypsinolysis and MS analysis of the samples was carried out in the same manner as described previously, with the assistance of Dr Rosario Romero (MRC Harwell).

Accession number	Protein	Score
Q8BNF8	Vcp Transitional endoplasmic reticulum ATPase	243
Q04447	Ckb Creatine kinase B-type	137
Q76MZ3	**Ppp2r1a Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform**	87
P63328	**Ppp3ca Isoform 1 of Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform1**	78
Q80TL0	**Ppm1e Protein phosphatase 1E**	75
Q8BGW1	Fto	74
P46460	Nsf Vesicle-fusing ATPase	52
P11984	Tcp1 T-complex protein 1 subunit alpha A	51

Table 3-8: IP proteins from brain tissue lysate of wildtype mouse. Bold text: serine-threonine phosphatases, Yellow shading: ATP related proteins; Blue shading: FTO.

Accession number	Protein	Score
Q9Z1G4	Atp6v0a1 Isoform A1-II of V-type proton ATPase 116 kDa subunit a isoform 1	462
Q8BGW1	Fto	128
P62141	**Ppp1cb Serine/threonine-protein phosphatase PP1-beta catalytic subunit**	74
P62631	Eef1a2 Elongation factor 1-alpha 2	74
P63328	**Ppp3ca Isoform 1 of Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform**	70
Q9D6P8	Calml3 Calmodulin-like protein 3	68
Q99JX4	Eif3m Eukaryotic translation initiation factor 3	64
O88543	Cops3 COP9 signalosome complex subunit 3	62
P06240	Lck Proto-oncogene tyrosine-protein kinase	60
Q76MZ3	**Ppp2r1a Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform**	57
Q9WUA3	Pfkp Isoform 1 of 6-phosphofructokinase type C	54
Q91ZX7	Lrp1 Prolow-density lipoprotein receptor-related protein 1	51
Q99L43	Cds2 Phosphatidate cytidylyltransferase 2	49
P56480	Atp5b ATP synthase subunit beta, mitochondrial	48
Q8K0T0	Rtn1 Reticulon-1	43
Q6P5H2	Nes Isoform 1 of Nestin	43
P13595	Ncam1 Isoform 1 of Neural cell adhesion molecule 1	42
P80318	Cct3 T-complex protein 1 subunit gamma	41
O54865	Gucy1b3 Guanylate cyclase soluble subunit beta-1	41
P35802	Gpm6a Neuronal membrane glycoprotein M6-a	41
Q8BXR1	Slc7a14 Probable cationic amino acid transporter	40
Q9JJ59	Abcb9 Isoform 1 of ATP-binding cassette sub-family B member 9	40
Q7TMB8	Cyfip1 Isoform 2 of Cytoplasmic FMR1-interacting protein 1	40

Table 3-9: IP proteins from brain tissue lysate of homozygous I367F mouse. Bold text: serine-threonine phosphatases, Yellow shading: ATP related proteins; Blue shading: FTO.

Proteins identified in the negative IP reaction were deleted from the lists of proteins considered, and for analysis a score cut off of 40 was selected to remove proteins identified with low probability. FTO is identified in both of the positive IP samples, but unfortunately no significant overlap is observed with results from previous IP reactions for tagged species in cultured cells. Proteins identified in both positive IP samples

include serine-threonine proteases, and ATPase/ATP related proteins. Significantly more unique proteins were identified for the IP from I367F mice, however none of these appear to be related to the metabolic profile observed for these mice (Section 2.10)[242], and low scores are obtained for many of the proteins, suggesting a low likelihood that these are true hits.

3.10.2 FTO overexpression model

A further line of investigation at MRC Harwell was development of mouse models overexpressing FTO, to produce mice with either one additional (three copies of mFTO in total, FTO3) or two additional (four copies of mFTO in total, FTO4) copies of the FTO gene. Mice with four copies of the FTO gene have increased fat mass, due to increased food intake[243]. Additionally, mice overexpressing FTO have reduced glucose tolerance and altered plasma biochemistry, with elevated levels of fatty acids, and reduced liver function[243]. A mouse knockout model was also available (Dr Chris Church, MRC Harwell, unpublished), allowing comparison to samples with no mFTO expression.

To gain information about FTO across the spectrum of FTO production levels, a comparison was made between mice with zero, two or four copies of the FTO gene. Previous analysis of mFTO mRNA expression in tissues of the different mouse models suggested that skeletal muscle showed the greatest elevation of mFTO in the FTO4 mice[243], therefore this tissue was studied here, to maximise the mFTO quantity present.

A further complexity when working with tissue samples is the requirement for optimisation of lysis conditions for each tissue type. For this work an extraction method previously optimised by Dr Chris Church for extraction from skeletal muscle with the CelLytic™ NuCLEAR™ Extraction Kit (Sigma Aldrich) was used. Nuclear protein extracted from 100 mg skeletal muscle was used as the input for each IP reaction. Anti-mFTO IP reactions were carried out using the Eurogentec antibody, in the same manner as for previous anti-FTO IPs, but this time using magnetic Protein G Dynabeads, which were added to the lysate/antibody mixture.

Alongside a negative IP reaction on lysate from tissue from a wildtype mouse, the IP reaction carried out on cell lysate from FTO knockout mice was a further control. As no FTO is present, any interactions of the anti-FTO antibody with other proteins in solution will be non-specific, so immunoprecipitated proteins are likely false positives.

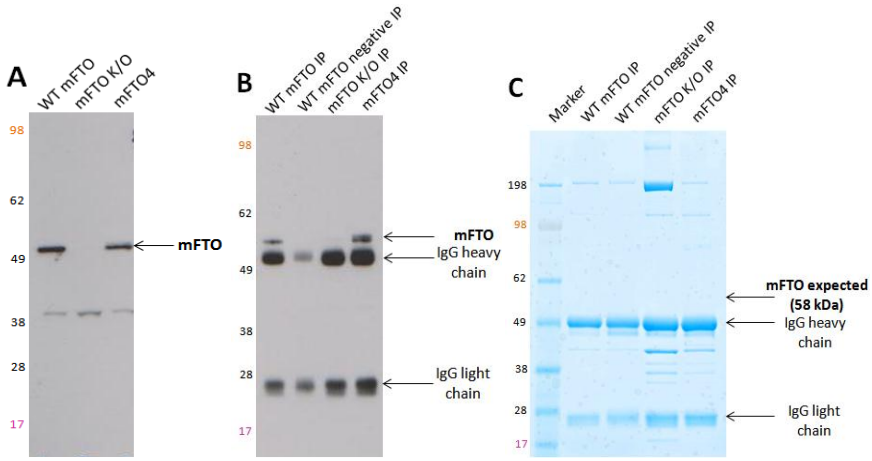


Figure 3-40: (A) Western blot showing level of FTO expression in different mouse models, (B) Western blot of immunoprecipitated protein samples, (C) IP products visualised by SDS-PAGE and stained using Colloidal Coomassie. 1 min exposures. k/o: mFTO knockout.

A western blot of the nuclear lysates of muscle samples from the different mouse models suggests that, despite the known overexpression of mFTO in mice with four copies of the FTO gene, the level of mFTO present in the nuclear extract does not appear to contain greatly elevated FTO (Figure 3-40A), although this could be due to unequal gel lanes loading. Additionally, a further band at ~40 kDa is visible in each lane of the blot, which could be due to non-specific detection of a different protein present in the nuclear lysate. Analysis of the immunoprecipitated samples by western blot shows that a small quantity of mFTO has been immunoprecipitated in this reaction, with an increased quantity of mFTO present in the FTO4 overexpression IP sample. Unfortunately however, the quantity of mFTO in the IP samples is quite low, as mFTO is not visible on the Colloidal Coomassie stained SDS-PAGE gel (Figure 3-40C).

In addition to the intense antibody bands present on the gel at ~50 kDa and ~25 kDa (Figure 3-40C), several additional bands are visible. Unexpectedly, significantly more bands are present in the mFTO knockout IP lane than in the other lanes. All bands present in the positive reaction lanes (WT mFTO and mFTO4) are also present in at

least one of the control lanes. In particular, the band at ~200 kDa is particularly intense in the mFTO deletion IP sample. Although proteins present at lower concentrations than are visible on stained gels can be observed by mass spectrometry, the presence of other proteins present at high, detectable levels when mFTO is not detectable suggests that this IP reaction has not been particularly successful, due to the non-specific detection of many other proteins. Therefore these samples were not further analysed by trypsinolysis. Overall it appears that IP of FTO from mouse tissues should be successful, but for effective IP reactions from cells a more efficient anti-FTO antibody is necessary.

3.11 Discussion

This chapter has described the production of a range of constructs, including point mutants and truncated constructs expressing individual protein domains, enabling cell based analyses of FTO. The majority of the constructs containing the *N*-terminal NLS sequence are predominantly targeted to the nucleus, the location of native FTO protein, although proteins with amino acid substitutions or truncations were transfected less efficiently than their wildtype counterparts.

Several literature methods were attempted to optimise nuclear extraction from cells overexpressing FTO, although extracts corresponding to the nuclear portion were difficult to obtain, and were not appropriate for successful IP reactions. IP protocols were tested and developed to enable investigations of the protein species with which the FTO protein interacts in cells.

Large scale IP reactions were carried out for FTO tagged with GFP at either the *N*- or *C*- terminus, or with an *N*-terminal 3xFLAG tag, using antibodies directed to the fusion tags, allowing proteomic analysis of the proteins interacting with FTO in cells. Although a large number of proteins were identified in the samples corresponding to binding to FTO, the scores obtained for the majority of these proteins were significantly lower than those for the target protein, often by orders of magnitude. Although, to enable analysis, FTO was greatly overexpressed in cells, whereas any interacting

species would be present at the normal endogenous level, it would be expected that any true interacting species would be enriched in the IP sample. A shortlist of proteins identified in multiple IP reactions was compiled, as proteins identified independently are less likely to interact non-specifically with the tag. Interaction with the top hit, the dihydrolipoyllysine-residue succinyltransferase component of the 2OG dehydrogenase complex was investigated, but this protein could not be identified in IP samples by western blot. However, as the normal location of this protein in the mitochondria, it is unclear how it may interact with FTO, an exclusively nuclear protein.

IP studies were also attempted for lysates obtained from FTO mouse model tissues, but unfortunately these did not yield similar proteins to those for tissue culture samples, and again difficulties were encountered with use of the FTO antibody in IP reactions.

As no clear protein hits were identified by these analyses, it is likely that FTO may not interact with other proteins *in vivo*, and may interact alone with its DNA or RNA substrates, acting via either a repair or a regulatory mechanism.

These initial studies have initiated cell based analyses of FTO to study its interactions with other cellular species, and thus to understand more about the nature of its potential roles in obesity, weight control, metabolism and the DNA damage response. Unfortunately the lack of a high-quality IP grade antibody precluded more detailed analysis of endogenous protein, or ChIP analysis. Therefore future studies should be directed towards development of more robust anti-FTO antibodies, and generation of mammalian cell lines stably expressing tagged FTO, work which is currently in progress.

Chapter 4 – *In vitro* and Cell-Based Studies of ABH Proteins

4.1 Introduction

Although ABH2 and ABH3 have been well studied (see Chapter 5), and the function of ABH8 has recently been elucidated, little is known about the roles of the remaining ABH family members. Additionally, the wider in-cell roles of the proteins are poorly understood. Learning more about the regulation and recruitment of the ABHs, and any other species in cells with which they may interact, is likely to assist with assigning a function for the proteins of currently unknown function.

ABH1 was the first human analogue of AlkB to be identified[162], and phylogenetic analysis suggests that of the nine currently identified AlkB homologues this has the highest similarity to AlkB in its catalytic domain[164]. Despite several groups being unable to observe any activity for ABH1 with ABH2/3 substrates[163,167], demethylation of 3meC, and inefficient 1meA demethylation, has been reported for ABH1[210], in addition to possible DNA lyase activity[211]. ABH1 was determined to localise in the mitochondria[210], possibly via an *N*-terminal targeting motif, although ABH1 is not identified in the "Mitocarta", a summary of all likely human mitochondrial proteins, determined by immunofluorescence and proteomics[324]. To attempt to summarise the current understanding of ABH1 in the literature, Table 4-1 contains information from each of the papers published to date describing the *in vitro* activity of ABH1 The table includes a short summary of the results of the papers, the methodologies used, and a brief discussion of the quality and relevance of the results obtained.

No clear-cut conclusions as to the role of ABH1 can be drawn from the two ABH1 mouse knockout studies[206-207], however one of these studies found an interaction with the histone deacetylase associating protein Mrj, suggesting a possible epigenetic role through histone modification[206]. Therefore further studies are required to more fully understand the function of ABH1.

Reference	Description	Construct used	Assay summary	Assay conditions	Subcellular localisation	Paper quality/comments
Wei et al., 1996	Initial description of ABH1 gene	*N*-terminal His-tag (pQE-9)	Transfection of ABH1 into MMS-treated AlkB knockout *E. coli* strain causes slight increase in survival.	Colonies of AlkB deficient *E. coli* transformed with ABH1 counted, after growth on plates containing 0-0.05% MMS.	n/a	ABH1 activity unclear, as increased *E. coli* survival not significant. Controls insufficient to determine whether survival attributable to ABH1 itself, or due to His-tag/ transfection process.
			No induction of ABH1 in MMS-treated human skin fibroblasts (although observed with AlkB) possibly suggesting a different regulation mechanism for ABH1.	ABH1 RNA quantity unaffected in skin fibroblasts (CCD-965SK) treated with 0-0.08% MMS.		
Duncan et al., 2002	Identification of demethylation of 1meA and 3meC in DNA by ABH2 and ABH3	*N*-terminal His-tag (pET-15b)	Transfection of ABH1 into AlkB knockout *E. coli* strain containing MMS-treated ssDNA phage does not complement the wildtype AlkB phenotype.	AlkB deficient *E. coli* transformed with pQE-9 ABH1 and grown with 100 µM IPTG for 3 h at 25 °C to induce ABH1 expression. Further transfected with MMS-treated M13 ssDNA phage, and phage reactivation monitored by colony counting.	n/a	Standard assays for AlkB, ABH2 and ABH3 do not show ABH1 activity, despite discussion that optimisation of assay conditions for ABH1 was attempted.
			Recombinant ABH1 does not release radioactivity from ^{14}C MeI treated poly(A) or poly(C) oligonucleotides, although release seen for AlkB, ABH2 and ABH3.	2 pmol AlkB, [14C] MeI-treated poly(A)/(C) (1200 cpm), 50 mM HEPES pH 8.0, 1 mM 2OG, 75 µM $Fe(NH_4)_2(SO_4)_2$, 2 mM ascorbate, 50 µg/ml BSA. 15 min, 37 °C. 11 mM EDTA quench, EtOH soluble material monitored by scintillation counting.		
Aas et al., 2003	Identification of demethylation of 1meA and 3meC in RNA by AlkB and ABH3	*N*-terminal His-tag (pET-28a)	No release of radioactivity from ^{3}H methylated ss or dsDNA or ssRNA by recombinant ABH1, although activity observed for AlkB, ABH2 and ABH3.	0.1 µg DNA/RNA, 30 min, 37 °C. EtOH precipitation, monitored by scintillation counting.	n/a	Standard assays for AlkB, ABH2 and ABH3 do not show ABH1 activity. Untagged ABH1 inactive, suggesting that His-tag used elsewhere does not interfere.
		Untagged, low copy-number pJB658 toluic acid inducible vector	Expression of ABH1 in AlkB deficient *E.coli* does not reactivate methylated DNA or RNA phages.	pJB658-ABH1 transformed into AlkB deficient *E. coli*, and expression induced for 3 h at 37 °C. Treated with MMS-methylated phage and colonies counted.		
Mishina et al., 2004	DNA/protein crosslinking studies using active site cytosine mutants	*N*-terminal His-tag (pET-28a)	ABH2 and ABH3 proteins crosslink specifically to modified 20mer ss/ds DNA oligonucleotides, but no binding observed for corresponding ABH1 protein.	Thiol tethered 20mer oligonucleotide, incorporating O^4-triazolyl-dU. 10 µM H231C ABH1 in 10 mM Tris HCl pH 7.34, 100 mM NaCl, incubated with oligonucleotide at 4 °C. Monitored by gel electrophoresis.	n/a	No discussion of ABH1 activity, but appears that ABH1 protein does not interact with ABH2/ABH3 DNA substrates.
Tsujikawa et al., 2007	Expression and subcellular localisation study of ABH constructs	*N*-terminal GFP fusions (pcDNA6.2/N-EmGFP-DEST) in HeLa cells	Endogenous ABH1 mRNA present in all 16 human tissues tested.	RT-PCR using specific primers.	EmGFP-ABH1 expressed throughout the cell.	GFP fusion at only one protein terminus for localisation studies, so slightly misleading.
Westbye et al., 2008	Identification of ABH1 as a mitochondrial protein, demethylating 3meC in ssDNA and ssRNA	*N*-terminal His-tag (pET-28a) - poor yield, therefore N-terminal Intein tagged ΔN3 ABH1 used instead, with tag cleaved (pTYB12)	ABH1 catalysed decarboxylation of 2OG, activity reduced for proteins containing point mutations of Fe/2OG binding residues.	2OG turnover: 100 µM 2OG (10% 1-[^{14}C]), 80 µM $Fe(NH_4)_2(SO_4)_2$, 4 mM ascorbate, 0.5 mg/ml catalase, 50 mM tris HCl pH 7.0. 1 h, 37 C. 50% TFA quench. $^{14}CO_2$ monitored by scintillation counting.	n/a	2OG oxygenase activity identified, with good controls. 3meC demethylation low (~4 fold less than observed for ABH3), therefore 3meC may not be the true substrate. Poly(rC) unstable therefore high background in assay, making activity unclear. No inhibitor controls. Strong evidence for mitochondrial localisation. Unable to replicate in vivo findings of Wei et al. (*E. coli* survival increase), again showing conflicting results.
			Low levels of 3meC demethylationof ssDNA/RNA observed, activity reduced for proteins containing point mutations of Fe/2OG binding residues.	[^{3}H] methylated poly(dC)/(rC) 50 mM .tris HCl pH 7.0, 100 µM $Fe(NH_4)_2(SO_4)_2$, 100 µM 2OG, 4 mM ascorbate/DTT. 20-60 min, 37 °C. EtOH soluble material monitored by scintillation counting		
				MMS methylated poly(dC)/(rC) 50 mM tris HCl pH 7.0, 10 µM $Fe(NH_4)_2(SO_4)_2$, 100 µM 2OG, 4 mM ascorbate/DTT. 20-60 min, 37 °C. Substrates digested to nucleosides and monitored by LC/MS/MS.	n/a	
		N-terminal EYFP fusion (pEYFP-N1) in HeLa cells	n/a	Transfection using FuGENE 6, then fixation, mounting and direct visualisation of YFP fluorescence.	EYFP-ABH1 predominantly nuclear, with some cytoplasmic staining.	
		C-terminal EYFP fusion (pEYFP-N1) in HeLa cells	n/a		Wildtype ABH1-EYFP and *C*-terminally truncated 1-119 ABH1-EYFP localised in mitochondria. ΔN27 ABH1-EYFP expressed diffusely throughout the cell.	
Muller et al., 2009	Identification of ABH1 as a DNA lyase acting at abasic sites	*N*-terminal His-tag (pET-28a for *E. coli* expression; pAcHLTb for baculovirus expression)	Purified ABH1 binds unmodified and methylated DNA and RNA.	Electrophoretic mobility shift assays on 1 % agarose gels; 10 µM ABH1, 0.3 µg linearised pUC18 DNA.	n/a	DNA interaction inconsistent with Mishina et al. Unclear how the lyase activity occurs as not due to 2OG oxygenase, and no cofactors or protein domains responsible identified. Possibly due to contaminating species, e.g. DNase or oxidative species - not fully ruled out in paper.
			MNNG methylated DNA degraded on incubation with ABH1 - likely due to spontaneous formation of abasic sites, which are lysed by ABH1, with product lacking 3'-phosphate. No reaction with reduced abasic site analogue, as β-elimination reaction intermediate not possible.	10 µM ABH1 with 0.5 µg MNNG-treated linearised pUC18 DNA; 3 h, 37 °C. Reaction independent of addition of ascorbate, iron, 2OG and EDTA, therefore not a 2OG oxygenase activity. *N*-/*C*-terminal ABH1 truncations insoluble, so not studied. Activity observed for proteins expressed from *E. coli* and baculovirus.		

Table 4-1: Summary of previous studies of ABH1, including a brief description of assays, and a comment on methodology and relevance of data obtained.

Very little is known about ABH7 in the literature. Despite the apparent identification of GFP-ABH7 throughout the cell[209], it is possible that ABH7 may also be a mitochondrial protein, as it is identified in the human Mitocarta[324]. The only study discussing *in vitro* analysis of recombinant ABH7 found that murine ABH7 has no activity with ABH2/3 substrates[173].

Both ABH1 and ABH7 are ubiquitously expressed in human tissues, and show the highest expression levels of mRNA in the spleen and pancreas, and are lowest in skeletal muscle[209]. As it is possible that, uniquely within the ABH family, ABH1 and ABH7 are both mitochondrial proteins, these were here studied in parallel. Recombinant ABH1 and ABH7 proteins were prepared, characterised, and activity studies were initiated. Proteomic analyses were carried out to attempt to identify proteins with which the proteins may interact in cells, to attempt to gain an insight into their functions.

4.2 Construct production

To enable recombinant protein production and expression of the proteins in mammalian cell lines, constructs were prepared containing the ABH1 and ABH7 sequences. Initial work on these proteins was carried out in collaboration with Alexander Drong.

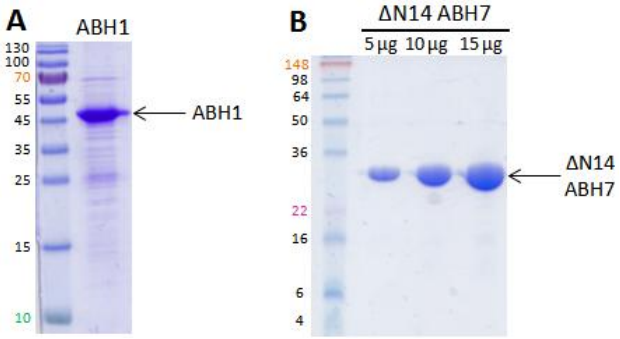


Figure 4-1: SDS-PAGE gel loaded with (A) 5 µg recombinant His-tagged ABH1, (B) recombinant ΔN14 ABH7, quantities as indicated.

A pET-28a (+) plasmid containing an *ABH1* cDNA fragment cloned into the *NheI/HindIII* restriction sites was provided by Dr Kirsty Hewitson. Alexander Drong optimised a purification protocol, with greatest ABH1 production observed in BL21 (DE3) cells incubated for 16 h at 28 °C, post induction with 0.5 mM IPTG.

Protein yields were approximately 0.7 mg per 1 l of culture, at ~90 % purity by SDS-PAGE (Figure 4-1A).

An expression vector for ABH7 was provided by Dr Stan Ng (Structural Genomics Consortium, Oxford) in the form of a pNIC28-Bsa4 vector containing the ABH7 insert, which was introduced *via* ligation independent cloning. The pNIC28-Bsa4 vector is based on the pET-28a(+) vector and also contains an *N*-terminal His-tag. Expression trials of the full length ABH7 showed good expression levels, but solely of insoluble protein. Following comparison to the sequences of crystallisation constructs for AlkB[192], ABH2[194] and ABH3[193], each of which are *N*-terminally truncated, an ABH7 construct containing residues 15-221 (ΔN14 ABH7) was developed by Alexander Drong, which comprised an analogous *N*-terminal truncation (Figure 4-2). This construct gave significant soluble protein expression, and therefore is used throughout when recombinant ABH7 protein is discussed. Optimal expression was obtained in *E. coli* incubated at 37 °C for 3 h post induction with 0.5 mM IPTG, resulting in an overall yield of 2 mg per 1 l of culture. The protein was judged to be >95 % pure by SDS-PAGE analysis (Figure 4-1B).

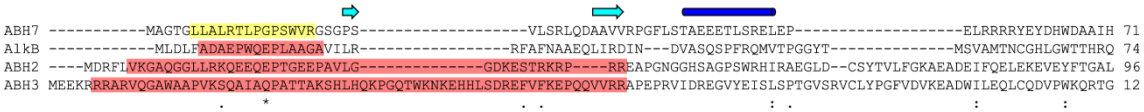


Figure 4-2: Alignment of ABH7, AlkB, ABH2 and ABH3 with prediction of ABH7 secondary structure. The secondary structure as predicted by Jpred3 is shown above the ABH7 sequence. Light blue arrows: β-sheet; dark blue cylinders: α-helix. Amino acids deleted for AlkB[192], ABH2[194] and ABH3[193] are highlighted in light red, with amino acids deleted for ABH7 highlighted in yellow. Figure prepared by Alexander Drong.

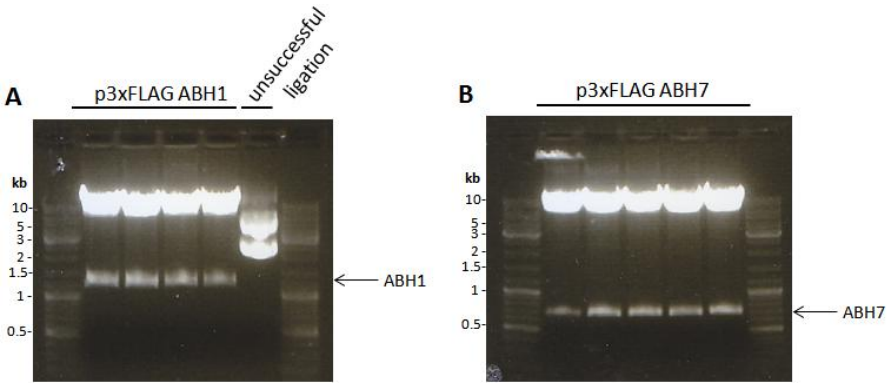


Figure 4-3: 1 % Agarose gels verifying of the presence of the (A) ABH1 and (B) ABH7 inserts in p3xFLAG CMV10 constructs.

The full length ABH1 and ABH7 constructs were cloned into the EGFP-C1 and EGFP-N1 constructs by Alexander Drong, allowing mammalian protein expression. The gene sequences were subcloned into the p3xFLAG CMV10 vector from the corresponding EGFP-C1 constructs via the EcoRI and KpnI restriction sites, allowing production of proteins expressing the *N*-terminal 3xFLAG sequence (Figure 4-3).

4.3 Recombinant protein investigation summary

4.3.1 Protein characterisation

Initial characterisation of the recombinant proteins was carried out in collaboration with Alexander Drong. Analytical gel filtration of the purified proteins showed that ΔN14 ABH7 was monomeric, whereas ABH1 exists predominantly as a monomer, with a low proportion of dimer. Circular dichroism suggests that the proteins were both well folded at 4 °C and 37 °C, displaying a high proportion of β-strand character as expected for 2OG oxygenases containing the DSBH fold.

T_m shift/°C	ABH1	ΔN14 ABH7
No metal	0.00	0.00
5 % DMSO	-0.09	-1.81
Mn(II)	0.66	-1.16
Fe(II)	0.84	-2.16
Co(II)	4.19	9.87
Ni(II)	5.83	-1.12
Cu(II)	n/a	-2.19
Zn(II)	0.38	-1.17

Table 4-2: T_m shifts obtained by DSF for ABH1 and ΔN14 ABH7, with addition of 5 % DMSO, or 50 µM metal (II).

The melting temperatures of the proteins with added metal (II) species were determined using DSF (Table 4-2). The melting temperatures of the two proteins are very similar, at 48.4 °C for ABH1 and 48.5 °C for ABH7, however the metal stabilisation properties appear to be different. ABH7 is destabilised by addition of most metals tested, but is significantly stabilised on addition of Co(II). However, stabilisation of ABH1 occurs for each of the metals added, with the greatest stabilisation occurring on addition of Ni(II)

and Co(II). As good melting profiles are observed for the proteins this DSF method could be used successfully for analysis of binding of 2OG mimics or other small molecules to the proteins.

4.3.2 Nucleic acid substrate investigation

As the true substrate of ABH1 is unclear, and ABH7 currently does not have a known substrate, investigation of a range of potential substrates was carried out by Alexander Drong. The 2OG turnover assay is useful for determination of activity with an unknown substrate, as a range of potential substrates may be incubated with the protein of interest. Elevation of 2OG turnover suggests that the added reagent is either directly modified by the enzyme, or that it stimulates uncoupled 2OG turnover by binding to the protein, which in either case can yield useful information about the protein's activity.

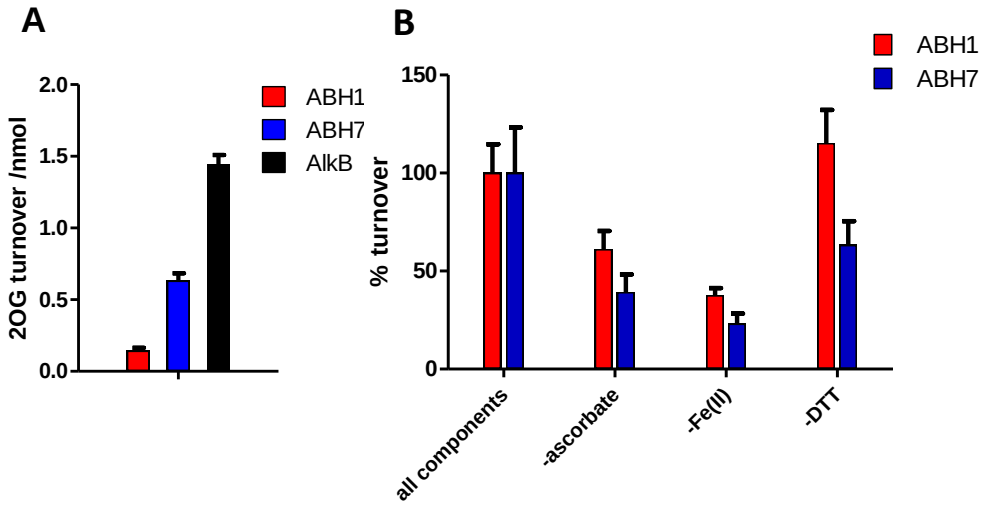


Figure 4-4: 2OG turnover of ABH proteins. (A) nmol 2OG turned over under standard assay conditions; (B) % 2OG turnover calculated relative to standard assay conditions (addition of all assay components). 30 minute incubations.

Both ABH1 and ABH7 recombinant proteins showed 2OG turnover in the absence of any potential substrates, although to a lesser extent than AlkB (Figure 4-4A). This turnover was found to be dependent on the presence of the assay components Fe(II) and ascorbate, with ABH7 having a greater co-substrate dependence than ABH1 (Figure 4-4B). The activity of ABH1 appeared to be slightly stimulated in the absence of DTT, suggesting that this alternate reducing agent may have a slight inhibitory effect on ABH1. Studies on PHD2, which has a high dependence on ascorbate for optimal

hydroxylation activity, have shown that DTT is not able to fully replace ascorbate in promoting activity[293].

As the ABH proteins with currently identified substrates are nucleic acid modifying enzymes, 2OG turnover in the presence of a range of singly methylated nucleosides (Appendix A) was initially studied. As a control, AlkB was also incubated with the tested methylated nucleosides, as it is known that 2OG turnover is elevated in the presence of the 1meA nucleoside[325].

The previously identified substrates of AlkB and ABH2/3, 1meA, 3meT and 3meC were tested to determine whether the substrate range of ABH1 and ABH7 overlaps with that of the previously studied enzymes. The corresponding 1meG nucleoside was not commercially available therefore was not tested here. Further methylated nucleosides were tested alongside these, including the epigenetic markers 5meC and N^6meA, the RNA capping modification 7meG, and the highly carcinogenic lesion O^6meG. Incorporation of damaged DNA bases is prevented by sanitation of the nuclear and mitochondrial dNTP pools, usually through hydrolysis to the monophosphates[128], but it is possible that a direct repair methodology may also be employed for this purpose, preventing the necessity of base-resynthesis.

As expected, the 2OG turnover of AlkB was stimulated by addition of the 1meA nucleoside (Figure 4-5), but no stimulation was observed on incubation with other methylated nucleosides, including a 3meC nucleoside[177,325]. However this was not true for incubation of ABH1, as neither 3meC nor 1meA cause an increased 2OG turnover. This is potentially surprising as a previous report has reported that 1meA is a substrate of ABH1, although the preferred substrate is 3meC[210]. However, in other previous studies no activity was observed with 1meA and 3meC oligonucleotide substrates[163,167]. No elevation of 2OG turnover was observed with the other methylated nucleosides tested, however this does not preclude the existence of a nucleic acid substrate for either ABH1 or ABH7. Many further nucleic acid modifications are known, particularly in tRNA, in which to date over 100 modifications have been identified[82]. Additionally, a

longer sequence may be required in order to create the required interactions between the substrate and protein, as the activity of AlkB is significantly increased with longer substrates[326].

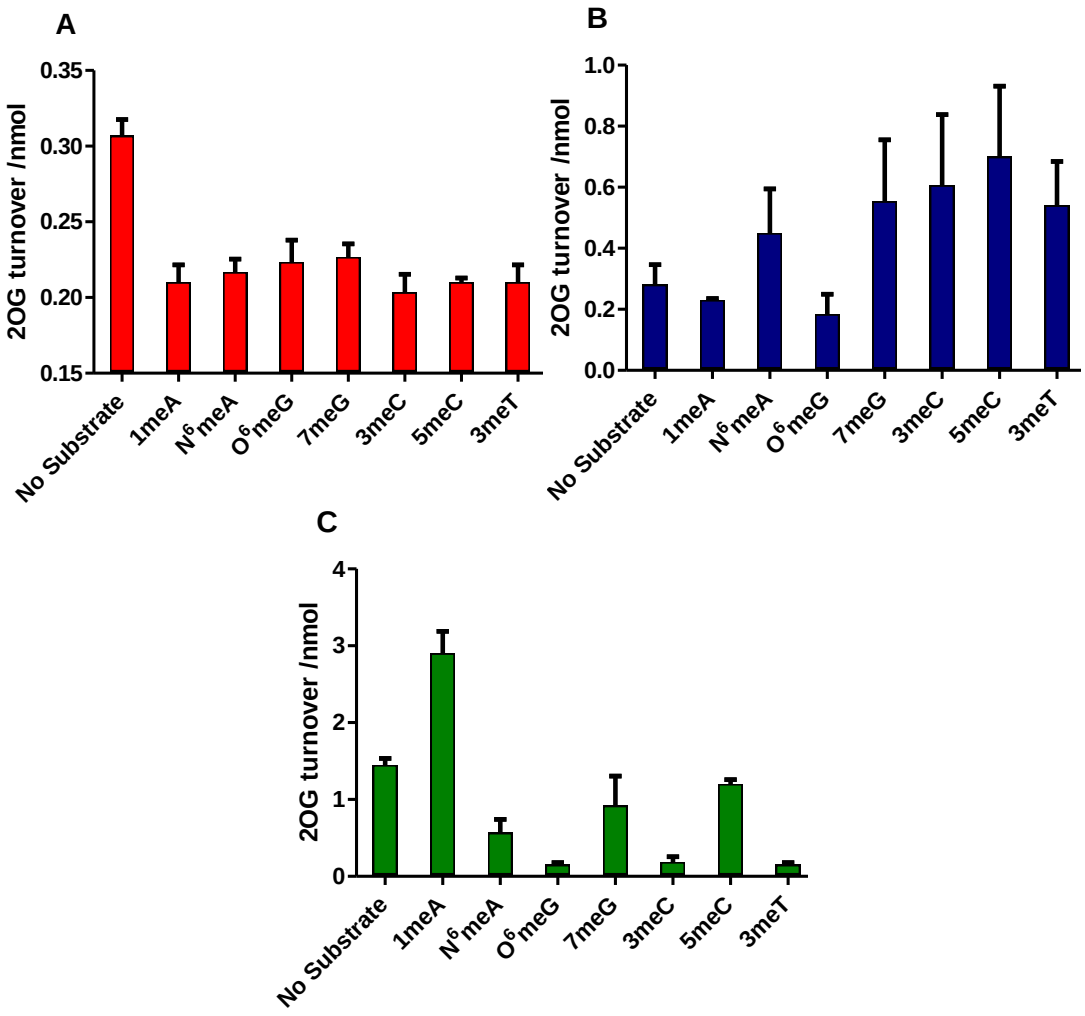


Figure 4-5: 2OG turnover of (A) ABH1, (B) ABH7 and (C) AlkB proteins with nucleosides. Data represent the mean of triplicate experiments, with error bars showing the standard error of the mean. Incubation time 30 min.

4.3.3 Histone substrate investigation

As a number of human 2OG oxygenases have histone substrates[25-26], the possibility of a histone derived substrate for ABH1 and ABH7 was also investigated by Alexander Drong. Additionally, a previous publication describing a mouse ABH1 deletion suggested that from the observed phenotype it is likely that ABH1 has a histone substrate[207]. Initially a 2OG turnover assay was carried out with addition of a bulk histone preparation, although no significant elevation of 2OG turnover was observed. This does not conclusively show that histones are not a substrate of the two proteins, as

lack of activity could be due to the lack of a required binding partner in the *in vitro* reaction mixture, or the use of a sub-optimal buffer or pH.

An alternative approach was taken, testing binding of modified histone peptides to the proteins, using AlphaScreen™ (Perkin Elmer) technology. This chemiluminescence-based approach has low reagent consumption yet high sensitivity, and has been optimised for analysis of the JMJD histone demethylase family[327]. Binding to a library of *N*-terminally biotinylated histone peptides (Alta Bioscience) was studied. As the majority of the peptides in the library have modified amino acids (including acetyl-Lys/Arg, methyl-Lys/Arg, and phospho-Ser) it can be determined whether the proteins recognise these histone marks, either as a substrate or as a binding partner. Emission of light occurs upon binding of the peptide to the proteins, which is detected as counts per second. Unfortunately no significant elevation of counts above baseline was observed for either ABH1 or ABH7, particularly when compared to results observed for the known histone demethylase JMJD2E (counts for binding peptides are typically 3 orders of magnitude greater than nonbinders/baseline). This, together with the lack of 2OG turnover, suggests that neither ABH1 nor ABH7 has a histone substrate.

As no definitive results were obtained from the initial investigations of the recombinant ABH1 and ABH7 proteins, a complementary approach was initiated, to study the proteins in mammalian cell lines.

4.4 ABH1 antibody

When these studies were initiated no ABH7 antibodies were commercially available, however an anti-ABH1 mouse monoclonal antibody was available from Sigma-Aldrich, which was tested in a range of applications.

Initially, whole cell lysates from 293T cells transfected with either EGFP-ABH1 or EGFP-hFTO were analysed by western blot. As a positive control to verify that the overexpressed proteins were present in the lysates, one pair of samples was probed using anti-GFP (Figure 4-6, α-GFP lanes), with bands observed at the expected masses of the GFP-tagged proteins. The same samples were probed using the α-ABH1

antibody, using a range of dilutions recommended by the supplier (Figure 4-6). Bands corresponding to both EGFP-ABH1 and ABH1 can be seen at each of the dilutions tested, suggesting that the *C*-terminal EGFP tag does not interfere with the recognition of ABH1 by the antibody. The detection of ABH1 using the protein-specific antibody also appears to be more efficient than when using the anti-GFP antibody, as observed by the more intense bands on probing with α-ABH1. Cells overexpressing hFTO are also likely to express endogenous ABH1, and a band at the expected mass of untagged ABH1 can be observed on detection with a 1 in 500 α-ABH1 dilution.

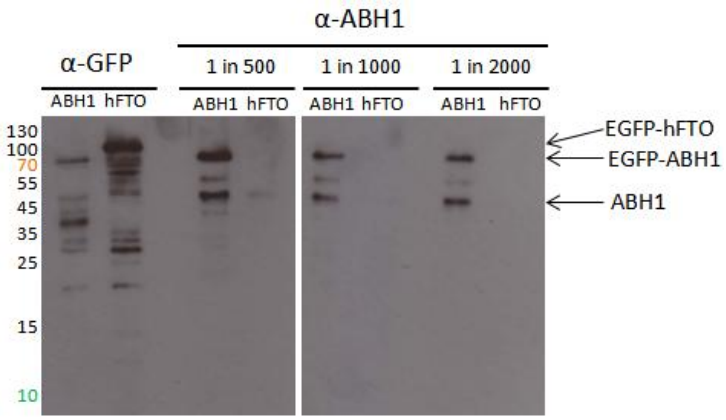


Figure 4-6: Test of detection of ABH1 in cell lysates by commercial ABH1 antibody. Cell lysates from 293T cells transiently transfected with either ABH1-EGFP (lanes marked ABH1) or EGFP-hFTO (lanes marked FTO). Detection using α-GFP or a range of concentrations of α-ABH1. Exposure 30 min.

Detection of ABH1 in a cell lysate overexpressing FLAG-ABH1 was also attempted, with detection successful in the presence of the 3xFLAG tag (Figure 4-7). Furthermore, recombinant ABH1 was detected successfully across a range of concentrations. This also shows that several of the bands running at lower mass than the mass of ABH1 in the SDS-PAGE gel of the recombinant protein (Figure 4-1A) are due to protein degradation, as they are also recognised by the α-ABH1 antibody (Figure 4-7).

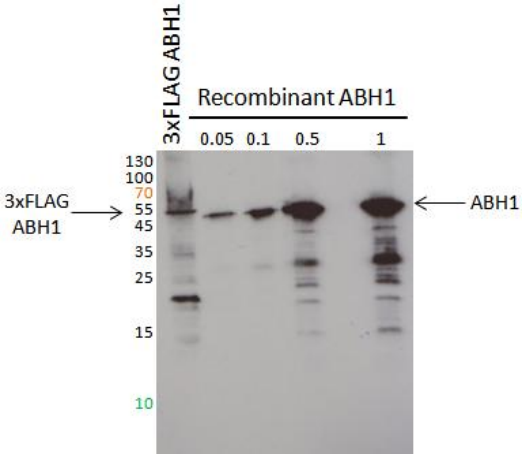


Figure 4-7: Test of detection of ABH1 using commercial ABH1 antibody. Lane 1: Detection of 3xFLAG-ABH1 from cell lysate of 293T cells overexpressing 3xFLAG-ABH1; Remaining lanes: detection of recombinant ABH1, quantities as indicated. 1 min exposure.

Given the successful detection of denatured ABH1 protein constructs in western blots, a test IP reaction was carried out. A whole cell lysate from cells overexpressing FLAG-ABH1 was incubated with 40 µg α-ABH1 before incubation with Protein G beads. Samples were analysed by western blot, with detection using anti-FLAG. This IP reaction was successful, with a good proportion of ABH1 bound to the beads (Figure 4-8). Due to the expense of the antibody, this approach was not used for initial proteomic studies, in which tagged proteins can be used successfully. However access to an IP grade α-ABH1 antibody is very useful for potential future ChIP analyses, or IP of endogenous proteins.

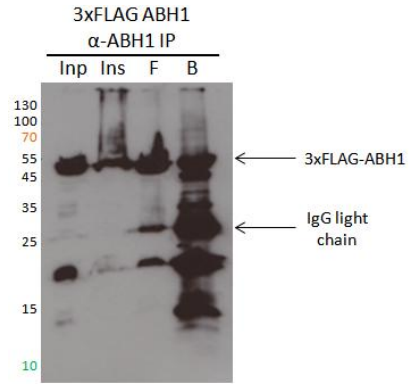


Figure 4-8: Test immunoprecipitation of 3xFLAG-ABH1 using commercial ABH1 antibody. 1.5 x 10^7 293T cells transiently transfected with 3xFLAG-ABH1. Inp: IP input; Ins: Insoluble protein; F: IP flow through; B: beads. 1 min exposure.

4.5 Subcellular localisation

4.5.1 Subcellular localisation of ABH1

Previous localisation studies carried out for ABH1 have been contradictory. The disparity between localisation sites is likely due to the use of large tags which may interfere with the subcellular targeting. In the report assigning ABH1 as a nuclear protein[206], the authors studied the localisation of ABH1 *N*-terminally tagged with the Myc transcription factor, the presence of which may obscure the mitochondrial targeting sequence. A further *N*-terminally tagged ABH1 species, in this case tagged with GFP, was used in the study suggesting that ABH1 is found throughout the cell[209]. The mitochondrial localisation of ABH1 was assigned based on visualisation of ABH1 directly detected with an α-ABH1 antibody, and the same localisation was observed for ABH1 *C*-terminally tagged with EYFP, although the corresponding *N*-terminally tagged species was again seen to be found throughout the cell, but predominantly in the nucleus[210]. This therefore suggests that the localisation of ABH1 is highly sensitive to the addition of fusion tags, particularly at the *N*-terminus. However, analysis of the ABH1 sequence for mitochondrial targeting sequences using common software tools does not confidently identify any strong mitochondrial targeting motifs either at the *N*-terminus, or through the remainder of the sequence.

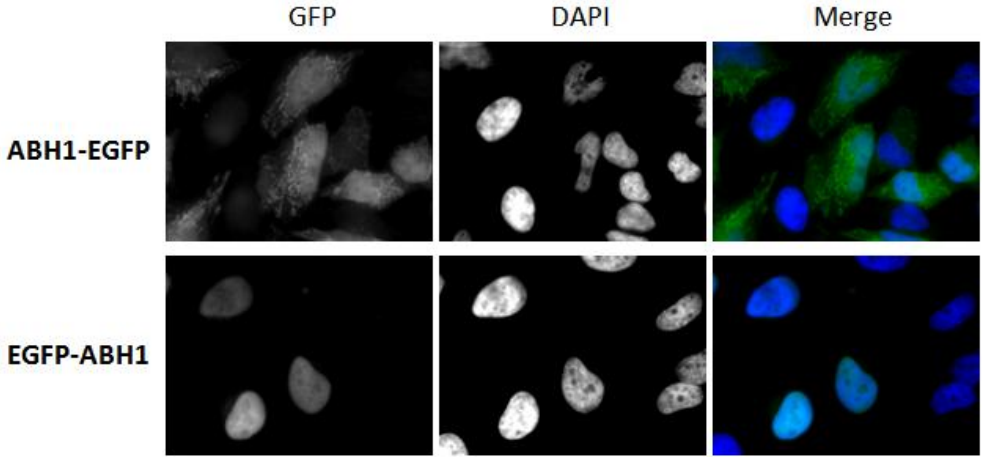


Figure 4-9: Localisation of N1 and C1 EGFP-tagged ABH1 constructs in HeLa cells, detected by direct fluorescence of GFP. Transfection efficiencies, ABH1-EGFP: 67 %, EGFP-ABH1: 66 %.

To further investigate this disparity in the literature, the localisation patterns of the available ABH1 constructs (*N*- and *C*-terminally GFP tagged, and *N*-terminally FLAG-tagged) were determined alongside the native protein, which was stained using the commercial anti-ABH1 antibody.

As observed by Westbye et al. ABH1-EGFP is present in the nucleus, whereas EGFP-ABH1 appears to be present at a low concentration in the nucleus, but with an additional punctuate pattern through the cytoplasm (Figure 4-9). FLAG-ABH1 is also found through the cytoplasm, although any distinct features are less resolvable than those observed for ABH1-EGFP (Figure 4-10). This would suggest that either the small (~2 kDa) FLAG-tag at the *N*-terminus does not interfere with the mitochondrial localisation of ABH1, but that detection via the FLAG antibody may not be sensitive enough to identify any discrete species present in the cytoplasm, or that the protein is present throughout the cell, as observed previously[209].

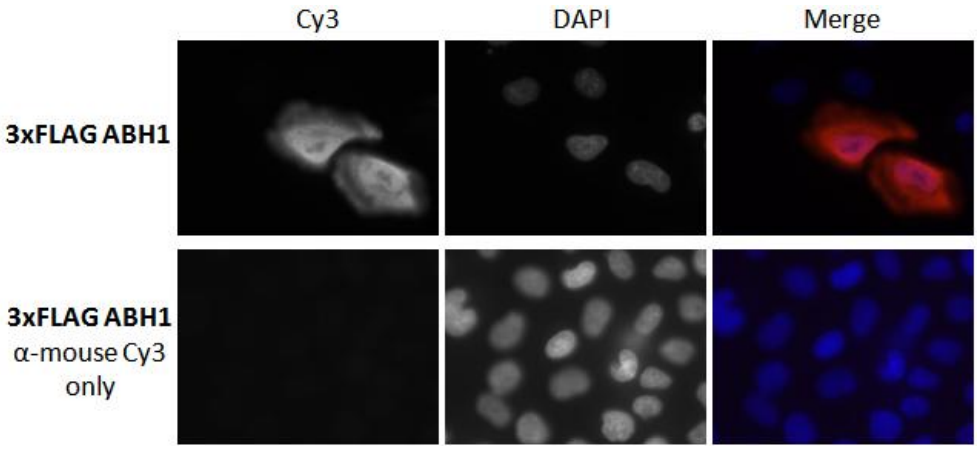


Figure 4-10: Localisation of FLAG-tagged ABH1 in HeLa cells detected using anti-FLAG antibody. Control probed using only secondary anti mouse Cy3 antibody. Transfection efficiency 33 %.

To further study the localisation of ABH1 in the cell, direct detection of ABH1 using an αABH1 antibody was carried out (Figure 4-11). Initially the antibody concentration was optimised to maximise the signal obtained above the background noise, the signal obtained on staining with the secondary antibody alone, and cross-talk between channels. Detection of endogenous ABH1 reveals only faint dots throughout the cell, but with no significant increase in intensity in the region of the nucleus.

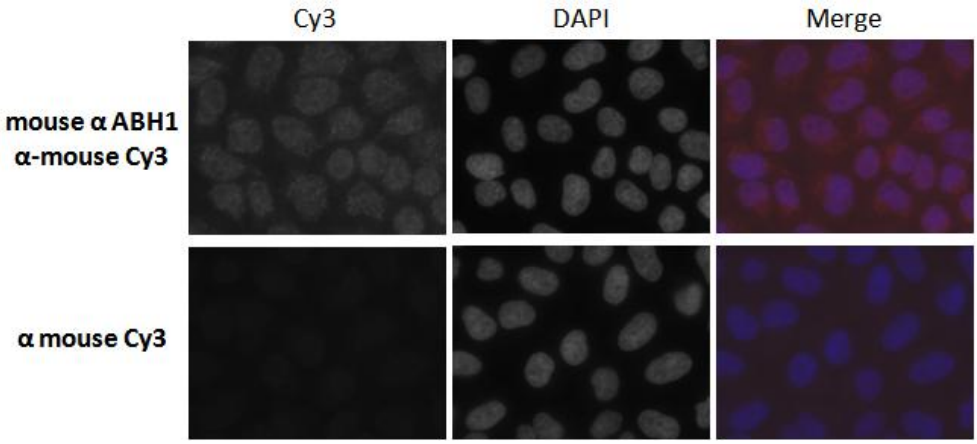


Figure 4-11: Localisation of ABH1 in HeLa cells, detected using anti-ABH1 antibody. Control probed using only secondary anti mouse Cy3 antibody.

It is possible that endogenous ABH1 is expressed at a very low level, although in the previous study of expression of ABH proteins no comment was made about relative expression levels[209]. No information about the protein/mRNA expression levels in different tissues is included in the Human Protein Atlas[328]. Endogenous ABH1 was previously observed in 293T cell lyates (Section 4.4), whereas HeLa cells have been studied here by microscopy. However, HeLa cells have also been shown to endogenously express ABH1[210]. It is unclear whether the signal observed is due to the presence of ABH1 in defined locations within the cell corresponding to certain organelles, or whether it is caused by poor staining by the antibody, which has only bound weakly to the protein present. Although the ABH1 antibody can recognise both native folded ABH1 in an IP reaction, and fully denatured ABH1 in a western blot, it may be less efficient in detection of protein fixed with formaldehyde, as the recognised epitopes may be crosslinked and thus destroyed.

4.5.2 Subcellular localisation of ABH7

Immunofluorescence of the ABH7 EGFP-tagged constructs showed that in both cases the GFP signal is present throughout the cell (Figure 4-12). However, ABH7-EGFP is present to a lesser extent in the nucleus, and the signal through the cytoplasm is punctuate, suggesting that the protein may be restricted to certain subcellular organelles. Localisation data published in the Human Protein Atlas using an anti-ABH7 antibody suggests that ABH7 is present in "vesicles" in the cytoplasm[328].

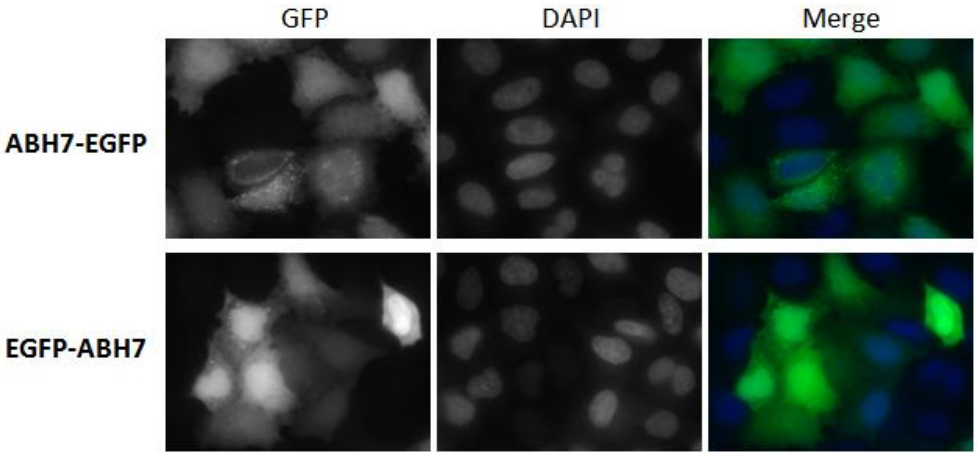


Figure 4-12: Localisation of N1 and C1 EGFP-tagged ABH1 constructs in HeLa cells, detected by direct fluorescence of GFP. Transfection efficiencies, ABH7-EGFP: 40 %, EGFP-ABH7: 76 %.

Analysis of the ABH7 amino acid sequence using MitoProt II[329] predicts an 86 % probability for the export of ABH7 to mitochondria, through identification of a likely mitochondrial targeting sequence in the first 20 *N*-terminal amino acids. Therefore it is possible that fusion of the large GFP-tag to the *N*-terminus of ABH7 prevents detection of the targeting sequence.

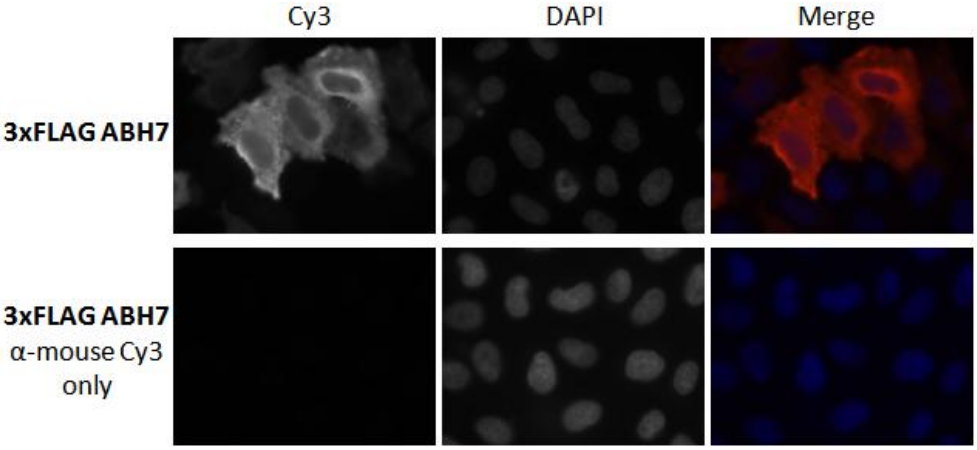


Figure 4-13: Localisation of FLAG-tagged ABH7 in HeLa cells detected using anti-FLAG antibody. Control probed using only secondary anti mouse Cy3 antibody. Transfection efficiency 39 %.

The smaller *N*-terminal 3xFLAG tag does not appear to alter the localisation of ABH7, as discrete foci are again observed through the cytoplasm, (Figure 4-13) suggesting that the smaller tag does not obscure the likely mitochondrial targeting sequence.

4.5.3 Further analysis of ABH1 and ABH7 localisation

To further investigate the localisation of the ABH proteins, colocalisation studies were undertaken by Alexander Drong, using commercially available fluorescent markers that

selectively accumulate in specific subcellular organelles. MitoTracker® (Invitrogen) is cationic, and accumulates in the mitochondria due to the negative mitochondrial membrane potential, whereas LysoTracker® (Invitrogen) is a weakly basic amine species, which is held in the lysosomes by pH partitioning. Both species may be viewed in the Texas Red channel, allowing orthogonal detection of GFP signals.

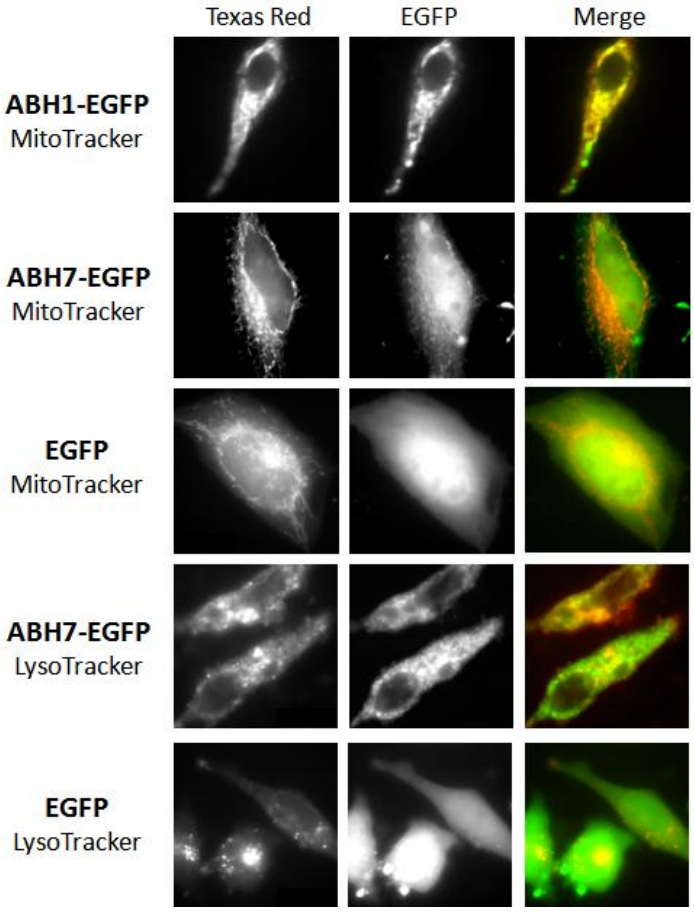


Figure 4-14: Analysis of colocalisation of GFP species with MitoTracker and LysoTracker dyes in HeLa cells, detected by direct fluorescence.

Staining with the chemical dyes revealed clear organelle structures (Figure 4-14). The structures visible on staining with MitoTracker® appear to colocalise with the signals for ABH1-EGFP and ABH7-EGFP, as shown by the yellow colour produced in the merged signal, although ABH7-EGFP is again additionally expressed in the nucleus. In contrast, ABH7-EGFP does not appear to colocalise with the LysoTracker® signals, and a control experiment shows no specific localisation of EGFP in either organelle. This therefore lends further support to the assignment of ABH1 and ABH7 as mitochondrial proteins.

4.6 Cell lysis

Based on the previous assignment of ABH1 as a mitochondrial protein[210], and the localisation studies carried out here (Section 4.5), initial attempts were carried out to prepare a mitochondrial protein extract from transfected cells. As the subcellular localisation was observed to be altered by the presence of the large GFP-tag at different ends of the protein only the *C*-terminally GFP-tagged proteins, which show a more defined localisation pattern, were investigated here.

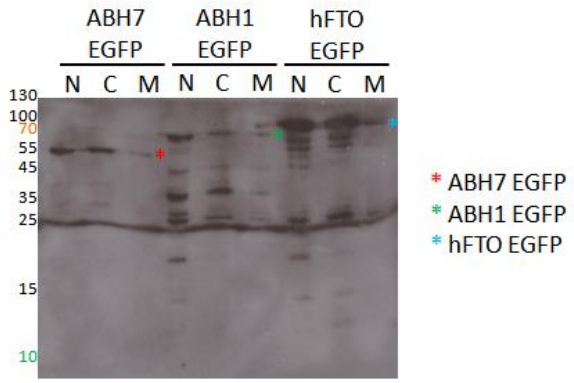


Figure 4-15: Mitochondrial extraction kit used for extraction of ABH1-EGFP, ABH7-EGFP and hFTO-EGFP proteins from 9.5 x 10^5 transiently transfected 293T cells. N: nuclear fraction; C: cytoplasmic fraction; M: mitochondrial fraction. 30 min exposure.

293T cells were transfected with either the ABH1-GFP or ABH7-GFP constructs, alongside hFTO-GFP as a comparison for extraction of a known nuclear protein, and protein extraction attempted using a ThermoScientific Mitochondrial Extraction kit. Briefly, the cells were lysed then the nuclei pelleted. Following incubation with a proprietary buffer, the mitochondria were separated from the other cytosolic components by centrifugation. Analysis of the GFP-tagged samples by western blot (Figure 4-15) shows a similar extraction pattern for each of the three proteins analysed, with the majority of the protein released in the initial nuclear and cytosolic fractions, with a smaller amount in the final "mitochondrial" fraction. As this pattern is observed for hFTO-GFP (known to be a nuclear protein) as well as the ABH proteins, it appears unlikely that the fractionation process has been successful.

Additionally it can be observed that, despite using the same quantity of cells and transfection protocol for the three constructs, the expression level of the ABH proteins

is low compared to that of hFTO (Figure 4-15). Expression of endogenous ABH1 was previously observed to be low when studied by immunofluorescence (Section 4.5.1). Some proteins are more highly expressed in cells grown for longer time-periods post-transfection, therefore this time can be optimised[330]. As expression of the ABH1-EGFP protein in particular was observed to be relatively low, a timecourse was carried out, harvesting cells at 24 h intervals post-transfection. As a comparison, hFTO-EGFP expression was followed, as this is observed to be well expressed after 24 h. Although the expression level of hFTO-EGFP was consistent through the time course, the expression level of ABH1-EGFP decreased over time (Figure 4-16), with significantly less protein expressed after 72 h. Therefore the 24 h post-transfection incubation period appears to be optimal, and was used for subsequent investigations of ABH1.

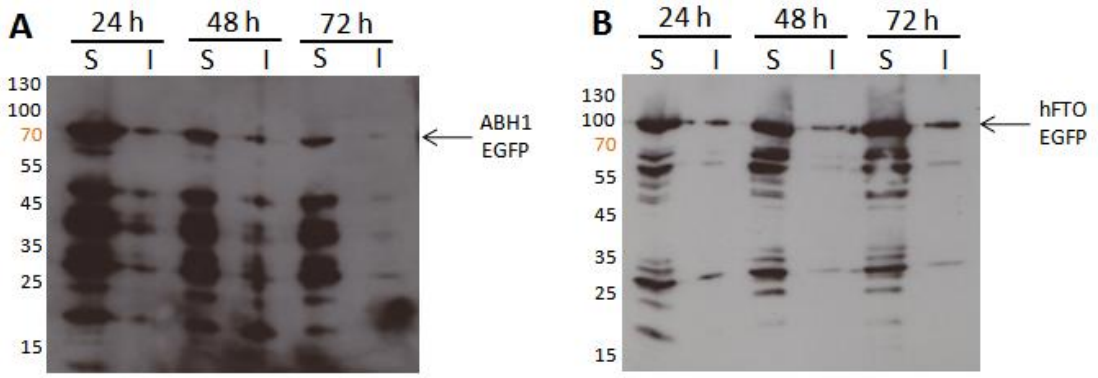


Figure 4-16: Timecourse following protein expression after transient transfection in 9.5 x 10^5 293T cells. (A) ABH1-EGFP, 30 min exposure; (B) hFTO-EGFP, 1 min exposure. S: soluble protein; I: insoluble protein.

Further subcellular fractionation was attempted using protocols described by Westbye et al. in the publication describing ABH1 as a mitochondrial protein[210]. Initially, a nuclear extraction protocol was attempted to determine whether the non-nuclear ABH proteins would be excluded from the nuclear fraction, using hFTO-GFP and GFP constructs for comparison. Very little protein extracted from cells transfected with the ABH proteins was found in the nuclear fraction (Figure 4-17A), although the same was true for hFTO-GFP (Figure 4-17B), despite it being known to be a nuclear protein. The majority of the GFP-tagged ABH proteins appeared to be released in the initial cytoplasmic fractions, in agreement with their presence in cytosolic species.

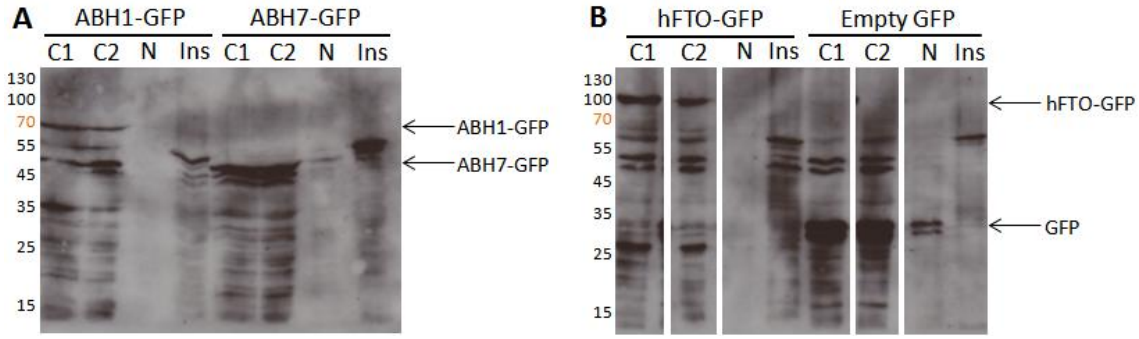


Figure 4-17: Nuclear extraction protocol carried out with 2.5 x 10^7 transiently transfected 293T cells. (A) ABH1-GFP and ABH7-GFP; (B) hFTO-GFP and GFP. C1: first cytoplasmic fraction; C2: second cytoplasmic fraction; N: nuclear fraction; Ins: Insoluble protein fraction. 1 hour exposures.

Subcellular fractionation to obtain mitochondria was also described by Westbye et al., and was attempted here for ABH1, alongside hFTO and GFP controls. The GFP construct used has been observed by immunofluorescence to be present throughout the cell (Figure 3-6, Chapter 3), and this is corroborated by the extraction profile (Figure 4-18A), which shows good expression of GFP, with protein present in each of the subcellular fractions. Unexpectedly, hFTO-GFP is most predominant in the wash fractions, but with only a small proportion present in the nuclear fraction, again showing the difficulties encountered in obtaining a pure nuclear lysate. However, very little hFTO-GFP is present in the lysed mitochondria fractions. Despite the good expression of the GFP and hFTO-GFP fractions, relatively little ABH protein was obtained. Unfortunately, very little ABH1 was found in the either the initial mitochondrial fraction, or the final lysed mitochondrial fractions, with a comparable protein quantity in the wash fraction.

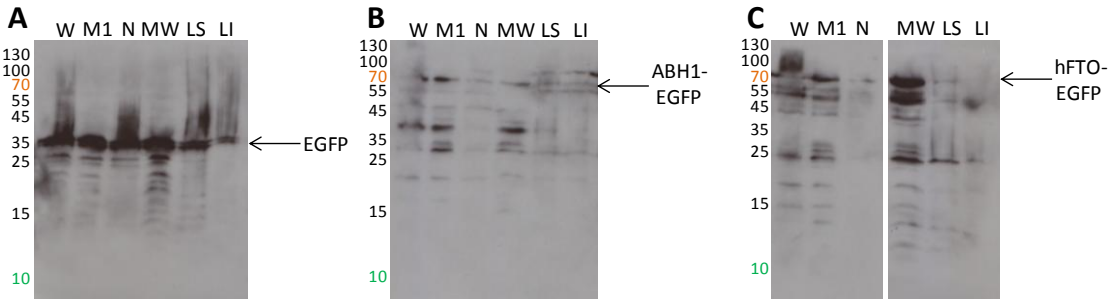


Figure 4-18: Mitochondrial extraction protocol carried out with 2.5 x 10^7 transiently transfected 293T cells. (A) EGFP, (B) ABH1-EGFP, (C) hFTO-EGFP. W: wash fraction; M1: crude mitochondria; N: nuclear/cell debris fraction; MW: mitochondrial wash fraction; LS: lysed mitochondria, soluble fraction; LI: lysed mitochondria, insoluble fraction. 1 hour exposures.

As no clear protein fractionation pattern was observed for either ABH1 or ABH7, these analyses cannot be used to provide further information about the true subcellular localisation of the proteins. Additionally, due to the significant protein loss at each stage of the extraction protocols, it was decided to use a whole cell lysate for subsequent IP and proteomic analyses, as previously employed for studies on hFTO (Section 3.4).

4.7 IP tests

To determine whether ABH1 and ABH7 IP reactions would be successful an initial small scale trial was attempted for ABH1-GFP and ABH1-GFP using GFPTrap beads. The western blot (Figure 4-19) shows that GFP-tagged proteins are efficiently bound to the beads, although significant degradation of immunoprecipitated ABH7-GFP to GFP is observed (Figure 4-19, ABH7 bead lane).

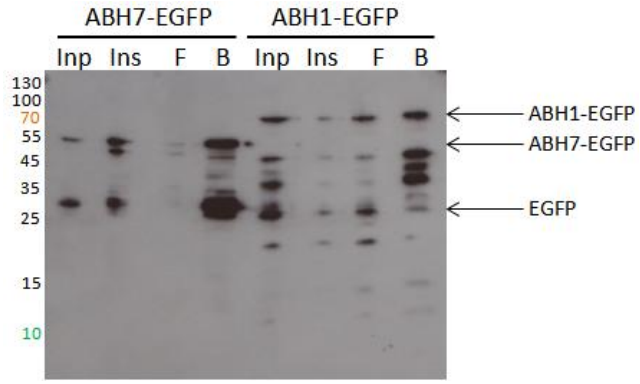


Figure 4-19: Test of ABH1-EGFP and ABH7-EGFP immunoprecipitation by GFPTrap beads from 9.5 x 10^5 293T cells transiently transfected with either N1-EGFP ABH1 or N1-EGFP ABH7. Whole cell lysate used as IP input. 30 min exposure.

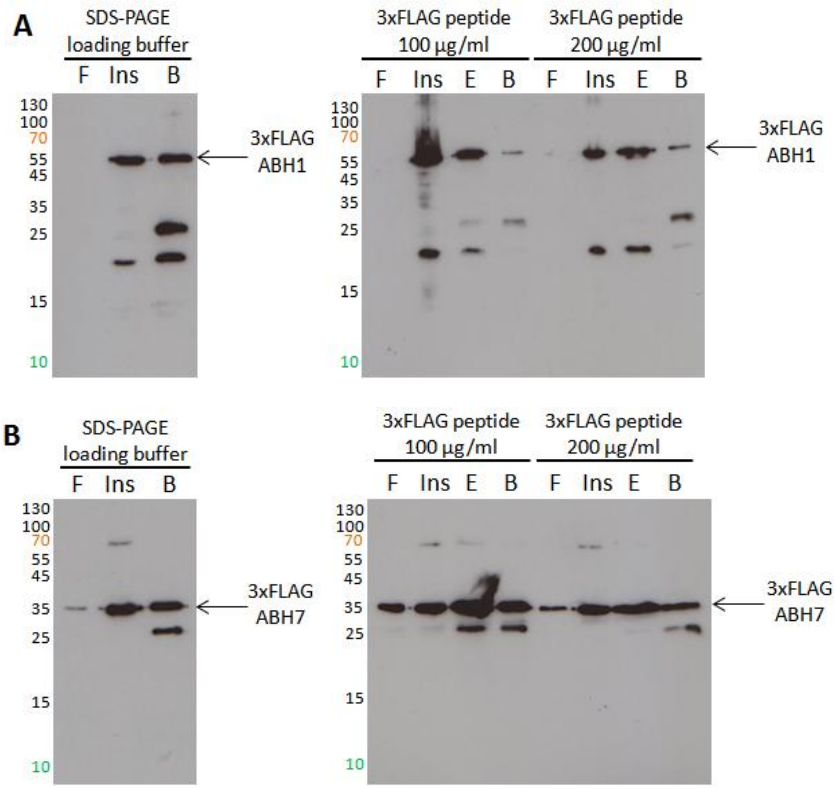


Figure 4-20: Immunoprecipitation of 3xFLAG tagged (A) ABH1 and (B) ABH7, testing methods for protein elution from beads. 9.5 x 10^5 293T cells transiently transfected with FLAG-tagged constructs, and immunoprecipitated using α-FLAG beads. Protein eluted from beads by addition of SDS-PAGE loading buffer or 3xFLAG peptide, as indicated. F: IP flow through; Ins: insoluble protein fraction; E: protein eluted from beads using 3xFLAG peptide; B: protein eluted from beads by addition of SDS-PAGE loading buffer. Exposures 1 min.

Immunoprecipitation of the FLAG-ABH constructs was also tested. In a further attempt to facilitate proteomic analyses, a gentler elution approach was tested for elution from the beads. The standard method previously used for protein elution was boiling the beads in SDS-PAGE loading buffer. This disrupts any interactions between the protein tag and antibody, denatures proteins, and breaks the bonds between the solid support beads and the antibody. When using the anti-FLAG beads this can complicate western blot analysis, as the bands corresponding to the antibody chains can be very intense and can interfere with the protein bands to be analysed. It was hoped that gentler elution would reduce the quantity of antibody protein released from the beads, while reducing elution of proteins that may bind non-specifically to the solid support beads themselves. The alternate approach taken here was addition of 3xFLAG peptide to the bead-target protein complex. When present in excess this competes with the tagged protein for the

antibody sites, causing a specific release of the FLAG-tagged protein from the beads. Different concentrations of the 3xFLAG peptide were tested, to determine a sufficient quantity to occupy all of the potential antibody binding sites.

Elution of the FLAG-tagged ABH proteins was successful using this alternate method, with both of the tested concentrations producing efficient elution of the proteins from the beads (Figure 4-21). In each case a small amount of the ABH proteins remained on the beads following incubation with the FLAG-peptide, but overall the eluted protein quantity was comparable to that for samples eluted with the original boiling method. This peptide elution technique was therefore applied in proteomic analyses of the FLAG-ABH proteins.

4.8 Large scale immunoprecipitation of ABH1 and ABH7

4.8.1 Anti-GFP immunoprecipitation

A similar approach to that used for analysis of hFTO binding partners (Section 3.9) was used here for analysis of species interacting with ABH1 and ABH7 in 293T cells. However, as the localisation characteristics of the proteins tagged at alternate ends were very different, again only the species tagged with GFP at the *C*-terminus were used here. Proteomic data previously obtained by Dr Alexander Wolf for cells transfected with the empty C1-EGFP vector was used as a control.

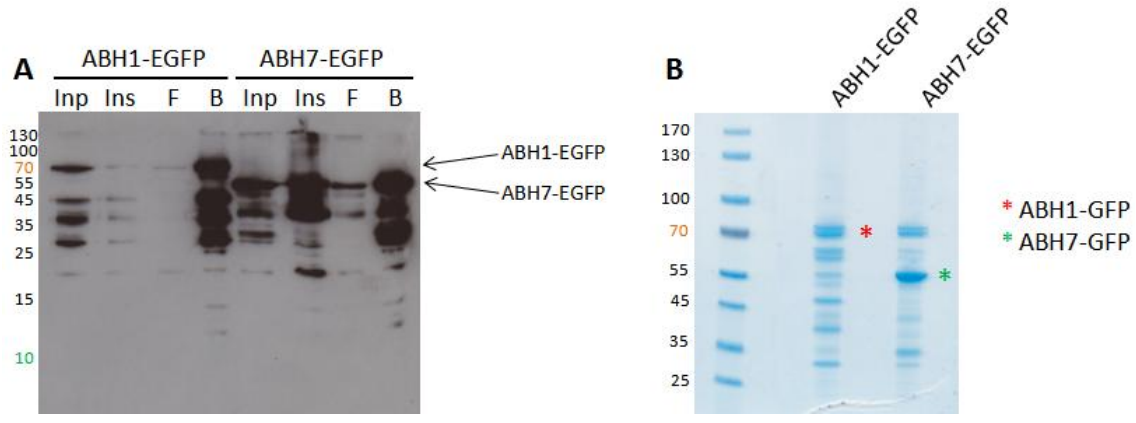


Figure 4-21: 2 x 10^8 293T cells transfected with either N1-EGFP ABH1 or N1-EGFP ABH7, and immunoprecipitated using GFPTrap beads. (A) Western blot testing expression and immunoprecipitation of ABH1-EGFP and ABH7-EGFP constructs. 1 µl each sample loaded. Inp: IP input; Ins: Insoluble protein fraction; F: Bead flow through; B: Beads. 30 min exposure. (B) 4-12 % NuPAGE gel stained with Colloidal Coomassie. Proteins eluted from beads using SDS-PAGE loading buffer, 35 µl each sample loaded per lane.

A western blot of the IP samples (Figure 4-21A) shows that large quantities of the GFP-tagged proteins are bound to the beads. Although no major band at the expected mass of ABH1-GFP is observed in the NuPAGE gel of the bead samples (Figure 4-21B), a clear band was observed by western blot, and the level of ABH1 protein expressed has been consistently less than for the other proteins studied. As previously, the gel lanes for both proteins were excised, divided horizontally into 20 equivalent portions, then individually trypsinised and analysed by mass spectrometry.

4.8.2 Anti-FLAG immunoprecipitation

Immunoprecipitation of the FLAG-tagged ABH proteins was also carried out, using as a control cells transfected with the empty 3xFLAG vector. Elution of the protein from the beads was achieved by incubation with 3xFLAG peptide, and a good quantity of protein could be seen to be recovered from the IP reactions (Figure 4-22). In this large scale experiment the recovery of the tagged protein in the eluted protein fraction was less successful than in the previous small scale tests, probably due to the increased volume of beads present. As previously (Figure 3-37) the only bands apparent on the western blot for the empty FLAG IP correspond to the light chain of the anti-FLAG antibody. Despite the use of the mild elution method there are still bands of reduced intensity corresponding to the antibody light chain in each of the eluted protein lanes. This is likely to be due to either carry-over of beads into the eluted protein sample, or elution of the antibody on treatment with the 3xFLAG peptide.

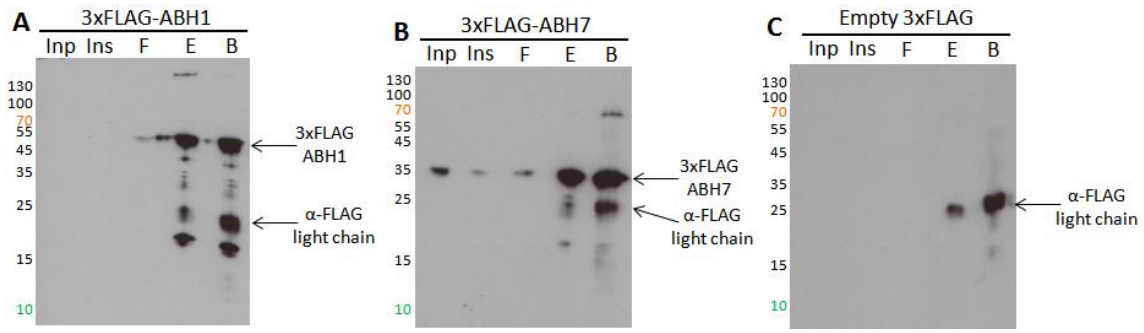


Figure 4-22: Western blot testing large scale expression and immunoprecipitation of (A) 3xFLAG-ABH1, (B) 3xFLAG-ABH7 and (C) empty 3xFLAG constructs. 2 x 10^8 293T cells transfected with either N1-EGFP ABH1 or N1-EGFP ABH7, and immunoprecipitated using GFPTrap beads. 1 µl each sample loaded. Inp: IP input; Ins: Insoluble protein fraction; F: Bead flow through; E: eluted protein; B: Beads. 1 min exposures.

As there were significant quantities of protein in the eluted protein and bead samples for each reaction, both samples were run on a NuPAGE gel (Figure 4-23). In each of the samples distinct bands corresponding to the proteins of interest are observed, showing that FLAG-ABH1 and FLAG-ABH7 have been successfully overexpressed. Additionally it can be seen that there are fewer proteins present at a lower concentration in the eluted protein lanes compared to the bead lanes, suggesting that the elution method attempted is more gentle and has hopefully only caused elution of more true protein interactors, reducing the background and allowing more straightforward data analysis.

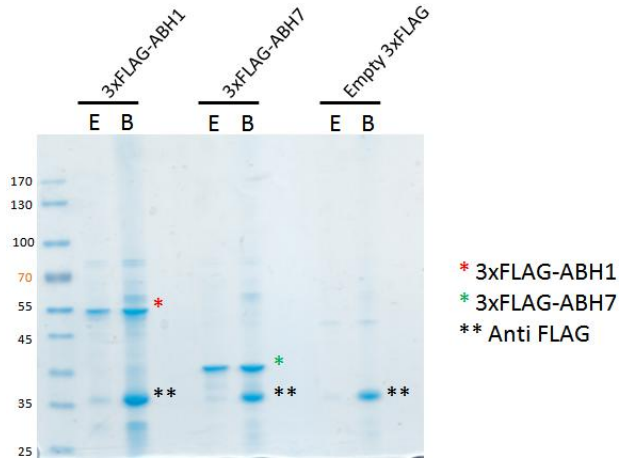


Figure 4-23: 4-12 % NuPAGE gel stained with Colloidal Coomassie. 2 x 10^8 cells transiently transfected with 3xFLAG-ABH1, 3xFLAG-ABH7 or empty 3xFLAG vector, and immunoprecipitated using anti-FLAG beads. E: Proteins eluted from beads by addition of 50 µl 150 µg/ml 3xFLAG peptide; B: Proteins eluted from beads by addition of 100 µl SDS-PAGE loading buffer, 35 µl each sample loaded per lane.

Due to the reduced concentrations of proteins present in the lanes of the gel, a slightly different approach was taken for sample analysis in this instance. The gel lanes in this case were excised, then each divided horizontally into 5 portions, which were then subjected to in-gel trypsinolysis, scaling up the reagent quantities used accordingly. MS analysis was carried out in the same manner as previously.

4.9 Proteomic analysis

4.9.1 ABH1 results

Accession number	Protein	Score
Q13686	Alkylated DNA repair protein alkB homolog 1	1948
P50502	Hsc70-interacting protein	132
P22695	Cytochrome b-c1 complex subunit 2, mitochondrial	99
P50402	Emerin	72
Q10567	AP-1 complex subunit beta-1	70
Q16775	Hydroxyacylglutathione hydrolase, mitochondrial	46
Q7Z6J4	FYVE, RhoGEF and PH domain-containing protein 2	43
P09601	Heme oxygenase 1	42
P62158	Calmodulin	42
O00255	Menin	41
Q9BQI7	RPH and SEC7 domain-containing protein 2	40
Q9H583	HEAT repeat-containing protein 1	39
Q2WGJ9	Fer-1-like protein 6	39
P20062	Transcobalamin-2	38
Q9Y4A5	Transformation/transcription domain-associated protein	37
O60235	Transmembrane protease serine 11D	37
Q7Z7A1	Centriolin	36
Q9Y5M8	Signal recognition particle receptor subunit beta	35
P28072	Proteasome subunit beta type-6	34
Q9UHD9	Ubiquilin-2	32
Q9NVE4	Coiled-coil domain-containing protein 87	32
O60885	Bromodomain-containing protein 4	31
O00624	Sodium-dependent phosphate transport protein 3	31
P63211	Guanine nucleotide-binding protein G(T) subunit gamma-T1	31
Q8NDB2	B-cell scaffold protein with ankyrin repeats	31
P62979	Ubiquitin-40S ribosomal protein S27a	31
Q5VZ89	DENN domain-containing protein 4C	30
Q5TFE4	5'-nucleotidase domain-containing protein 1	30
Q6W2J9	BCL-6 corepressor	30
Q9HDC9	Adipocyte plasma membrane-associated protein	30
Q9NUU6	Protein FAM105A	30
Q76FK4	Nucleolar protein 8	30

Table 4-3: IP proteins from ABH1-GFP (N1 EGFP) overexpression and anti-GFP pulldown. Blue shading: ABH1; Yellow shading: mitochondrial proteins; Red text, ubiquitin associated proteins.

LC/MS/MS analysis of tryptic peptides and MASCOT searches to identify interacting proteins was carried out as previously (Section 3.9), by Dr Holger Kramer. Proteins identified in the ABH1 samples, but not in the corresponding controls, are listed in

Table 4-3 and Table 4-4. As observed with FTO, some ubiquitin-associated proteins were identified in the GFP IP reaction, suggesting that ABH1 may be ubiquitinated in cells. However, despite the identification of ABH1 as a mitochondrial protein, very few mitochondrial proteins were identified in the IP reactions (shaded blue in Table 4-3). Additionally no proteins (barring the target ABH1 protein) were identified both of the IP reactions using the two different tags. Together, these results suggest that the IP reactions have not been successful in identifying proteins with which ABH1 may interact in cells.

Accession number	Protein	Score
Q13686	Alkylated DNA repair protein alkB homolog 1	2685
Q5TCZ1	SH3 and PX domain-containing protein 2A	41
B2RXH2	Lysine-specific demethylase 4D-like	36
Q96G23	LAG1 longevity assurance homolog 2	36
P49454	Centromere protein F	34
O43374	Ras GTPase-activating protein 4	32

Table 4-4: IP proteins from FLAG-ABH1 overexpression and anti-FLAG pulldown. Blue shading: ABH1.

4.9.2 ABH7 results

Proteomic analyses were carried out in the same manner for ABH7, with identified proteins listed in Table 4-5 and Table 4-6. A considerable proportion of the proteins identified are mitochondrial, possibly supporting the mitochondrial localisation of ABH7. However ATP synthase, identified in the FLAG-ABH7 IP reaction (Table 4-6), was previously seen to interact with I367F FTO in an IP from mouse tissue (Table 3-9). Again, no proteins were identified in the IP products for both GFP and FLAG tags.

Accession number	Protein	Score
Q9BT30	Alpha-ketoglutarate-dependent dioxygenase ABH7	3105
Q9UNE7	E3 ubiquitin-protein ligase CHIP	81
P14735	Insulin-degrading enzyme	75
O95831	Apoptosis-inducing factor 1, mitochondrial	72
P42704	Leucine-rich PPR motif-containing protein, mitochondrial	65
P35520	Cystathionine beta-synthase	62
Q99653	Calcium-binding protein p22	59
Q9H9B4	Sideroflexin-1	58
P45880	Voltage-dependent anion-selective channel protein 2	56
P29692	Elongation factor 1-delta	52
P46379	Large proline-rich protein BAT3	51
Q96HS1	Serine/threonine-protein phosphatase PGAM5, mitochondrial	49
P00338	L-lactate dehydrogenase A chain	47
P29966	Myristoylated alanine-rich C-kinase substrate	47
O75489	NADH dehydrogenase [ubiquinone] iron-sulfur protein 3, mitochondrial	47
Q9NS69	Mitochondrial import receptor subunit TOM22 homolog	44
Q13228	Selenium-binding protein 1	43
O43399	Tumor protein D54	41
Q9UNM6	26S proteasome non-ATPase regulatory subunit 13	41
P50993	Sodium/potassium-transporting ATPase subunit alpha-2	41
P00167	Cytochrome b5	41
P98173	Protein FAM3A	38
P54819	Adenylate kinase 2, mitochondrial	38
P08195	4F2 cell-surface antigen heavy chain	38
P40925	Malate dehydrogenase, cytoplasmic	37
Q8WYB5	Histone acetyltransferase MYST4	37
P11441	Ubiquitin-like protein 4A	36
Q8TB73	Fibronectin type-III domain-containing protein C4orf31	36
Q16819	Meprin A subunit alpha	35
Q8N4Q1	Mitochondrial intermembrane space import and assembly protein 40	35
Q15738	Sterol-4-alpha-carboxylate 3-dehydrogenase, decarboxylating	34
Q9HCN8	Stromal cell-derived factor 2-like protein 1	33
Q2TAA2	Isoamyl acetate-hydrolyzing esterase 1 homolog	33
Q9UJZ1	Stomatin-like protein 2	33
Q9ULK5	Vang-like protein 2	32
Q15637	Splicing factor 1	31
P05388	60S acidic ribosomal protein P0	31
P11177	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	30
Q9P2E3	NFX1-type zinc finger-containing protein 1	30

Table 4-5: IP proteins from ABH7-GFP (N1 EGFP) overexpression and anti-GFP pulldown. Blue shading: ABH7; Yellow shading: mitochondrial proteins; Red text, ubiquitin associated proteins.

Accession number	Protein	Score
Q9BT30	Alpha-ketoglutarate-dependent dioxygenase ABH7	3102
P06576	ATP synthase subunit beta, mitochondrial	85
P15531	Nucleoside diphosphate kinase A	58
P49184	Deoxyribonuclease-1-like 1	35
Q9UDY6	Tripartite motif-containing protein 10	34
Q63ZY3	KN motif and ankyrin repeat domain-containing protein 2	34
Q6ZN66	Guanylate-binding protein 6	30

Table 4-6: IP proteins from FLAG-ABH7 overexpression and anti-FLAG pulldown. Blue shading: ABH7; Yellow shading: mitochondrial proteins

4.9.3 Comparison of ABH1 and ABH7 results

To assess the likelihood of the proteins identified being false positives, decoy searches were run (Table 4-7). This suggests that ~3 % of the proteins identified for the FLAG-ABH samples may be false positives (above the identity threshold), however the likelihood is higher for the IP samples obtained from the lysate of cells transfected with the empty FLAG vector.

Peptide match %	FLAG-ABH1		FLAG-ABH7		Empty FLAG	
	Eluted	Beads	Eluted	Beads	Eluted	Beads
Above identity threshold	2.72	1.49	0.95	3.05	6.47	3.2
Above identity or homology threshold	11.45	8.13	7.92	10.19	24.16	12.35

Table 4-7: Percentage identification rate for MASCOT decoy search analysis of IP samples for FLAG constructs.

Overall, few proteins were identified in the anti-FLAG IP samples for FLAG-ABH1 and FLAG-ABH7, and additionally only likely contaminant proteins (e.g. keratin and heat shock proteins) were identified in both the eluted and bead samples. Therefore the proteins identified from the eluted and bead samples were combined above.

As both ABH1 and ABH7 are thought to be mitochondrial, and tagged species were selected such that the tag should not interfere with this subcellular localisation, it may be expected that a significant proportion of the proteins identified would be mitochondrial but, particularly for ABH1, this is not observed. Furthermore, as no proteins were identified on IP with both GFP and FLAG for either ABH1 or ABH7 it is unlikely that the proteins identified are true interactors. The interacting proteins identified have low scores, and do not appear to represent any specific protein classes.

4.10 Discussion

This chapter describes the expression, purification, and characterisation of ABH1 and ABH7 proteins. Preliminary activity studies did not yield any likely substrates among a wide range of histone sequences, or methylated nucleosides. However, future *in vitro* analyses should focus on investigation of further potential substrates including

methylated oligos, tRNA modifications, and oxidative nucleic acid lesions particularly with respect to the location of the proteins in the mitochondria. Crystallographic studies of ABH7 are ongoing, which may yield more information about the role of this protein. As these initial *in vitro* studies were not fruitful, an alternate approach was taken to investigate the proteins in cells.

Immunofluorescence studies confirmed the assignment of ABH1 as a mitochondrial protein, and identified for the first time ABH7 as mitochondrial. The mitochondrial assignment of ABH7 would be strengthened by deletion mutagenesis of the likely mitochondrial targeting sequence, and immunofluoresecnce studies analysed by confocal microscopy.

Immunoprecipitiation of GFP and FLAG-tagged ABH1 and ABH7 proteins was carried out, and the proteins interacting with FTO were analysed by proteomics. Unfortunately no significant protein interactors could be confidently identified for either protein, possibly suggesting that neither protein interacts with significantly with other proteins in cells. However, tests carried out with a commercially available αABH1 antibody suggested that a direct IP using this antibody should be possible, therefore this should be the focus of future studies, to remove the background interactions associated with binding to the tags. An ABH7 antibody is now also commercially available, so the corresponding ABH7 studies could be carried out using this reagent. Use of the αABH1 and αABH7 antibodies would also facilitate ChIP and crosslinking studies, which would allow a more broad analysis of species that may interact with the ABH proteins in cells.

Chapter 5 ABH Purification and Characterisation

5.1 Introduction

The *E. coli* DNA demethylase AlkB has significant structural and functional homology to the human ABH2 and ABH3 proteins. These are 2OG oxygenases with a preference for the repair of 1meA lesions[163,326], using a direct reversal mechanism[158-159]. AlkB has a broader substrate range than the human homologues, efficiently removing a number of more bulky base adducts, and acting on both ssDNA and RNA[127]. ABH3 repairs ssDNA methyl lesions, although little repair is observed *in vivo*[174]. In contrast it appears that ABH2 has a preference for dsDNA, and acts as the major repair protein for 1meA lesions *in vivo*[174]. Although AlkB is part of the well studied *E. coli* adaptive response, less is known about the regulation and role of ABH2 in *H. sapiens*. To date it is known that ABH2 is a nuclear protein that co-localises with proliferating cell nuclear antigen (PCNA) at replication foci in S phase[167], via an *N*-terminal binding motif[168].

As AlkB is more "tractable" than ABH2 or ABH3 it was decided to use it as a model system for the study of this family of DNA demethylases. Following investigation of AlkB, the studies may allow comparison to the human homologue ABH2. Each of the ABH family members appears to have a flexible *N*-terminal region[192,194] which shows less homology within the family[164]. The similarity between the *C*-terminal regions may be seen in the alignment of AlkB with ABH2 and ABH3 (Figure 5-1). As highlighted, the proteins each contain the conserved functional 2OG oxygenase residues, including the conserved R residue in the β8 strand, which is characteristic of the AlkB family[157]. Additionally several other regions of high similarity or identity are observed through the sequence. This similarity suggests that methodologies developed are likely to be applicable among the ABH family, here initially for further analysis of ABH2, and later for the study of the other ABH proteins, including ABH3 and ABH8[217].

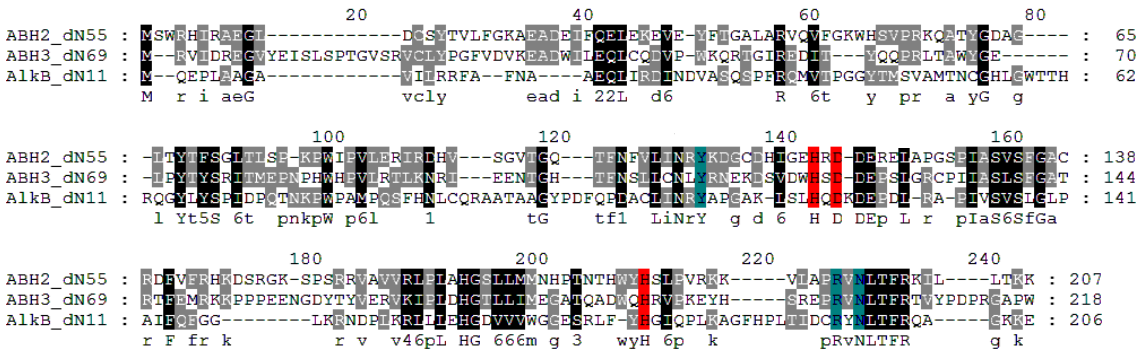


Figure 5-1: Sequence alignment of *N*-terminally truncated ABH constructs used in crystallography, comparing the *C*-terminal oxygenase domains. Numbering used refers to the position in the amino acid sequence of the truncated constructs. Red: conserved iron binding residues; Blue: 2OG binding residues. Alignment produced using TCoffee and GeneDoc.

To allow studies on these members of the ABH family, AlkB and ABH2 proteins were purified and characterised. The activity of the proteins produced was initially investigated by a standard radiolabelled 2OG turnover assay, then a coupled assay detecting the formaldehyde released in demethylation reactions was developed and optimised allowing improved analysis of the enzymatic activity.

5.2 Construct and protein production

5.2.1 AlkB protein constructs and purification

Full length AlkB constructs were cloned by Dr Richard Welford[325], by insertion of the AlkB sequence into the NdeI and BamHI sites of the pET-24a(+) (producing untagged protein) and pET-28a(+) (producing *N*-terminally His_6-tagged protein) vectors. Expression of the AlkB protein from the pET-24a(+) construct was optimised by Dr Richard Welford, using 0.2 mM IPTG to induce expression in *E. coli*, followed by growth for 4 h at 28 °C.

Previously published crystal structures of AlkB were obtained using a construct truncated by 11 amino acids at the *N*-terminus, as hydrogen/deuterium exchange experiments revealed that the *N*-terminus is conformationally flexible[192]. To enable crystallography, the corresponding AlkB constructs were produced from the full length pET-24a(+) and pET-28a(+) constructs using deletion mutagenesis to remove the nucleotide sequence corresponding to amino acid residues 1-11 of the AlkB sequence.

This construct is referred to as ΔN11 AlkB throughout. Expression trials were carried out to determine the optimal expression conditions for production of ΔN11 AlkB in *E. coli*, with greatest soluble expression found for overnight expression at either 15 °C (pET-24a) or 28 °C (pET-28a) following induction with 0.2 mM IPTG (Figure 5-2, Figure 5-3).

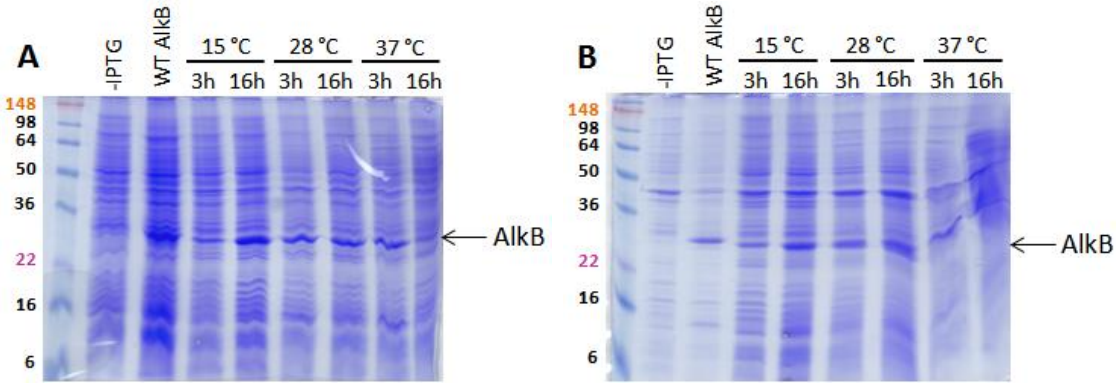


Figure 5-2: Expression trials of ΔN11 AlkB pET-24a(+) in BL21 (DE3) *E. coli*. (A) Soluble protein, (B) Insoluble protein. Expression induced with 0.2 mM IPTG, then cells grown under the temperature and time conditions indicated.

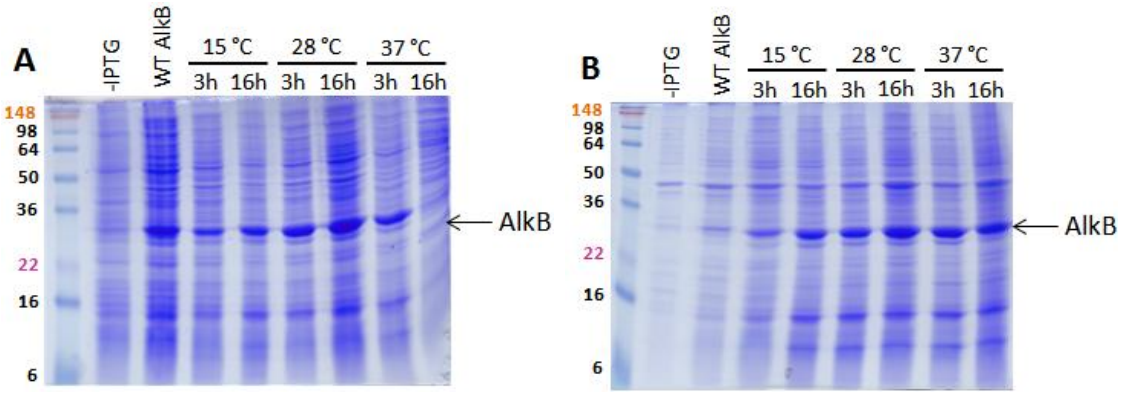


Figure 5-3: Expression trials of ΔN11 AlkB pET-28a(+) in BL21 (DE3) *E. coli*. (A) Soluble protein, (B) Insoluble protein. Expression induced with 0.2 mM IPTG, then cells grown under the temperature and time conditions indicated.

Cation exchange chromatography was used for initial purification of pET-24a(+) AlkB constructs, with crude AlkB protein eluted by application of an NaCl gradient. Fractions containing the highest levels of pure AlkB, judged by SDS-PAGE analysis (Figure 5-4A), were pooled and concentrated.

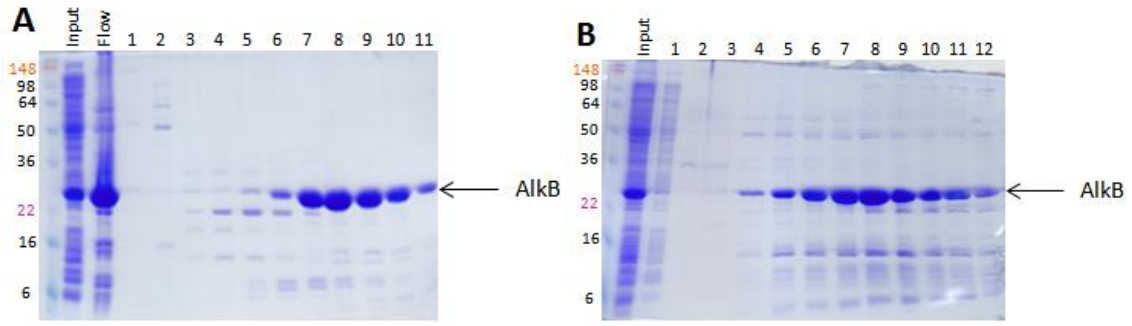


Figure 5-4: SDS-PAGE gels obtained for a typical AlkB protein purification. (A) Samples obtained from cation exchange purification. Input: Cell lysate used as column input; Flow: Protein flowing through column without binding; 1-11: Fractions eluted from column, fractions 7-10 pooled and concentrated. (B) Samples obtained from gel filtration purification. Input: Gel filtration column input; 1-12: Fractions eluted from column, fractions 6-11 pooled and concentrated.

An alternate method was employed for purification of His-tagged ΔN11 AlkB expressed from the pET-28a(+) construct. Nickel affinity chromatography was carried out, exploiting the presence of the affinity tag. Following elution from the column, SDS-PAGE analysis again showed the appropriate fractions to be pooled (Figure 5-5A), and the His-tag was cleaved by overnight incubation with thrombin.

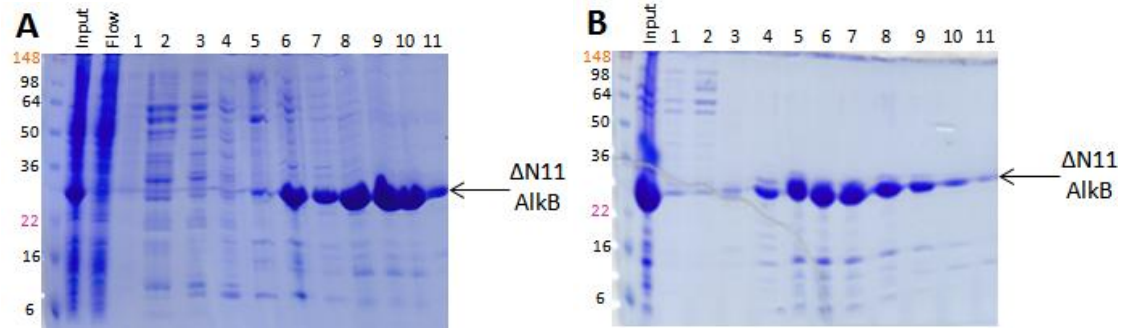


Figure 5-5: SDS-PAGE gels of a typical nickel affinity purification of ΔN11 AlkB pET-28a. (A) Samples obtained from nickel affinity purification. Input: Cell lysate used as nickel column input; Flow: Protein flowing through column without binding; 1-11: Fractions eluted from column, fractions 7-11 pooled and concentrated. (B) Samples obtained from gel filtration purification. Input: Gel filtration column input; 1-11: Fractions eluted from column, fractions 5-9 pooled and concentrated.

Following these initial purification steps, proteins were further purified by gel filtration chromatography (Figure 5-4B, Figure 5-5B). Purified proteins (Figure 5-6) were concentrated and exchanged into an appropriate buffer for storage at -80 °C. Depending on the intended purpose, proteins were stored in either 50 mM Tris, 500 mM NaCl, 1 mM DTT, pH 7.5, or 50 mM HEPES, pH 7.5. Typical protein yields from 1 litre of

culture were approximately as follows: AlkB pET-24a, 3.9 mg, ΔN11 AlkB pET-24a, 7.2 mg, ΔN11 AlkB pET-28a, 5.9 mg.

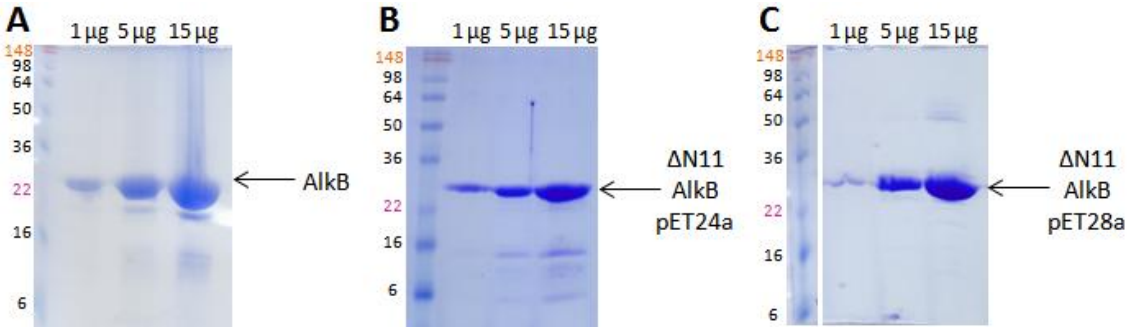


Figure 5-6: SDS-PAGE gels of purified AlkB proteins. (A) AlkB pET-24a; (B) ΔN11 AlkB pET-24a; (C) ΔN11 AlkB pET-28a, quantities as indicated. Purity of proteins is ~95 %, judged by SDS-PAGE.

5.2.2 ABH2 protein constructs and purification

A construct containing the full length ABH2 sequence cloned into the NdeI and NotI sites of pET-28a(+) was obtained from Dr Kirsty Hewitson. Expression of ABH2 was found to be optimal in cells grown for 4 hours at 37 °C following induction with 0.2 mM IPTG. As is the case for AlkB, the published crystal structures of ABH2 use an *N*-terminal truncation, in this case the deletion of residues 1-54[194]. The corresponding truncation, to produce ΔN55 ABH2, was cloned by Dr Alexander Wolf. Expression trials also carried out by Dr Alexander Wolf identified optimal expression of ΔN55 ABH2 in BL21 (DE3) cells grown for 16 h at 18 °C following induction of expression with 0.2 mM IPTG.

Both ABH2 constructs were purified using nickel affinity purification followed by gel filtration. However, the His-tag could not be successfully cleaved from the ABH2 proteins due to the presence of an internal thrombin cleavage site, which produced cleaved protein. Yields of proteins obtained per litre of bacterial culture were 0.7 mg for ABH2, and 3.7 mg for ΔN55 ABH2. Both proteins were purified to a purity of ~95 %, as judged by SDS-PAGE analysis (Figure 5-7).

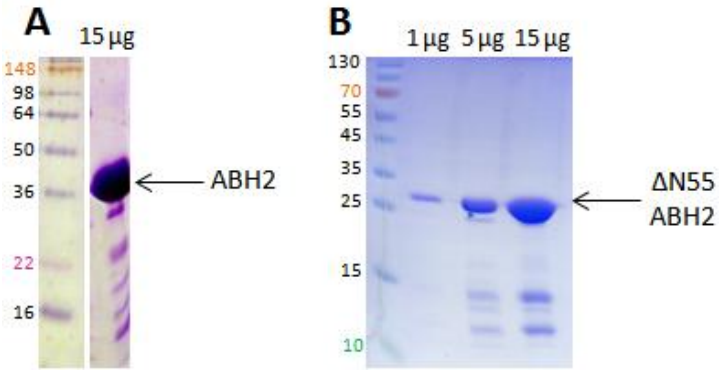


Figure 5-7: SDS-PAGE gels of purified ABH2 proteins. (A) ABH2; (B) ΔN55 ABH2, quantities as indicated.

5.3 Protein characterisation

5.3.1 Circular Dichroism

To verify that the purified proteins were folded in solution, and that the *N*-terminally truncated proteins have a similar secondary structure to the full length wildtype proteins, circular dichroism spectra of the purified AlkB and ABH2 proteins were obtained. The spectra of AlkB and ΔN11 AlkB have a very similar overall shape (Figure 5-8), as would be expected from the small deletion of 10 residues from the protein terminus.

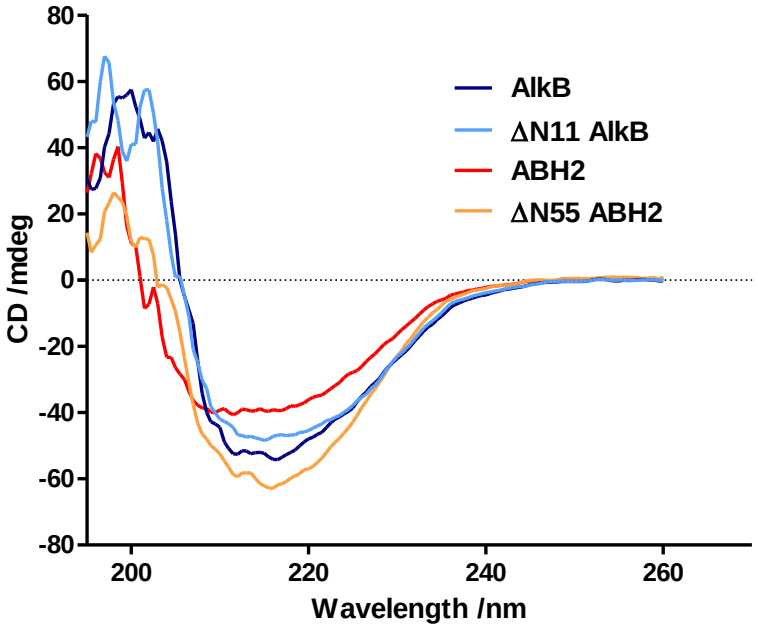


Figure 5-8: Circular dichroism spectra for AlkB and ABH2 proteins (0.5 mg/ml), recorded at 4 °C. Data points represent the mean of triplicate experiments.

A more significant difference is observed between the full length ABH2 and ΔN55 ABH2 constructs. The spectrum of the full length protein has increased intensity at ~205 nm compared to the ΔN55 construct (Figure 5-8). Comparison to the Fasman curves presented previously (Chapter 2, Figure 2-36) suggests that the full length protein has a greater proportion of α-helix whereas the truncated construct has a higher proportion of β-sheet secondary structure content, which is supported by deconvolution of the data obtained. It is unsurprising that there is a significant difference in the overall secondary structure of the protein as this construct represents a large deletion; about 20 % of the protein is truncated (55 out of 261 residues). Of the protein sequence remaining in the truncated construct much is known to comprise the DSBH core, therefore giving an increased proportion of β-sheet character. Use of the JPred3 structural prediction programme[331] to suggest any likely secondary structure features in residues 1-55 of the ABH2 sequence reveals that the only confidently predicted structure in this region is a short α-helix comprising residues 2-6. The remaining sequence is predicted to be generally unstructured, forming random coils. This lack of defined structure reveals a rationalisation for use of an *N*-terminally truncated construct for the published crystal structure for ABH2, as flexible regions are unlikely to successfully crystallise.

5.3.2 Non-denaturing Mass Spectrometry

To investigate more precisely the masses of the proteins purified, and assess their suitability for use in subsequent analysis, non-denaturing mass spectra of the AlkB proteins were obtained with Dr Esther Woon. These spectra (Figure 5-9) show that the calculated masses of the proteins are in good agreement with the masses observed for the native proteins. Some 2OG oxygenases are observed by mass spectrometry to co-purify with the Fe and 2OG co-substrates[332]. An additional peak is present in the spectrum of pET28a ΔN11 AlkB at a mass difference of +62 Da to the main protein peak, possibly due to the presence of a transition metal ion (e.g. M_r(Fe) = 55.8 Da; M_r(Ni) = 58.7 Da), which may have bound during the purification.

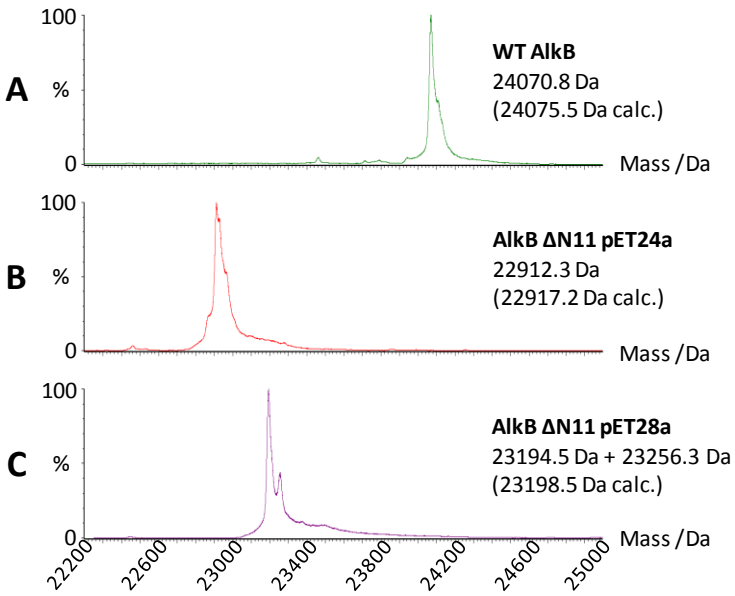


Figure 5-9: Non-denaturing mass spectra of purified AlkB proteins. (A) Full length AlkB pET-24a; (B) ΔN11 AlkB pET-24a; (C) ΔN11 AlkB pET-28a.

5.3.3 Differential Scanning Fluorimetry

Differential scanning fluorimetry (DSF) was used to determine the melting temperatures of the purified proteins. As results from DSF may be usefully applied for selection of conditions and compounds that stabilise proteins to improve crystallisation attempts, the *N*-terminally truncated protein constructs designed for use in crystallisation were analysed here.

To determine that appropriate melting curves could be obtained with this group of proteins, an initial screen was carried out varying the concentration of protein present. The results for AlkB (Figure 5-10) show that appropriate curves were obtained for protein in the range 2-10 μM, with consistent T_m values obtained between the triplicates at each of the protein concentrations. However, the T_m derived shows a marked decrease as the protein concentration is increased, showing the importance of maintaining the protein at a defined concentration throughout to ensure consistent T_m values between assays. To minimise the quantity of protein necessary, while still obtaining consistent results, 2 μM protein can be used for further analysis of small molecule testing.

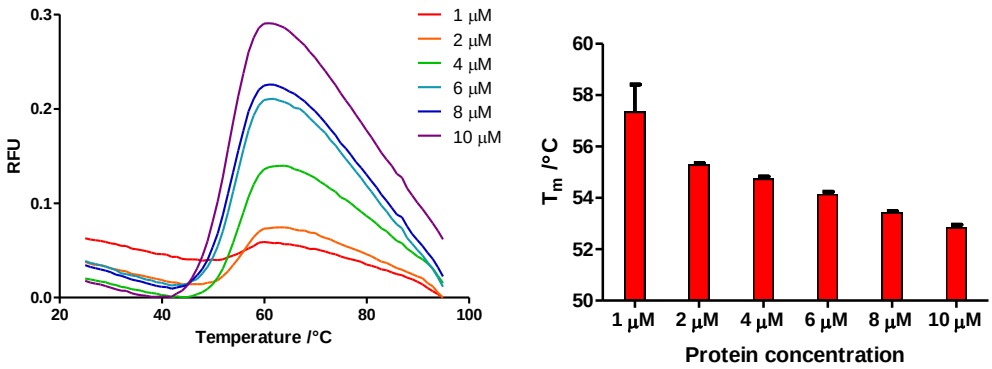


Figure 5-10: Differential scanning fluorimetry of ΔN11 AlkB at a range of concentrations; represented as melting curves and derived T_m values. Data represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

The tolerance of the proteins for different metals was also tested, using metal (II) salts to mimic the Fe(II) usually present in the active site of the proteins. As some metals can have a significant effect on the shape of the resulting DSF curve, a protein concentration of 4 μM was used here, to allow for reduction in magnitude of the curve while still enabling data analysis. Furthermore, as an initial test showed that T_m shifts obtained in the presence of DMSO in addition to the metal were more significant than those observed for either DMSO or metal alone, the metals were screened both with and without DMSO. Addition of a small proportion of DMSO to the assay mixture is required when adding small organic molecules that often have poor solubility in aqueous solution. Therefore for potential extension of this technique into a useful screening assay it is necessary for the protein to be able to withstand a certain amount of DMSO in this screening technique.

The T_m values obtained (Figure 5-11) show that addition of Co(II) and Ni(II) to the protein causes stabilisation of the protein compared to the reference sample containing no added metal or DMSO (T_m shift values summarised in Table 5-1). However, inspection of the melting curves shows that this increase in melting temperature is concomitant with a decrease in the fluorescence intensity, making extraction of the T_m from the raw data more difficult, and less accurate, as the region of the curve over which the Boltzmann function is to be fitted becomes very small. Interestingly, addition of zinc to the protein causes significant destabilisation, possibly due to the slightly

larger ionic radius[333] that may prevent effective binding in the active site. Neither Mn(II) nor Fe(II) produce major T_m shifts with AlkB, however the melting curves (Figure 5-11) are well defined. Although Fe(II) is the metal present naturally, Fe(II) is very readily oxidised to Fe(III) in oxygenated solution, therefore on addition of compounds and heating of the solution it is unlikely that the Fe(II) state would be maintained. Oxidation of Mn(II) to Mn(III) is less facile, therefore maintenance in the Mn(II) form is more likely, and addition of Mn(II) would be more appropriate in future studies. Cu(II) was also tested, but no T_m shift could be obtained as the data could not be fitted to a Boltzmann sigmoidal curve. This could be due to oxidative damage to the protein caused by generation of ROS by Cu(II), or through thiol oxidation[334].

Generally, the addition of DMSO destabilises the protein, significantly in the cases of Fe(II) and Ni(II), and although stabilisation of AlkB is observed in the presence of Co(II), the further addition of DMSO causes destabilisation. A slight stabilising effect is observed with Mn(II) in the presence of DMSO, making this condition the preferred option for future screening of small molecules.

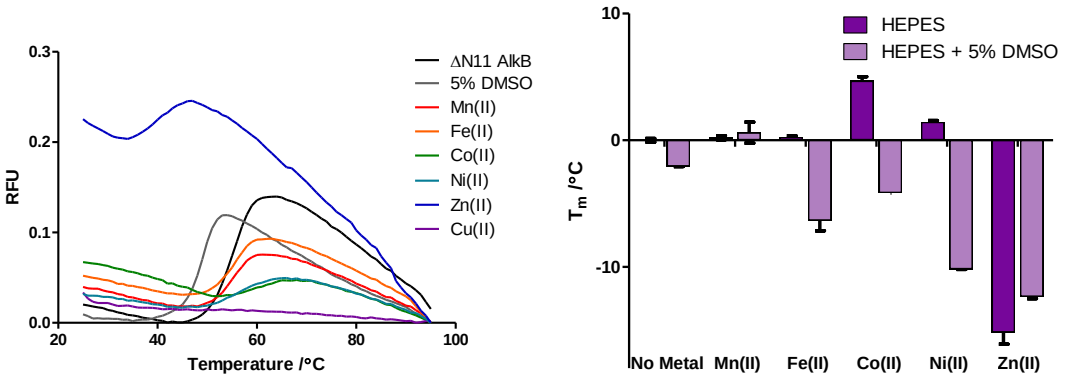


Figure 5-11: Differential scanning fluorimetry of ΔN11 AlkB (4 µM) with addition of 50 µM metal (II) and/or 5 % DMSO (v/v); represented as melting curves and derived T_m shift values. Data represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

A similar screen of metal (II) salts was carried out with ΔN55 ABH2. Overall, each of the metals tested cause protein destabilisation (Figure 5-12, T_m shifts summarised in Table 5-1), with all cases leading to a negative T_m shift. The smallest destabilisation is observed with Mn(II) and Ni(II), although the melting curve obtained with Mn(II) is of

greater magnitude, facilitating data analysis. As Mn(II) was also the most favourable metal tested with AlkB, use of Mn(II) for both proteins would allow direct comparisons. Additionally, screens of a wide panel of 2OG oxygenases carried out by Inga Pfeffer (CJS lab, unpublished results) found that Mn(II) was generally the most favourable metal for addition in DSF analyses. Therefore use of Mn(II) in these studies would facilitate comparison of results across the 2OG oxygenase family.

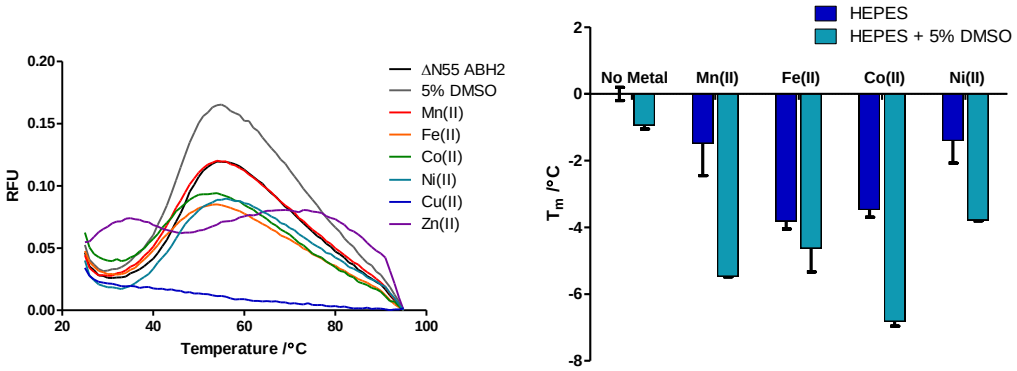


Figure 5-12: Differential scanning fluorimetry of ΔN55 ABH2 (4 µM) with addition of 50 µM metal (II) and/or 5 % DMSO (v/v); represented as melting curves and derived T_m shift values. Data represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

As with AlkB, Zn(II) appeared to significantly destabilise the ABH2 protein, here causing a double peaked curve which could not be fitted to a standard two state transition Boltzmann function. Again no T_m shift could be obtained with Cu(II).

T_m shift /°C	AlkB	AlkB + 5 % DMSO		ABH2	ABH2 + 5 % DMSO
No metal	0.00	-1.04		0.00	-0.95
Mn(II)	0.19	-1.01		-1.49	-5.48
Fe(II)	0.25	-4.27		-3.82	-3.99
Co(II)	4.68	-2.10		-3.48	-1.65
Ni(II)	1.43	-8.11		-1.41	-2.00
Cu(II)	n/a	n/a		n/a	n/a
Zn(II)	-15.18	-10.30		n/a	n/a

Table 5-1: T_m shifts for 4 µM ΔN11 AlkB and ΔN55 ABH2 with addition of 50 µM (II) metals, and/or 5 % DMSO (v/v).

5.4 Activity assays

5.4.1 Choice of substrate

The aim of the work described in this section was to obtain consistent and reliable kinetic and inhibition data. To enable this it was necessary to choose an appropriate AlkB substrate, and based on this optimise a suitable assay to follow the reaction. Previously it has been shown that the positively charged 1meA and 3meC methylated bases are the preferred substrates of AlkB[106,176] with 1meA possibly being the optimal AlkB substrate[325-326,335]. For this reason, and due to the possibility of more straightforward synthesis of the 1meA unit[336], the 1meA lesion was selected for use in the majority of assays.

The simplest AlkB substrate would be the individual 1meA base, nucleoside or nucleotide, and analysis of different portions of the nucleotide structure has revealed the minimal scaffold that is demethylated. Although it has been shown that the 1meA nucleoside can significantly stimulate 2OG turnover by AlkB it is not a substrate[325], as HPLC analysis shows that it is not demethylated[315,325]. Low uncoupled 2OG turnover activity was also observed with 1-methyl adenine, suggesting that the sugar moiety increases substrate recognition[325]. A different study detecting formation of demethylated products by HPLC found that 1-methyl adenosine monophosphate (1me-dAMP(5')) is the minimal substrate required for demethylation, containing the structural features required for productive binding and reaction at the active site[326]. Dimeric and trimeric sequences containing 1meA were also shown to be substrates of AlkB, however activity with longer (100+ nucleotides) oligonucleotide sequences was found to be much higher[326]. Substrates containing a phosphate 5' to the lesion show significantly higher activity than those lacking this phosphate[326], and this strong preference is rationalised by the crystal structure of AlkB in complex with a dT(1me-dA)dT trimer, which shows that R161 makes a salt bridge to the 5' phosphate of 1meA[192]. Crystal structures of AlkB crosslinked to a longer 13mer dsDNA fragment show a slightly different orientation for the oligonucleotide in the active site (although still retaining this important salt bridge)

most likely because the shorter trimer is unable to reach across the DNA binding surface[194]. However, direct binding of the oligonucleotide occurs to the DNA immediately flanking the methylated base.

The majority of previous assays investigating AlkB activity have used poly(monobase) oligonucleotides, which have been subjected to random methylation by treatment with the S_N2 methylating agents [^{14}C]-methyl iodide (MeI)[158], or *N*-[^{3}H]methyl-*N*-nitrosourea (MNU)[159]. The enzymatic activity may then be tracked by detection of released ethanol-soluble radioactive material. This approach requires the use of oligonucleotides containing only one base, and it is likely that multiple modifications may be introduced through the oligonucleotide, at sites other than that of the target modification. For example, on MeI treatment of poly(dA) the principal product is 1meA, but significant quantities of 3meA and 7meA are also produced[158]. Furthermore, the modifications produced will not have the same pattern for every oligonucleotide molecule present, therefore formation of a non-homogeneous population will complicate detailed data analysis. Other studies have used a poly(dT) background, with insertion of a limited number of different bases in the sequence, which can be modified chemically[158], as T is the least reactive base, therefore is less frequently modified[337]. At the rate of modification occurring in the genome it is unlikely that there will be multiple methylation modifications in the same region of oligonucleotides, particularly for short (~20mer) sequences. Even depurination, one of the most common modifications (at a rate of 2000-10000 per day[93]) occurs at a very low rate when compared to the size of the human genome (~3 billion bp)[338]. Furthermore, modification at different sites on the same base, which may occur on direct reaction with high concentrations of a methylating reagent, is extremely rare *in vivo*[97].

Although the use of oligonucleotides formed from only one base is somewhat artificial, as no secondary structure can be formed, their use can simplify detection methods, and facilitate more direct comparison between different modified bases. For this reason a poly(dT) 15mer oligonucleotide containing a single modified 3meC base was used in a

previous inhibition study in which the reaction was followed by capillary electrophoresis[339]. Here the 3meC lesion was positioned as the third base of the sequence from the 5' end, to ensure that this was well separated from the 3' fluorescein group. Despite the relative proximity of the lesion of the 5' end of the oligonucleotide, activity should still be optimal as the most striking structural requirement is for a 5'-phosphate, which is present, and as crystal structures show binding only to the bases directly adjacent to the lesion[194].

Oligonucleotides of 15-18 bases were used as substrates as they are a good compromise between use of a longer substrate for maximising activity in an environment more similar to that found *in vivo*, while simplifying synthesis, in which failed incorporations can form shorter sequences requiring separation. Furthermore, use of a shorter sequence will facilitate resolution of the methylated substrate and non-methylated product (differing in mass by 14 Da) in methods such as HPLC and mass spectrometry.

The decision was therefore made to use an 18mer DNA oligonucleotide substrate synthesised to contain one specific methylation lesion at a defined position within the sequence, in a mixed base background to allow formation of secondary structure: GCXAGGTCCCGTAGTGCG, where X is the modified base. This substrate maintains the modified base at the 3rd position, as had been used previously for fluorescent detection[339]. Assays carried out using similar sequences have also been reported while this work was underway, using 15mer[196], 16mer[340] and 29mer[335] oligonucleotide substrates containing specific individual modified bases. No detailed studies of the preference of the demethylation reaction for specific residues flanking the lesion have been carried out, but it appears that the reaction is sequence independent[175].

Oligonucleotide synthesis, using a solid-phase supported approach[341], was carried out by ATDBio, using a 1meA phosphoramidite synthesised by Anders Sejr Hansen. Typically, the final step of the synthetic process involves treatment with concentrated ammonia to cleave the newly synthesised oligonucleotide from the solid resin, with concomitant removal of protecting groups. However, 1meA is susceptible to an

ammonia-induced Dimroth rearrangement to the more stable N^6meA isomer[342] which is not a substrate of AlkB (although 1,N^6-ethenoadenine (εA) is successfully repaired[178-179]). Therefore to prevent this unwanted side-reaction the synthesis required the use of "ultra mild" solid phase synthesis conditions, using more labile protecting groups and resin linkages[336,343]. Similar oligonucleotides were also synthesised containing the different modified bases for further use in assays.

5.4.2 2OG turnover assay

As an initial test to determine the activity of the proteins produced, the 2OG turnover capability was assessed using the previously described radiolabelled 2OG turnover assay (Section 2.4.1). The proteins were incubated under standard assay conditions with a single-stranded 18mer oligonucleotide containing a 1-methyl adenine base, as this lesion has been identified as an optimal substrate for both AlkB[158-159] and ABH2[163].

Figure 5-13 shows that 2OG turnover is observed for each of the tested proteins, with a clear elevation of 2OG turned over in the presence of the oligonucleotide substrate. Uncoupled turnover by the proteins in the absence of substrate is expected, as self-hydroxylation of residue W178 in the active site of AlkB has been observed[344]. The 2OG turnover of ΔN11 AlkB in the absence of substrate is higher than that for the full length AlkB (full length 23 %; ΔN11 47.8 %). This may suggest a role for the *N*-terminal portion of the protein in preventing unproductive 2OG turnover. Inspection of the AlkB crystal structure reveals that the *N*-terminus is at the opposite side of the 3D protein structure from the entrance to the active site, therefore it is unlikely that any effect is caused via a direct interaction with the active site, but the presence of the additional sequence at the *N*-terminus may alter the overall 3D structure of the protein making turnover in the absence of the prime substrate more feasible. During development of the ΔN11 crystallisation construct, protein further truncated at the *N*-terminus (ΔN17 AlkB) was found to precipitate, suggesting that residues 11-17 are important for maintaining the correct protein structure[192], and form a short β-strand[195]. In contrast it appears that the *C*-terminal portion of the protein may be more important

for maintaining the enzymatic activity, as protein with a *C*-terminal His-tag is inactive[192].

Interestingly, the 2OG turnover of ABH2 is just over twice that of both AlkB constructs, but with a high (42 % of turnover in the presence of 1meA oligonucleotide) uncoupled proportion, suggesting that the protein is active with respect to turnover of 2OG, but is poorly capable of discriminating between whether the prime substrate is present or absent. Furthermore it has been reported previously that ABH2 has a preference for dsDNA over ssDNA substrates[167], although this is only observed in the presence of high concentrations of NaCl[173] or magnesium[175]. Therefore it is possible that on addition of the 1meA-containing single stranded oligonucleotide 2OG turnover may be stimulated, but that these cycles may be unproductive. ΔN55 ABH2 has been reported to be active with respect to 1meA demethylation, but no quantitative analysis of the turnover was reported[194].

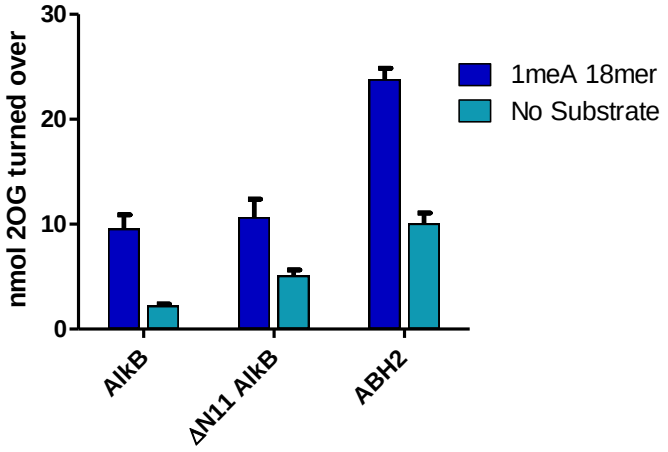


Figure 5-13: ^{14}C 2OG turnover by full length or *N*-terminally truncated AlkB, and the AlkB human homologue ABH2 in the presence or absence of an 18mer oligonucleotide containing a 1meA modified base. Reaction incubation time 30 minutes. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

The activity of AlkB was tested further by variation of the substrate present. An assay was carried out using a range of single stranded synthetic 18mer oligonucleotides, each with a different methylated base lesion at the same site in the sequence. The lesions investigated are structural analogues of the 1meA modification (Appendix A), with the methyl group present at the ring nitrogen, and have all been previously identified as substrates of AlkB[158-159,176-177]. As a comparison to this, a further oligonucleotide was

tested, containing 5meC, an important epigenetic marker. Here the methyl group is bound to a carbon atom at position 5 of the pyramidine ring, rather than the N-atom at the 3-position. Therefore demethylation of this group would require breaking of the stronger C-C bond, rather than the N-C bond present in the known AlkB substrates. Hydroxylation may be catalysed by AlkB, but it is unlikely that demethylation of this methyl group will follow. The hydroxylation of the 5meC methyl group is now known to be catalysed by the 2OG oxygenase TET1[51], forming the stable 5-hydroxymethyl cytosine (5hmeC) base[80].

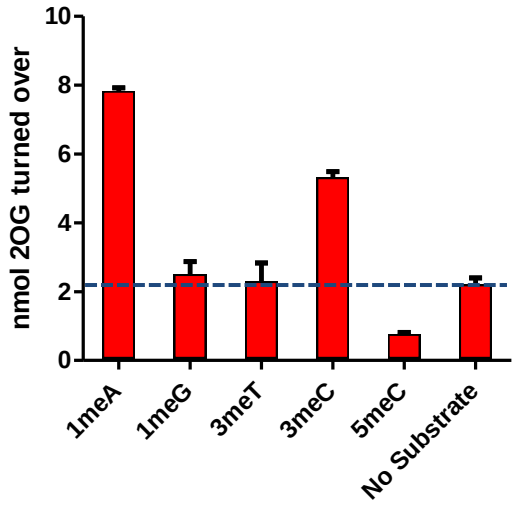


Figure 5-14: ^{14}C 2OG turnover by full length AlkB in the presence or absence of 18mer oligonucleotides containing a single methylated base at position 3. Reaction incubation time 15 minutes. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean. Blue dashed line indicates nmol 2OG turned over in the no substrate control reaction.

The highest 2OG turnover is observed for the oligonucleotide containing 1meA, the preferred substrate. An elevated 2OG turnover compared to the uncoupled background turnover is also observed with 3meC (68 % of the turnover in the presence of 1meA containing oligonucleotide). However, very low elevation above background 2OG turnover was observed with oligonucleotides containing 3meT and 1meG. These lesions were identified as AlkB substrates later than the original 1meA and 3meC lesions[177], and are less commonly formed and repaired *in vivo*[106]. The oligonucleotide containing the 5meC modification appears to repress 2OG turnover, as the turnover is reduced below the baseline level seen in the absence of substrate. It is possible that the DNA

strand may bind in the AlkB active site, but is unable to react. However on binding it may block entry or release of 2OG reducing the uncoupled turnover of 2OG. It has been reported that 1meG and 3meT nucleosides do not stimulate uncoupled 2OG decarboxylation[177], despite the observation of uncoupled turnover for 1meA, and that no demethylation of trimer oligonucleotides containing 1meG and 3meT is observed[200], suggesting that the interaction of these species with the active site may be weaker.

5.4.3 Discussion of limitations of the 2OG turnover assay

The radiolabelled 2OG turnover assay described above is useful for following the turnover of the 2OG co-substrate of the 2OG oxygenases. A particularly valuable feature of this assay is that it may be used when no substrate is known. Initially, 2OG oxygenase family members may be identified by excluding certain assay components or adding known generic oxygenase inhibitors, and observing decreases in the 2OG turnover. Additionally, possible substrates may be added, with observation of elevated 2OG turnover suggesting that the added reagent may be a substrate; this approach was successful for the identification of the substrate of FTO[165].

Despite these advantages, the assay also has a number of drawbacks. The assay methodology is not technically straightforward, and is very low-throughput. Even in the hands of experienced users the results obtained can be erratic and highly variable, often with a high background signal. There is therefore a requirement for high concentrations of reagents (particularly the expensive modified oligonucleotides) in order to observe significant signal above the high baseline noise and uncoupled turnover. The assay is also indirect, as the ^{14}C detected arises from turnover of the 2OG co-substrate, rather than of the prime substrate. As "uncoupled" turnover of the 2OG co-substrate is well documented for many members of the 2OG oxygenase family[5] (including DAOCS[345], JMJD2E[346] and PHD2[347]) this must be accounted for, and the indirect measurements are not ideal for detailed kinetic analysis.

A potentially preferable radioactive-based turnover assay has been reported, which uses [5-^{14}C]-oxoglutarate as the co-substrate, and releases ^{14}C-labelled succinate, which may

be detected by derivatisation with 2,4-dinitrophenyl hydrazine[348]. A more high-throughput usage of this methodology has also been reported[349], suggesting that the assay may be more amenable to use for inhibitor testing, despite the lengthy derivatisation steps required to obtain a readout. However the assay again only follows 2OG turnover, rather than the direct modification of the prime substrate, therefore uncoupled turnover of 2OG not dependent on reaction with the substrate would be difficult to dissect from coupled turnover. Optimisation of this methodology was attempted in this laboratory by Dr Elena Sanchez Fernandez, but the approach was found to be unsuccessful due to very high background signals, possibly due to degradation of the [5-^{14}C]-oxoglutarate starting material, or the presence of impurities.

5.4.4 Possible alternative assays

An assay exploited recently for other 2OG oxygenases uses AlphaScreen™ (Perkin Elmer) technology. This antibody-based chemiluminescence technique can be miniaturised to the nanomolar scale, and has been applied successfully for kinetic studies of histone demethylases[327], and more recently for the HIF hydroxylases (unpublished results, Dr Akane Kawamura, Dr Nathan Rose). The assay output is the emission of light upon binding of a specific antibody to either the substrate or the product of the reaction. Excluding antibodies to the established epigenetic marker 5meC, or more recently the "sixth DNA base" 5hmeC, few antibodies to modified DNA nucleosides are commercially available. Unfortunately antibodies are not available to the less stable, and less common, 1meA modification (or its analogues) precluding detection of the reaction substrate. It is possible that *in vitro* a specific substrate could be used containing in its sequence a modified base, but not the unmodified version, meaning that when the methylation lesion was removed the unmodified base would be revealed, and could be detected by an antibody specific to the unmodified base. However this approach would not be appropriate in more realistic or lengthy sequences, which would contain the unmodified base, and it would not be possible to directly compare between modified bases in the same background sequence.

A further approach, useful for assaying double stranded oligonucleotides, is the use of methylation sensitive restriction enzymes. Oligonucleotides are designed containing a specific methylation in a restriction site. Once this methylation is enzymatically removed the restriction enzyme can cleave the sequence, and the results visualised by agarose gel electrophoresis. This approach has been used successfully for assays of AlkB and ABH2[173-174,239], however it is again not a high-throughput approach, and often involves the use of [^{32}P]-labelled oligonucleotides, further reducing the throughput.

NMR has been used to follow the activity of other 2OG family members, both through the turnover of 2OG to succinate, and through direct detection of substrate modification[346-347]. This technique is particularly useful for quantitatively following the progress of the reaction, but requires many hours of valuable machine time, and is thus not well suited to screening of large numbers of possible inhibitors.

5.4.5 Chromatography based techniques

Various mass spectrometric techniques were also attempted or considered. Reacted oligonucleotides may be digested to the component nucleosides, either by acid hydrolysis[158] or enzymatically, using nuclease P1, snake venom phosphodiesterase and alkaline phosphatase[350]. Detection is then possible by 2D thin layer chromatography (TLC), as was used for the identifications of base J[41] and 5hmeC[80]. A similar method has been more recently reported using HLPC to separate nucleosides, followed by ESI to observe the intact nucleosides, and collision induced dissociation (CID) to break the glycosidic bond and allow observation of the base[351], however again these methods are not high-throughput.

HPLC of the intact oligonucleotide followed either by UV or mass spectrometric detection was carried out. The LC conditions were optimised for the 18mer 1meA oligonucleotide to enable separation of the two species differing in mass by 14 Da, however further optimisations would be required for each oligonucleotide used considering the base sequence, oligonucleotide length and the modified base present in the sequence. Both product and reagent were observed simultaneously, with

quantification possible by integration of the chromatographic peaks. However, preparation of samples prior to analysis was lengthy as protein was removed from the reaction sample by precipitation on addition of methanol. The oligonucleotide was then dried and resuspended in a smaller volume of liquid, to facilitate analysis. During this process some DNA was lost, precluding quantitative comparison. As the post-reaction processing steps are significant this method is not particularly rapid.

A similar LCMS method (using ESI-TOF MS detection) for analysis of the reaction of AlkB has been published, using a 16mer oligo with the lesion at the 9th position of the oligo. The authors obtained good signal using only 20 pmol oligonucleotide, however the study was a proof of principle analysis to directly verify repair of the spectrum of lesions by AlkB, rather than a high-throughput screening method[340]. Since the recent identification of the 5hmeC modification, several new highly sensitive mass spectrometric methods have been published, primarily for analysis of the proportion of 5hmeC in genomic DNA samples[352]. In the future these techniques may have further utility for more sensitive analysis of the transient and less common damaged bases.

5.4.6 FDH-coupled assay

An assay method that has been successfully applied for some 2OG oxygenases involves coupling the formation of the formaldehyde co-product (from demethylation of the DNA base) to production of a fluorophore. This method uses *Pseudomonas putida* formaldehyde dehydrogenase (FDH), which oxidises formaldehyde to formic acid, using the coenzyme nicotinamide adenine dinucleotide (NAD^+) as a co-substrate (Figure 5-15)[353]. In carrying out this reaction, the non-fluorescent NAD^+ is reduced by FDH to NADH, which has a fluorescence emission peak at 460 nm[354]. The appearance of NADH in solution can thus be tracked to monitor the progress of the AlkB demethylation reaction. The absorbance spectra of NAD^+ and NADH also differ, with an additional peak for NADH at 340 nm[355]. However, the fluorescence characteristic of NADH is used rather than the absorbance as the sensitivity of the assay is increased by

an order of magnitude due to the significantly reduced background signal for fluorescence[356].

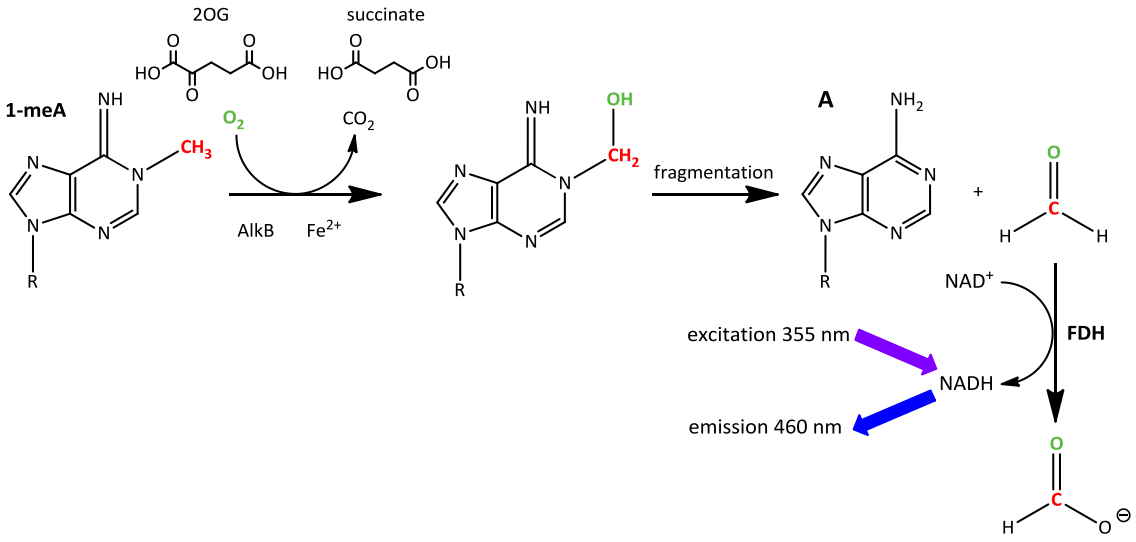


Figure 5-15: Scheme for AlkB demethylation coupled to the FDH-catalysed oxidation of formaldehyde to formic acid. This primary reaction is coupled to the reduction of NAD^+ to NADH, allowing fluorescence detection of the NADH produced by excitation at 355 nm and emission at 460 nm.

Although this assay does not directly identify the oligonucleotide itself, it is still able to directly follow the productive catalytic assay as the formaldehyde present in the assay mixture will only be produced by the demethylation of 1meA, with the carbon of the formaldehyde molecule originating from the methyl group lesion.

The use of this assay to follow the reaction of AlkB was initially published by Roy et al., who investigated some of the important considerations for establishment of the assay[335]. The use of the common buffer Tris.HCl was found to be incompatible with the FDH reaction as, due to the presence of three primary alcohol groups in the structure, this is a substrate for FDH[357]. As an alternative buffer, HEPES is used here. An important consideration for coupled assays is ensuring that the second reaction is coupled effectively to the first, and here the significantly higher specific activity of FDH than AlkB means that no "lag" period is observed[335]. Additionally, NAD^+ is used at a saturating concentration to ensure that the reaction is dependent only on the concentration of formaldehyde.

In this previous study the assays were run as timecourses in batches of four reactions, running an assay triplicate alongside a blank reaction[335]. This was again not a particularly efficient or high-throughput approach, so this was further developed. In this laboratory an FDH-coupled assay had previously been optimised to follow the reaction of the histone demethylase JMJD2E, which also releases formaldehyde on demethylation of methylated lysines[358]. To increase the throughput of the assay, a fluorescence plate reader was used to record data, allowing several samples to be run simultaneously, and monitored as a timecourse with readings recorded every 30 s. Signals were found to be most sensitive when using 100 µl reactions in 384 well plates, although due to the limitation of how many assays can be manually initiated simultaneously, and how regularly readings could be taken, a maximum of 24 wells could be monitored per assay. The use of this assay format also allows for the possibility of miniaturisation for true high-throughput application, and this has been successfully achieved for the histone demethylases to identify novel potent inhibitors[359-360]. The previous FDH-coupled study of AlkB used the alternative coenzyme 3-acetylpyridine adenine dinucleotide ($APAD^+$) rather than NAD^+, observing increased stability and sensitivity[335], but no significant advantage was observed on testing in JMJD2E assays (personal communication, Dr Nathan Rose), therefore NAD^+ was used instead.

5.5 FDH assay development

5.5.1 Initial assay trials

To determine whether the purified AlkB is capable of demethylating 1meA-containing oligonucleotides, and whether this assay is an appropriate method for following this reaction, an initial FDH-coupled assay was carried out using standard assay conditions, previously used for JMJD2E[358].

The typical method previously used involves two reagent master mixes: a "substrate mix" containing FDH, NAD^+, substrate and 2OG, and an "enzyme mix" containing enzyme, iron and ascorbate. The enzyme mix is placed into individual wells of a 384

well plate, then initiation of the reaction occurs on addition (using a multichannel pipette) of the substrate mix, giving a final assay volume of 100 µl in each well. The reagents are mixed, then immediately read in a fluorescence plate reader. Previous studies on AlkB showed that extended incubation of AlkB with iron and ascorbate led to specific protein cleavage[315]. Therefore an alternate setup was used to prevent incubation of these reagents, using a substrate mix containing 500 µM NAD^+, 40 µM methylated oligonucleotide, 200 µM 2OG, 10 µM iron and 100 µM ascorbate, and an enzyme mix of 2 µM AlkB with 0.1 U FDH per reaction.

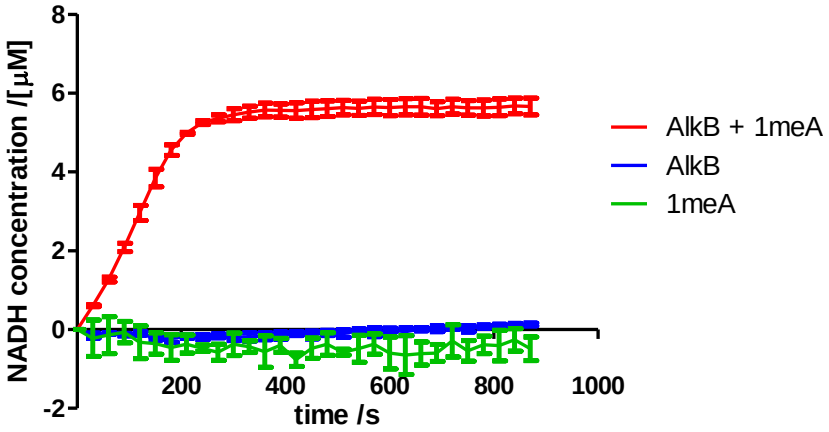


Figure 5-16: AlkB demethylation timecourse assay followed using a FDH-coupled assay. Reactions were carried out with all assay components (AlkB + 1meA, red), without oligonucleotide substrate (AlkB, blue), or without AlkB (1meA, green). Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

Conversion of results from relative fluorescence units to the concentration of NADH produced (which should be equivalent to the concentration of formaldehyde produced, and therefore the amount of 1meA oligonucleotide demethylated) is possible using formaldehyde calibration curves (prepared by Dr Nathan Rose and Dr Akane Kawamura). FDH assays were run using a range of known formaldehyde concentrations, to give the RFU observed per µM formaldehyde consumed, recorded as the concentration of NADH produced. The values generated were used throughout for conversion of relative fluorescence units to NADH concentration.

The results obtained for an AlkB timecourse assay (Figure 5-16) show that the reaction has occurred successfully under these conditions, with a clear increase in signal

corresponding to formation of NADH due to the reaction of FDH with the produced formaldehyde. Following an initial lag period due to initiation of the reaction and allowing time for mixing of the reagents, the reaction has an approximately linear phase in the 30-180 s time period. In the absence of either AlkB or the oligonucleotide substrate a baseline signal level is observed, with no significant increase occurring over the course of the assay, suggesting that the increase observed in the reaction containing both AlkB and the methylated oligonucleotide substrate is due to the true enzymatic reaction, rather than a side reaction occurring between other assay components (for example the reaction buffer), or formaldehyde already present or produced by a different means.

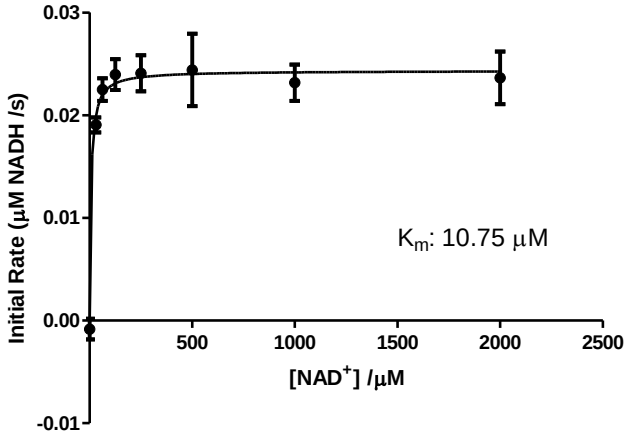


Figure 5-17: Variation of NAD^+ concentration for AlkB reaction with a 1meA oligonucleotide, followed using a FDH-coupled reaction. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

An assay was carried out varying the concentration of NAD^+ (Figure 5-17), showing that with the standard NAD^+ concentration (500 μM) the reaction is zero order, therefore is sufficient for the coupled reaction. No inhibition is observed at concentrations of up to 2 mM NAD^+ (50 fold greater than the concentration of oligonucleotide used).

These initial assays were run using conditions based on those for JMJD2E[358]. However within the 2OG oxygenase family, individual members have very different co-substrate requirements, for example PHD2 has a strong requirement for ascorbate[293], and the K_m of FIH for 2OG is higher that that for other similar oxygenases[361-362]. Therefore,

assuming that a steady-state reaction occurs (Equation 5-1), and that the Michaelis-Menten equation can be applied (Equation 5-2), it is necessary to determine the Michaelis-Menten constant (K_m) values (Equation 5-3) for each of the assay components for the reactions of both AlkB and ABH2. To carry out valid kinetic and inhibition assays it is necessary to ensure that the assay components are well in excess of K_m, so the reaction rate is not limited.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P$$

Equation 5-1: Assumed enzyme reaction mechanism for the Michaelis-Menten model, using the steady state approximation where $k_2 >> k_{-2}$. E: enzyme; S: substrate; ES: enzyme-substrate complex; P: product; k_x: rate constants for reaction stage indicated.

$$\frac{d[P]}{dt} = \frac{V_{max}[S]}{K_m + [S]}$$

Equation 5-2: Michaelis-Menten equation for enzyme kinetics, based on the model in Equation 5-1. [P]: product concentration; V: reaction rate; V_{max}: maximum rate achieved by the system under saturating substrate concentrations; [S]: concentration of substrate S; K_m: substrate concentration at which the reaction rate is half of V_{max}.

The K_m can give an indication of the affinity of enzymes for a substrate, with a lower K_m suggesting that the maximal catalytic efficiency is achieved at low substrate concentrations. Therefore the ES complex is held together very tightly and rarely dissociates without the substrate first reacting to form product. The rate constant k_2 (Equation 5-4) is also referred to as k_{cat}, the turnover number, representing the maximum amount of substrate the enzyme is capable of turning over per second.

$$K_m = \frac{k_2 + k_{-1}}{k_1}$$

Equation 5-3: Definition of the Michaelis-Menten constant (K_m), where k_x are rate constants as described in Equation 5-1.

$$\boldsymbol{V_{max} = k_2[E]_0}$$

Equation 5-4: Definition of the maximal reaction rate (V_{max}) under saturating substrate conditions, where k_2 is the rate constant described in Equation 5-1, and $[E]_0$ is the initial enzyme concentration.

5.5.2 Determination of kinetic parameters for AlkB and ABH2 reactions

FDH-coupled assays were carried out for both AlkB and ABH2 using an 18mer 1meA-containing oligonucleotide substrate (Section 5.4.1), in each case varying the concentration of one of the assay components. Each of the other assay components was maintained in excess to ensure that these were not rate limiting. This allowed determination of the K_m values for each of the assay components, through fitting the results obtained to the Michaelis-Menten equation, using GraphPad Prism, and extracting the kinetic parameters from the fitted curves. This method was found to be more consistent than a linear plot analysis, although obtained results were comparable between the two methods. Although initial kinetic studies of the substrates of AlkB and mouse ABH2 (discussed below) have been published previously, no kinetic analyses of the co-substrate requirements of either enzyme have been published to date. Results are depicted in Figure 5-18, with the kinetic parameters obtained summarised in Table 5-2.

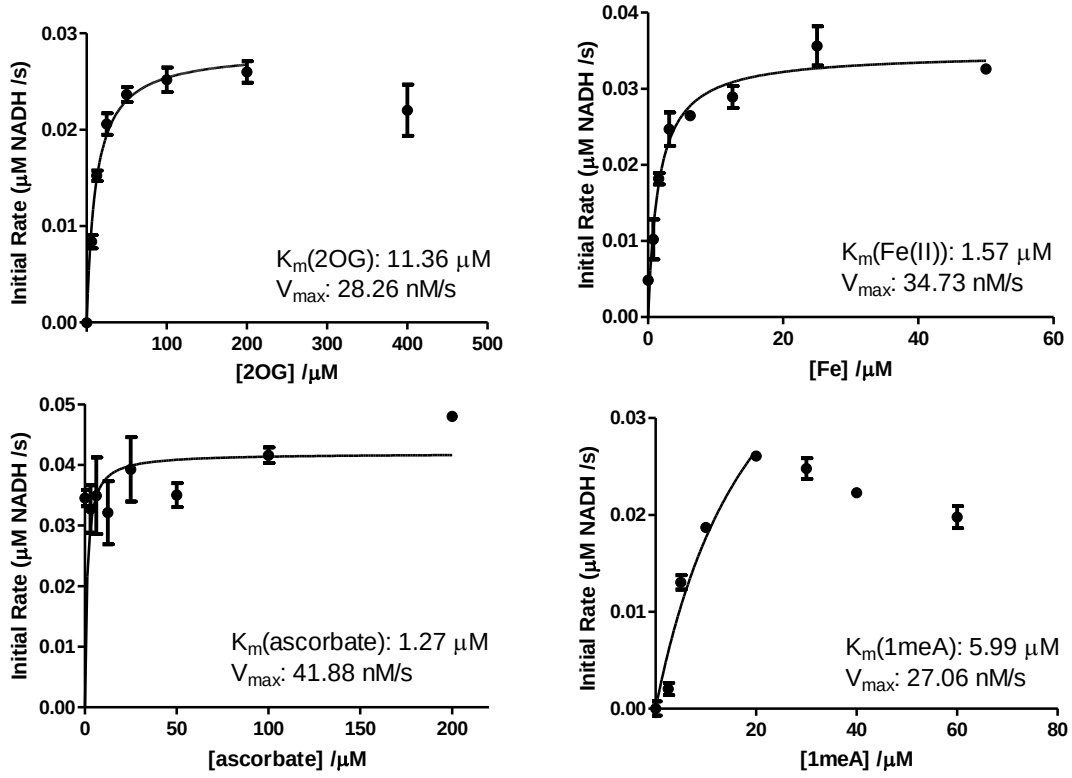


Figure 5-18: Michaelis-Menten plots for AlkB assay components, determined using the FDH-coupled assay. Kinetic parameters are indicated on the graphs, and are derived from fitting the Michaelis-Menten equation to the data obtained, using GraphPad Prism. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

Substrate	K_m /μM		V_{max} /μmol s^{-1}		k_{cat} /s^{-1}		k_{cat}/K_m /M^{-1} s^{-1}	
	AlkB	ABH2	AlkB	ABH2	AlkB	ABH2	AlkB	ABH2
18mer 1meA oligonucleotide	6.0 ± 1.9	4.2 ± 1.5	0.027 ± 0.002	0.0069 ± 0.0005	0.014 ± 0.001	0.0035 ± 0.0003	2340	825
2OG	11.4 ± 1.3	4.0 ± 1.2	0.028 ± 0.0008	0.0075 ± 0.0003	0.014 ± 0.0004	0.0038 ± 0.0002	1230	934
Fe(II)	1.6 ± 0.29	1.8 ± 0.5	0.035 ± 0.001	0.010 ± 0.0006	0.017 ± 0.0007	0.0052 ± 0.0003	10800	2860
Ascorbate	1.3 ± 0.55	177 ± 37	0.042 ± 0.005	0.023 ± 0.003	0.021 ± 0.0005	0.012 ± 0.001	16500	67.8

Table 5-2: Kinetic data obtained from FDH assays for the reactions of AlkB and ABH2 with an 18mer 1meA oligonucleotide. Data are derived from fitting the Michaelis-Menten equation to the data obtained using GraphPad Prism, and are presented as the mean of triplicate experiments ± the standard error of the mean.

The reaction of AlkB with variation of the iron concentration follows standard Michaelis-Menten kinetics across the range tested. Varying the ascorbate concentration appears to have little effect on the activity of AlkB, which contrasts with a previous study of AlkB in which the activity was observed to be strongly dependent upon the presence of ascorbate[325]. However Welford et al. followed the activity of AlkB with the 1meA nucleoside, which is not demethylated, using the radiolabelled 2OG turnover assay described previously. Therefore it is possible that in these assays the uncoupled reaction has a high dependence on ascorbate, as it is postulated that the likely role of ascorbate is recovery of the Fe(II) centre from the oxidised Fe state[292], which would be likely to occur more frequently when no primary substrate is present. Inhibition of AlkB at high concentrations of 2OG was also reported in this previous study[325], which is again observed here. Determination of the K_m for 2OG was achieved by excluding the highest concentration of 2OG used (400 μM) from the values fitted to the curve (as shown in Figure 5-18). Additionally, a reduction in activity at elevated substrate concentrations was observed, possibly due to substrate or product inhibition, a phenomenon which has been observed for PHD2[274]. This decreased activity therefore makes assigning a K_m difficult, as determination of which points should be fitted to the Michaelis-Menten curve is not straightforward. However the curve was fitted to data points at lower concentration to obtain an approximate K_m, excluding the possible inhibition effects.

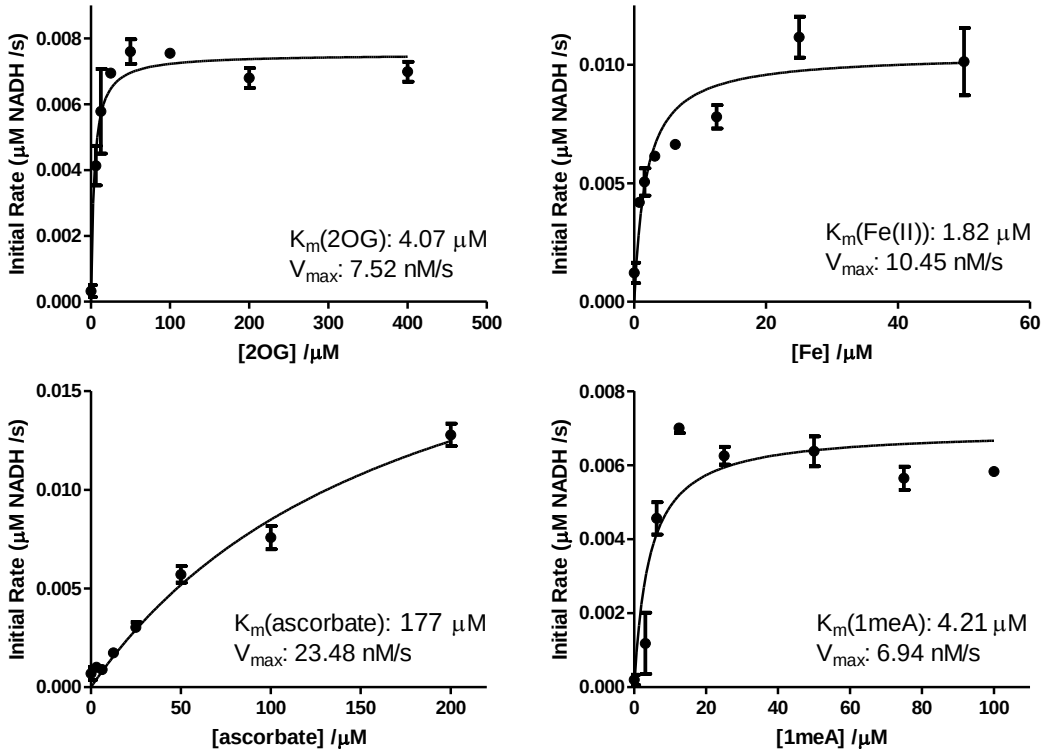


Figure 5-19: Michaelis-Menten plots for ABH2 assay components, determined using the FDH-coupled assay. Kinetic parameters are indicated on the graphs, and are derived from fitting the Michaelis-Menten equation to the data obtained, using GraphPad Prism. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

A similar set of reactions was carried out with ABH2 to determine the K_m values for each of the assay components (Figure 5-19, summarised in Table 5-2). Inhibition of ABH2 is not observed at high concentrations of 2OG, and the K_m is less than half that observed with AlkB, suggesting that 2OG is bound more tightly by ABH2 during the catalytic cycle. Furthermore, as inhibition is not observed at high 2OG concentrations this may suggest that the activity of ABH2 may be inhibited by a different class of inhibitors from the simple 2OG mimics previously studied for AlkB[325,339], suggesting the possibility of selectivity within the 2OG-dependent oxygenase family. The iron requirement of ABH2 is very similar to that of AlkB, with similar K_m values obtained. For both proteins the K_m for iron is just over 1 μM, in reactions where the enzyme concentration is at 2 μM, suggesting that the proteins have purified with Fe present in the active site, as has been observed previously for AlkB[363], and other 2OG oxygenases[332] although this was not clear from the mass spectra obtained for the purified AlkB protein used here (Section 5.3.2). It is possible that ABH2 may show

slight inhibition at higher substrate concentrations, but this is not as apparent as with AlkB. The most striking difference between the reactions of AlkB and ABH2 is the ascorbate dependence, as the K_m of ABH2 for ascorbate is over 100 times that observed with AlkB, with a ~250-fold lower k_{cat}/K_m. Therefore the ABH2 reaction is strongly dependent on the presence of ascorbate, similar behaviour to that of other oxygenases, such as PHD2[293].

Protein	Substrate	K_m /μM	k_{cat} /s^{-1}	k_{cat}/K_m /M^{-1} s^{-1}	Method	Reference
AlkB	T(1meA)T	2.8 ± 0.9	0.123 ± 0.001	43300	HPLC	Koivisto et al.
	methyated poly(dA)	1.4 ± 0.2	0.195 ± 0.022	143000		
AlkB	ss 19mer 1meA	5.4 ± 0.9	0.062 ± 0.003	11480	FDH assay	Roy et al.
	ds 19mer 1meA	6.2 ± 1.3	0.052 ± 0.004	8390		
	ss 19mer 3meC	3.4 ± 0.6	0.037 ± 0.005	10880		
	ds 19mer 3meC	9.3 ± 2.4	0.055 ± 0.006	5910		
ΔN11 AlkB	T(1meA)T	1.4 ± 0.8	0.045 ± 0.013	31670	HPLC	Yu et al.
	T(3meC)T	24 ± 5	0.350 ± 0.067	15000		
	T(εA)T	60 ± 14	0.0022 ± 0.0008	33		
	CA(1meA)AT	0.06 ± 0.01	0.090 ± 0.022	1617000		
	CA(3meC)AT	0.29 ± 0.03	0.380 ± 0.170	1305000		
ΔN11 AlkB	ss 18mer 1meA	6.0 ± 1.9	0.014 ± 0.001	2340	FDH assay	This study
hABH2	ss 18mer 1meA	4.2 ± 1.5	0.0035 ± 0.0003	825		
mABH2	ss 24mer 1meA	0.018 ± 0.023	3.3 ± 0.27	18000000	Restriction digest	Lee et al.
	ds 24mer 1meA	0.032 ± 0.073	13.7 ± 2	42900000		
	ss 24mer 3meC	0.082 ± 0.022	1.06 ± 0.12	12900000		
	ds 24mer 3meC	0.17 ± 0.027	8.84 ± 0.026	52900000		

Table 5-3: Summary of kinetic data previously reported for substrates of AlkB and *Mus musculus* ABH2 (mABH2). ss: single stranded, ds: double stranded. Data taken from[173,200,326,335].

Previously published kinetic data for AlkB and *Mus musculus* ABH2 (mABH2) with a range of substrates is summarised in Table 5-3, alongside the data obtained here. Although it is difficult to make quantitative comparisons between the data due to the different assay methodologies and conditions used, it is useful to qualitatively assess the results, particularly through comparison of k_{cat}/K_m, a measure of the catalytic efficiency of the enzyme under sub-saturating conditions. This shows many of the previously described substrate preferences of the two enzymes. AlkB prefers ssDNA, with greater efficiency with longer sequences, and is most efficient with 1meA substrates, whereas

ABH2 has a preference for 1meA in ssDNA, and 3meC in dsDNA, but overall is more efficient with dsDNA.

The K_m of AlkB for the 1meA oligonucleotide tested in this study is very similar to the published value obtained with a 19mer 1meA ssDNA oligonucleotide (6 μM vs. 5.4 μM)[335], although the k_{cat} values obtained here are lower than previously published. The results obtained in this study suggest that ABH2 has a lower substrate turnover than AlkB, as shown by the lower k_{cat} and k_{cat}/K_m values. In contrast, the published values for mABH2 are significantly higher than those previously published for AlkB, or are observed here with human ABH2. A restriction enzyme digest method was used to obtain these data, which used very low reagent concentrations (picomolar enzyme and nanomolar substrate) which will affect the magnitude of the results obtained. Furthermore, the two ABH2 proteins tested are derived from different organisms, therefore despite their homology (75 % identity) their activities may differ slightly.

Component	AlkB	ABH2
Enzyme	2 μM	2 μM
18mer 1meA oligonucleotide	20 μM	20 μM
2OG	100 μM	150 μM
Fe(II)	10 μM	10 μM
Ascorbate	100 μM	320 μM
NAD	5 mM	5 mM
FDH	0.1 U	0.1 U

Table 5-4: Summary of optimal assay conditions for FDH-coupled reactions for AlkB and ABH2 with 18mer oligonucleotide substrates.

Following determination of the kinetic parameters, the results were used to establish optimal assay conditions, ensuring that the reagents were in excess of K_m, and would not limit the reaction rate (Table 5-4). Although the K_m for 2OG of ABH2 is lower than that for AlkB, a higher concentration of 2OG was chosen for use in ABH2 reactions. In these activity assays it was decided to maintain a large excess of 2OG to allow the reaction to proceed fully, but as inhibition of AlkB was observed at high 2OG concentrations a lower concentration was used in this case. As a true K_m could not be

obtained for the 1meA oligonucleotide substrate with AlkB, the substrate concentration at which maximal activity was achieved was used for further assays.

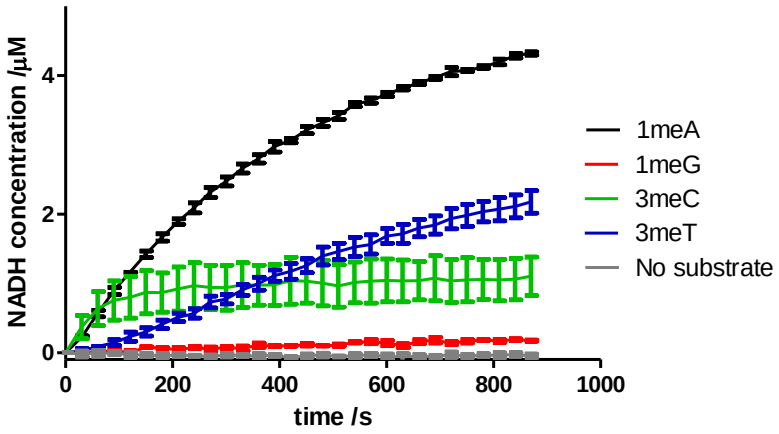


Figure 5-20: FDH-coupled assay for ABH2 with 18mer oligonucleotide substrates (20 µM), each containing a different methylated base. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

To compare the activity of ABH2 with different base modifications an FDH assay was carried out using oligonucleotides with the same base background, but with different base modifications (Figure 5-20). Greatest activity was observed with the oligonucleotide containing 1meA, suggesting that this modification is the preferred substrate of ABH2. This has been observed previously, with a previous study reporting four-fold greater activity with 1meA than 3meC[163]. No significant activity above the background enzyme-only background was observed with the oligonucleotide containing the 1meG base, but no activity has been reported for ABH2 with 1meG[177]. As it appeared that the 3meT-containing oligonucleotide may be a good substrate of ABH2, this was further tested in an assay using a range of oligonucleotide concentrations. However this approach was unsuccessful, as it appeared that overall the activity was substantially lower than that with the 1meA oligonucleotide, with significant substrate/product inhibition observed at higher substrate concentrations, as had been observed previously for AlkB with the 1meA oligonucleotide.

5.5.3 N-*Terminally truncated constructs*

As the *N*-terminally truncated constructs of both AlkB and ABH2 are produced in a higher yield, and have previously been successfully crystallised, it would be useful to

use these proteins in activity assays, to be able to compare results obtained more directly with crystal structures.

The activity of the two AlkB constructs was found to be very similar (Figure 5-21), in agreement with previous reports of the activity of the ΔN11 construct (although no quantitative comparison of the activities of the proteins was given)[192]. The initial rate of the *N*-terminally truncated protein is 82 % that of the full length protein, and after the 15 min incubation 88 % of the total turnover is recorded. This is slightly different from the previous observation made using the 2OG turnover assay (Section 5.4, Figure 5-13), in which the ΔN11 construct showed 112 % of the turnover of the full length construct, although after a 30 min incubation. However it is likely that this was due to the increased uncoupled 2OG turnover for the ΔN11 construct observed in the 2OG turnover assay, highlighting the difficulty in interpreting data from assays that follow the reaction of the prime substrate indirectly.

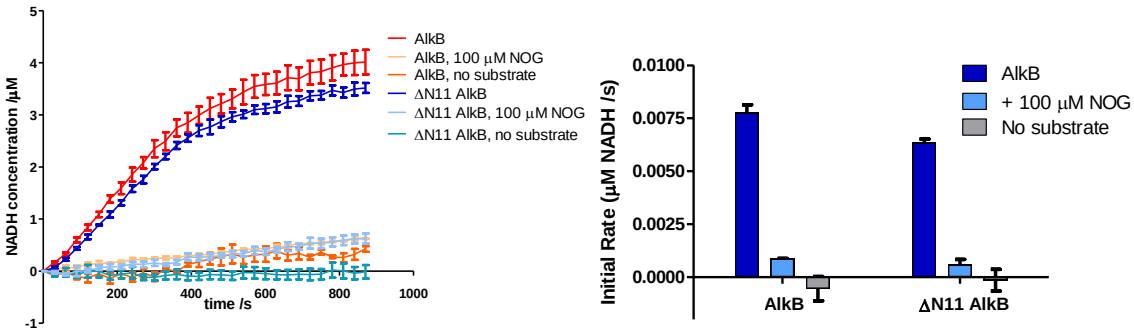


Figure 5-21: Comparison of the relative activities of the full length and ΔN11 AlkB constructs in an FDH-coupled assay, with an 18mer 1meA oligonucleotide substrate. Data are represented as the reaction timecourse, and the initial reaction rates. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

As may be anticipated from the significantly larger *N*-terminal truncation for the ΔN55 ABH2 protein, the truncated ABH2 protein shows a significantly reduced activity as compared to the full length protein, with the initial rate only 29 % (and the total turnover during the 15 min incubation 36 %) that of the full length protein (Figure 5-22). The initial report of the ΔN55 ABH2 construct describes demethylation activity, but again no quantitative comparison to the full length wildtype protein was made[194]. The activity of the ΔN55 ABH2 construct is still distinguishable from the reaction

baseline, obtained in the absence of the oligonucleotide substrate, but is lower than that for the wildtype protein treated with *N*-oxalylglycine, a 2OG mimic, and a generic 2OG oxygenase inhibitor. Therefore the truncated ABH2 construct is not well suited for use in activity assays.

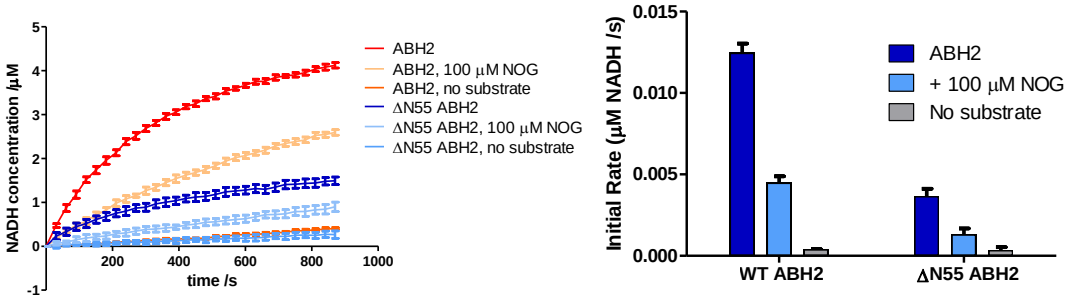


Figure 5-22: Comparison of the relative activities of the full length and ΔN55 ABH2 constructs in an FDH-coupled assay, with an 18mer 1meA oligonucleotide substrate. Data are represented as the reaction timecourse, and the initial reaction rates. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

5.5.4 Assay miniaturisation

The main limiting factors for carrying out higher-throughput screens are the cost of the modified oligonucleotide substrate, and the availability of small molecule inhibitors. Therefore having established the required conditions for assays of the two proteins further optimisation was attempted, through miniaturisation. However, despite wishing to reduce the quantities of reagents used, it is still important to obtain good quality activity data. In order to assess the quality and validity of data obtained in the assays standard statistical methods were used for analysis.

Firstly, the signal to noise ratio can be evaluated (Equation 5-5). To obtain good data the signal to noise ratio should be maximised to obtain the greatest difference between any background signal (noise) and the signal obtained for positive assay reactions. Despite the value of this parameter, it does not account for all important aspects of the assay under question. Therefore when establishing assays for use in high-throughput screens the Z-factor (Equation 5-6) has been described to allow evaluation of the validity of an assay, and the likelihood of successful identification of hit compounds from a high-throughput screen[364]. The Z-factor is also referred to as the screening

window coefficient, and reflects the signal dynamic range (the difference between mean signals) and variation of the data obtained in assays, assessed by the "separation band", the difference between 3 standard deviations of the negative and positive values. An interpretation of obtained Z/Z'-factor values is summarised in Table 5-5. However, it must be remembered that the Z-factor is a theoretical parameter, designed for interpretation of results obtained in a model system, from automated, robotic assays. Therefore it does not include any allowance for additional errors occurring in non-ideal manual assays, where pipetting and mixing variations are likely to be greater. Thus Z-factor values derived for manual assays, as carried out here, will underestimate errors.

$$Signal\ to\ noise\ ratio = \frac{\mu_p - \mu_n}{\sqrt{\sigma_p^2 + \sigma_n^2}}$$

Equation 5-5: Definition of signal to noise ratio[364]. μ_p: mean of positive control; μ_n: mean of negative control; σ_p: standard deviation of positive control; σ_n: standard deviation of negative control.

The Z'-factor is the corresponding Z-factor for control data, allowing evaluation of the suitability of new assays. To determine this, a "positive" value is defined as the maximum signal contained, for example from a reaction carried out under optimal assay conditions, and the "negative" value is defined as the baseline signal for a negative control reaction. In this case two negative situations where no reaction should be observed were compared, with reactions carried out in the absence of either enzyme or substrate.

$$Z' - factor = 1 - \frac{3(\sigma_p + \sigma_n)}{|\mu_p - \mu_n|}$$

Equation 5-6: Definition of Z'-factor[364]. μ_p: mean of positive control; μ_n: mean of negative control; σ_p: standard deviation of positive control; σ_n: standard deviation of negative control. The screening window is defined as the mean ± 3 standard deviations, and the separation band is defined as the difference between the means.

Z-factor value	Structure of assay	Conclusion
1	σ = 0 (no variation) or dynamic range → 0	An ideal assay
1 > Z ≥ 0.5	Separation band is large	An excellent assay
0.5 > Z > 0	Separation band is small	A double assay
0	No separation band – sample and control variation bands touch	A yes/no assay
< 0	No separation band – sample and control variation bands overlap	Screening essentially impossible

Table 5-5: Categorisation of screening assay quality using the calculated Z-factor. σ: standard deviation. Table adapted from Zhang et al.[364].

The initial assay volume used for assays of both AlkB and ABH2 was 100 µl, therefore a reduction of the assay volume to 20 µl, 25 µl and 50 µl was tested. Given the use of relatively small volumes other modifications were made to the assay conditions to reduce the likelihood of significant evaporation occurring during the course of the assay. Firstly, the temperature at which the assays were run was reduced from 37 °C to 25 °C, and secondly the 50 mM HEPES buffer used for the assays was supplemented with detergent (0.05 % Tween 20).

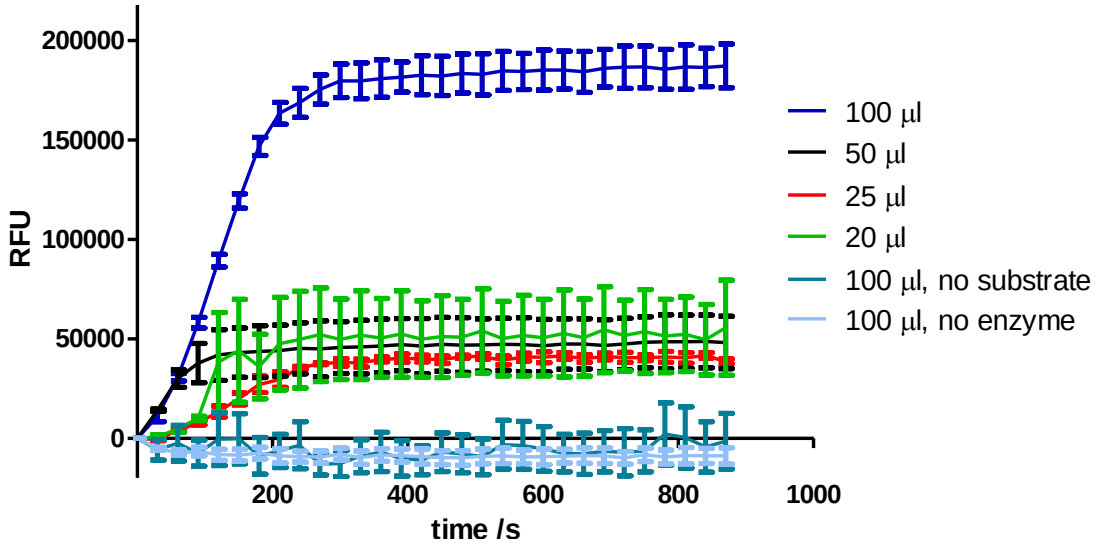


Figure 5-23: Miniaturisation test for the FDH-coupled reaction of AlkB with 1meA oligonucleotide, assays carried out in assay volumes indicated. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean. Results given are recorded as relative fluorescence units (RFU) to allow more direct comparison between assays carried out in different volumes, as conversion to the amount of NADH produced requires different calibration curves for each assay volume.

The timecourse results for the FDH-coupled reactions of AlkB with 1meA in the different assay volumes (Figure 5-23) show that the reaction remains successful at a

reduced temperature, and in the presence of detergent. It appears that the 100 µl assay volume is optimal, which is verified on considering the statistical analysis, with Z'-factor values of 0.50 compared to the no substrate control, and 0.73 compared to the no enzyme control. This suggests from the definitions of the Z'-factor that under these conditions this is an excellent assay.

To a first approximation it appears that graphically the results obtained from the smaller assay volumes are generally comparable, although a significant variation in the obtained data is observed for the 20 µl reaction. However, referral to the statistical analysis of the assays shows that the 25 µl reaction volume has overall higher signal to noise and Z'-factor values than the 50 µl assays, despite the reduced assay volume.

Initial rate	**Positive control**		**No Substrate as negative control**				**No Enzyme as negative control**			
(RFU/s)	Average	SD	Average	SD	S/N	Z'	Average	SD	S/N	Z'
100 µl	852.02	49.05	-16.08	96.03	8.05	0.50	-35.58	29.78	15.47	0.73
50 µl	484.63	97.98	0.91	2.94	4.93	0.37	1.77	45.81	4.46	0.11
25 µl	161.95	19.16	-12.09	17.92	6.63	0.36	-23.26	4.53	9.41	0.62
20 µl	112.49	26.17	9.13	9.22	3.73	-0.03	-20.14	7.83	4.86	0.23

Table 5-6: Statistical analysis of FDH-coupled assay for AlkB with 1meA oligonucleotide substrate, carried out in different assay volumes. Shaded cells indicate conditions in which the statistical analysis of the results is satisfactory. RFU: Relative fluorescence units; SD: Standard deviation; S/N: Signal to noise ratio; Z': Z'-factor. Assays carried out in triplicate then analysed.

Although the results obtained for the 100 µl assay volume reactions are better than for the reduced volume cases, the 25 µl assay volume conditions offer a good compromise between the reduction in reagent required, and the quality of data derived. Therefore subsequent AlkB FDH-coupled assays are carried out in a 25 µl assay volume.

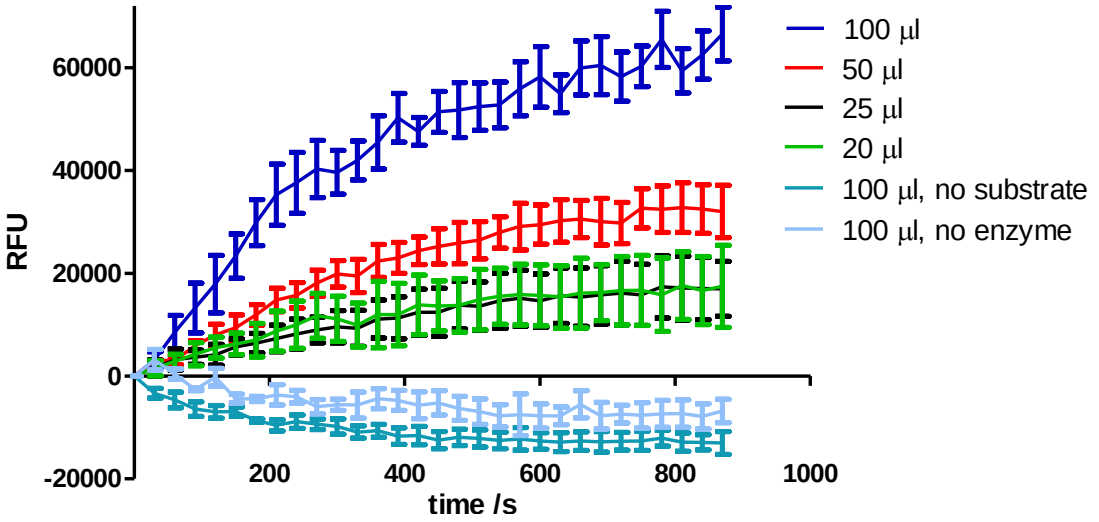


Figure 5-24: Miniaturisation test for the FDH-coupled reaction of ABH2 with 1meA oligonucleotide, assays carried out in assay volumes indicated. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean. Results given are recorded as relative fluorescence units (RFU) to allow more direct comparison between assays carried out in different volumes, as conversion to the amount of NADH produced requires different calibration curves for each assay volume.

A similar analysis was carried out for the reaction of ABH2 (Figure 5-24). Unfortunately the activity of ABH2 in this assay is significantly less than that of AlkB, as seen by comparing the lower maximal RFU signal for the 100 µl assays (~180000 for AlkB, but ~60000 for ABH2). Therefore the variation in the data obtained has a greater impact on the Z'-factor, as the separation band between the positive and negative values is smaller. This is reflected in the statistical analysis of the results (Table 5-7), which suggests that only the 100 µl reaction may be successfully used, but that it is not likely to be a particularly robust assay.

Initial rate	Positive control		No Substrate as negative control				No Enzyme as negative control			
(RFU/s)	Average	SD	Average	SD	S/N	Z'	Average	SD	S/N	Z'
100 µl	183.75	62.02	-41.99	3.18	3.64	0.13	-27.40	8.44	3.37	0.00
50 µl	65.14	25.24	-9.01	21.81	2.22	-0.90	-61.24	32.47	3.07	-0.37
25 µl	33.75	15.89	-6.54	3.68	2.47	-0.46	-29.77	15.60	2.85	-0.49
20 µl	39.41	19.61	10.95	12.55	1.22	-2.39	-35.05	15.66	2.97	-0.42

Table 5-7: Statistical analysis of FDH-coupled assay for ABH2 with 1meA oligonucleotide substrate, carried out in different assay volumes. Shaded cells indicate conditions in which the statistical analysis of the results is satisfactory. RFU: Relative fluorescence units; SD: standard deviation; S/N: Signal to noise ratio; Z': Z'-factor. Assays carried out in triplicate then analysed.

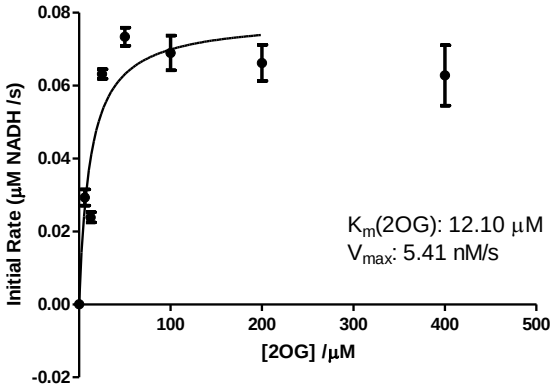


Figure 5-25: Michaelis–Menten plot for 2OG as a co-substrate of AlkB, using miniaturised and optimised assay conditions. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

To verify that the miniaturised assay for AlkB gave comparable kinetic parameters to those previously determined, an assay was carried out with AlkB varying the concentration of 2OG, to determine the K_m for 2OG under the altered conditions. The determined value of 12.10 μM (Figure 5-25) is very similar to the previous value of 11.36 μM, therefore the assay can be assumed to be similar under these optimised conditions.

5.6 Discussion

In this chapter the purification to high purity of full length and *N*-terminally truncated crystallisation constructs of recombinant AlkB and ABH2 proteins has been described. Initial characterisation revealed that the proteins are folded in solution, and are amenable to analysis by differential scanning fluorimetry, with optimal protein stabilisation in the presence of Mn(II).

Both proteins are DNA demethylases with a preference for the demethylation of 1meA lesions in oligonucleotides, and were observed to turn over the co-substrate 2OG in an indirect ^{14}C-labelled 2OG turnover assay. To directly monitor the reaction with the methylated substrate an FDH-coupled assay was developed, and the co-substrate dependence of the reactions determined. This revealed a significant difference in the ascorbate dependence of the proteins, with the activity of ABH2 appearing much more dependent on ascorbate than AlkB (AlkB K_m 1.3 μM vs ABH2 K_m 177 μM).

Furthermore, substrate inhibition was observed for AlkB at high concentrations with an 18mer oligonucleotide containing a 1meA lesion.

The FDH-coupled assay was subsequently optimised to allow more high-throughput assays to be conducted. This work therefore enables screening of small molecules for possible binding and inhibition studies of AlkB and ABH2 and subsequent crystallisation trials, studies that are described in the following chapter.

Chapter 6 - ABH Inhibitor Design

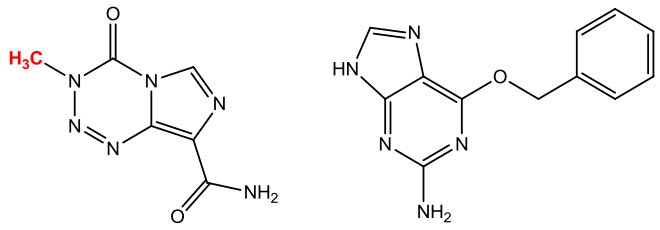
6.1 Introduction

Members of the 2OG oxygenase family are valuable therapeutic targets[365], due to their crucial roles in several cellular processes and reported links to conditions including cancer[366] and obesity[367]. Detailed inhibition studies of a number of 2OG oxygenases have been published, with much work focussing on the HIF hydroxylases[368-369], and the histone demethylases[360,370]. Tumours have low oxygen levels and require the initiation of adaption mechanisms such as angiogenesis via the hypoxic response for their survival. Inhibiting these processes in tumours (via inhibition of the HIF hydroxylases) could halt tumour growth[371]. There is also currently a large effort to develop selective chemical probes[372] for the *in vivo* study of oxygenases, particularly the histone demethylases[373], to dissect their individual roles in chromatin modification.

A major difficulty in identifying therapeutically beneficial inhibitors is extensive cross-reactivity due to poor selectivity. For example, humans have ~60 2OG oxygenases[374], therefore the use of simple 2OG mimics for inhibition of specific family members is unlikely to be successful, as such simple compounds are likely to act as generic inhibitors of all family members, leading to unwanted side-effects. Therefore, it is necessary to develop targeted inhibitors tailored to the individual structural features of the enzyme target, not only to deliver increased potency, but also to improve selectivity.

Many cancer therapies involve DNA damage, and alkylating agents are the oldest class of anticancer drugs in current use[375]. Temozolomide (Figure 6-1) is one of the most common alkylating agent pharmaceuticals, and forms mainly O^6meG and 7meG lesions[376]. These are repaired by O-6-methylguanine-DNA methyltransferase (MGMT)[377], but MGMT may be overexpressed in cancer cells decreasing the effectiveness of Temozolomide treatment[378]. Use of a "double therapy" combining methylating agents with an inhibitor blocking repair of the DNA lesions produced may increase the efficacy of the methylating agent. This has been attempted for Temozolomide with O-6-benzylguanine, an MGMT inhibitor[379] (Figure 6-1), and

although results in animal models were promising[380], human clinical trials did not show significantly increased therapeutic activity[381]. Use of a different class of methylating agent to produce lesions repaired by one of the ABH proteins, in combination with a specific ABH inhibitor, may have more success. AlkB is a tractable protein, so this model system can be used to study the inhibition of this class of DNA demethylases, with studies subsequently extended to include the ABH human homologues.



Temozolomide O-6-Benzylguanine

Figure 6-1: Structures of the methylating agent Temozolomide, and the MGMT inhibitor O-6-Benzylguanine.

Preliminary inhibition studies of AlkB have been published (Table 6-1), but many of the compounds tested to date are simple 2OG mimics known to be generic inhibitors of the 2OG oxygenase family, therefore are unlikely to be selective between 2OG oxygenase family members. The results were obtained using two different assay methods, either following 2OG turnover[325] (an indirect assay) or by capillary electrophoresis[339] (direct detection of demethylated product), which has likely led to discrepancies between results for some compounds tested by both methods. Quantitative results cannot be easily compared between the two assays due to the use of different methods, but a rank order of IC_{50} values may be compared. Although LNOA has the lowest IC_{50} of the compounds tested to date when the demethylation product is detected directly, its IC_{50} with the 2OG turnover assay is more than three times that of NOG, the second most potent compound by 2OG turnover. These differences are likely due to some compounds stimulating 2OG turnover that is not coupled to the demethylation of the prime substrate. In this situation the true activity of the enzyme will be overestimated, and thus inhibition underestimated. Therefore, direct detection of demethylation is much more reliable for use in inhibition and activity studies.

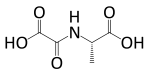
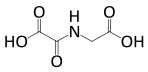
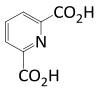
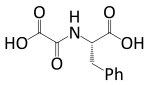
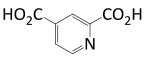
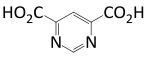
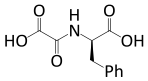
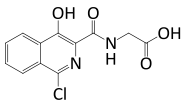
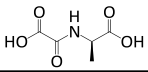
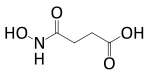
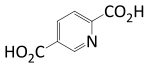
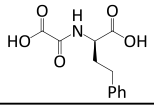
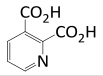
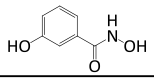
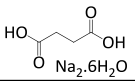
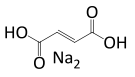
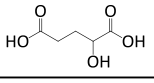
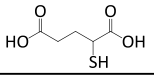
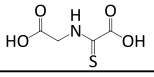
Structure	Compound name	IC_{50} /μM	
		2OG turnover	C.E.
	*N*-oxalyl-L-alanine (LNOA)	2390	35 ± 5
	*N*-oxalylglycine (NOG)	700	58 ± 7
	Pyridine-2,6-dicarboxylic acid (2,6-PDCA)		64 ± 8
	*N*-oxalyl-L-phenylalanine (LNOF)		75 ± 2
	Pyridine-2,4-dicarboxylic acid (2,4-PDCA)		84 ± 4
	Pyramidine-4,6-dicarboxylic acid		121 ± 5
	*N*-oxalyl-D-phenylalanine (DNOF)		126 ± 1
	FG2216		140 ± 18
	*N*-oxalyl-D-alanine (DNOA)	3330	149 ± 18
	4-hydroxyamino-4-oxobutanoic acid		191 ± 1
	Pyridine-2,5-dicarboxylic acid (2,5-PDCA)		193 ± 1
	*N*-oxalyl-D-homophenylalanine		260 ± 6
	Pyridine-2,3-dicarboxylic acid (2,3-PDCA)		479 ± 62
	LBE 332-6-3		>> 1000
	Succinate		No inhibition at 1000
	Fumarate		No inhibition at 1000
	(±)-2-hydroxyglutarate (2HG)	No inhibition at 4000	
	(±)-2-mercaptoglutarate	120	
	2-thiono(*N*-oxalylglycine)	810	

Table 6-1: Summary of previously reported inhibition data for compounds with AlkB: 2OG turnover[325] and capillary electrophoresis (C.E.)[339].

To date only one crystal structure of a small molecule (other than the co-substrate 2OG) bound to an ABH protein has been published, which is that of AlkB bound to succinate, the co-product of the catalytic reaction with 2OG. This shows that succinate binds in a monodentate manner at the metal centre[192]. The 2OG turnover inhibition study suggested that succinate is a poor inhibitor of AlkB (no inhibition at 1000 μM)[325], in contrast to observations for PHD2, which in a radiolabelled 2OG turnover assay has an IC_{50} for succinate of 19 μM[382]. Binding of succinate to FIH was additionally seen to occur in a monodentate manner, but again inhibition by succinate is not detected even at high succinate concentrations[382].

The aim of the work described here was the design and development of further generations of small molecule inhibitors of AlkB, which would hopefully have increased selectivity, specificity and potency compared to previously studied inhibitors. A dynamic-combinatorial mass spectrometry based screening method was used as a starting point for the development of potent small molecule inhibitors. Rationalisation for potency and protein stabilisation was possible through crystal structures of AlkB in complex with the three most potent compounds.

6.2 Inhibitor screening by FDH-coupled assay

6.2.1 Use of FDH assay for inhibition studies

The FDH-coupled assay has been successfully applied to identification of inhibitors of the JMJD2 histone demethylases, on both small[358] and large[359] scales, leading to the identification of compounds with high potency in cells[360].

Prior to testing of compounds (structures in Appendix C) in the fluorescence-based FDH-coupled assay precautions must be taken. Some small molecules have similar light absorption/emission characteristics to NADH, thus masking the NADH signal observed. An additional consideration is preincubation of compounds with the proteins investigated, as this can have a significant effect on the IC_{50} values obtained, both in the FDH assay[358] and for radiolabelled 2OG turnover[382]. Here, compounds tested were

preincubated with AlkB and FDH for 15 min at room temperature, prior to initiation of the reaction.

Furthermore, in a coupled assay a counter-screen of the compounds against the coupled enzyme (FDH) is necessary to verify that any decreased activity in the presence of small molecules is due to inhibition of AlkB, rather than FDH inhibition. A screen of selected 2OG mimics was carried out (Figure 6-2), but unfortunately under these conditions some compounds inhibited the activity of FDH. Compounds inhibiting FDH by >60 % (40 % residual activity, i.e. 2,4-PDCA, citrate, quercetin and ACG-180a), were determined to be inappropriate for use in the FDH-coupled assay. Compounds shown as 0 % activity (ACG-148a, ACG-263, FG-2216 and BB-287) have the same fluorescence characteristics as NADH, and gave maximal signal, so no FDH activity could be determined.

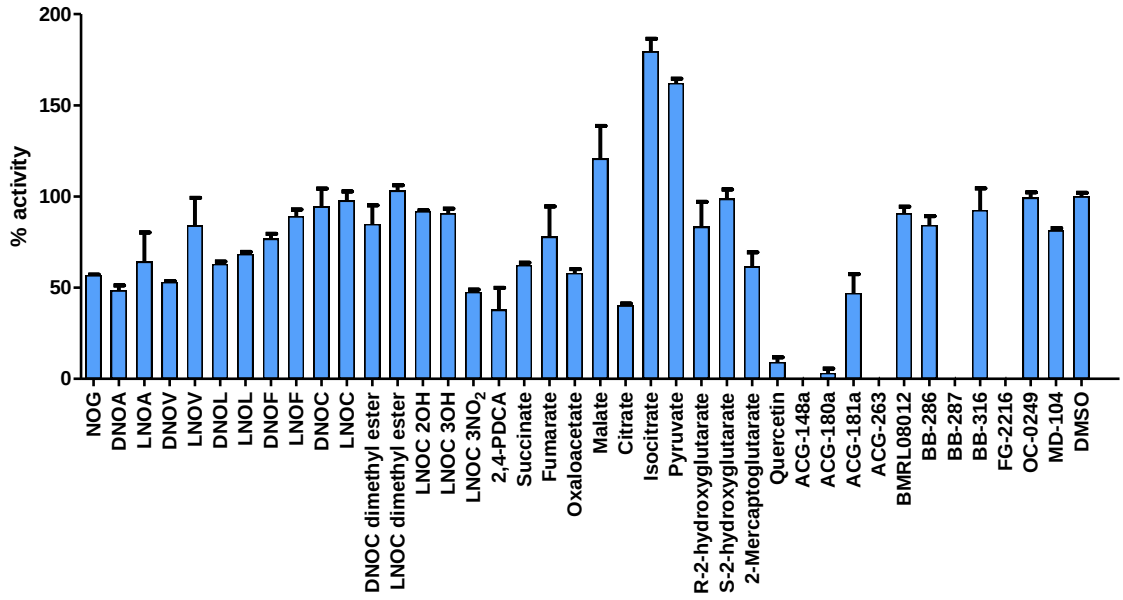


Figure 6-2: Initial reaction rate of FDH with formaldehyde after 15 min preincubation with 5 mM compounds. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean. Data are normalised to activity of FDH in the absence of small molecule (100 % activity), and activity in the absence of formaldehyde (0 % activity).

Since most small molecules are soluble in DMSO, the effect of DMSO on the activities of AlkB (Figure 6-3) and ABH2 (Figure 6-4) was investigated, both with and without the generic 2OG oxygenase inhibitor *N*-oxalylglycine (NOG). The presence of DMSO in fact stimulated the production of NADH for both AlkB and ABH2, with increased

initial rates observed for the demethylation reactions. To facilitate the use of poorly soluble small molecules, without greatly inhibiting FDH, a standard final concentration of 5 % DMSO (v/v) was used throughout.

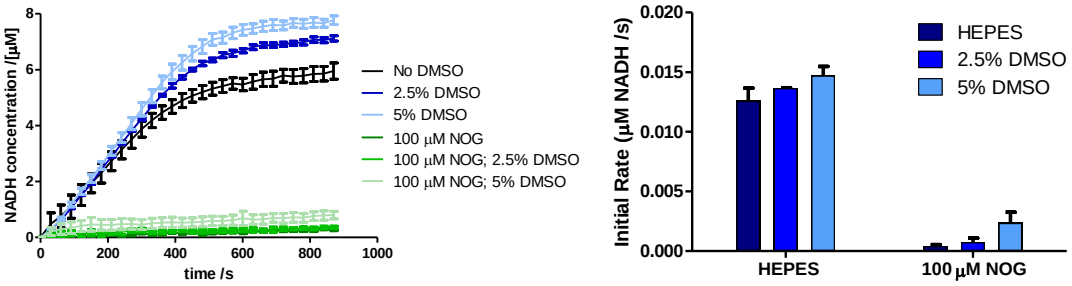


Figure 6-3: Comparison of the activity of ΔN11 AlkB in an FDH-coupled assay with an 18mer 1meA oligonucleotide substrate, at a range of DMSO concentrations (v/v). Data are represented as the reaction timecourse, and the initial rate. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

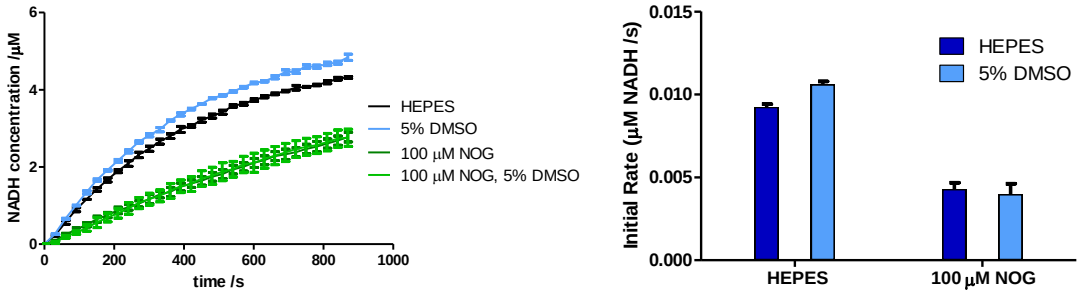


Figure 6-4: Comparison of the activity of ABH2 in an FDH-coupled assay with an 18mer 1meA oligonucleotide substrate, with or without 5 % DMSO (v/v). Data are represented as the reaction timecourse, and the initial rate. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

6.2.2 AlkB inhibition

To assess protein inhibition, the half maximal inhibitory concentration (IC_{50}) may be determined. The IC_{50} is a quantitative measure of inhibition, recording the concentration of compound at which the activity of an enzyme is reduced to half its maximal level. As the potential inhibitors tested are likely to be competitive with 2OG, inhibition assays are carried out maintaining 2OG at a concentration approximately equal to the K_m of 2OG, determined previously (Table 5-2). Figure 6-5 shows the dose-response curve for AlkB, which gives an IC_{50} value for NOG of 1.57 ± 0.33 µM. This value is low, suggesting that NOG is a good inhibitor of AlkB. In both of the previous inhibition studies NOG was the second most potent compound tested, although these assays were

carried out without preincubation, which can have a significant effect on the IC_{50} values obtained[358]. As NOG is a generic 2OG oxygenase inhibitor it is not appropriate for selective inhibition of AlkB, but its high potency means it may be used as the benchmark to which future inhibitors can be compared.

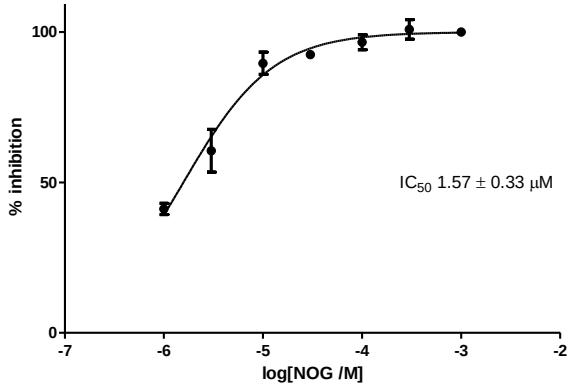


Figure 6-5: Inhibition of AlkB by *N*-oxalylglycine determined using an FDH-coupled assay with a 1meA oligonucleotide substrate (20 μM 2OG). Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean. IC_{50} value was calculated using log(inhibitor) compared with normalised response (variable slope), using GraphPad Prism.

6.2.3 ABH2 inhibition

The IC_{50} of NOG with ABH2 was determined (Figure 6-7), and found to be an order of magnitude greater than for AlkB (IC_{50} = 10.4 ± 2 μM), although given the slightly different assay conditions and K_m values for the reaction co-substrates this is not a significant difference. Within the 2OG oxygenase family varying NOG inhibition potencies are observed, for example the histone demethylase FBXL11 is only weakly inhibited by NOG (IC_{50} = 252 ± 50 μM), whereas NOG is a particularly potent inhibitor of PHD2 (IC_{50} = 0.8 ± 0.2 μM)[362].

Inhibition by further 2OG analogues was also investigated. Mutations in isocitrate dehydrogenases 1 and 2 (IDH1/2), which convert isocitrate to 2OG, have been found to have a gain-of-function effect allowing turnover of 2OG to *R*-2-hydroxyglutarate (*R*-2HG)[383]. These mutations are commonly identified in malignant glioblastoma and acute myeloid leukaemia[384], in which the accumulation of ~10 mM *R*-2HG may be observed in cells[383]. Elevation of either *R*- or *S*-2HG is also observed in 2-hydroxyglutaric aciduria, a rare neurometabolic disorder, although brain tumours are not observed in these patients, probably due to the severity of the primary symptoms[385].

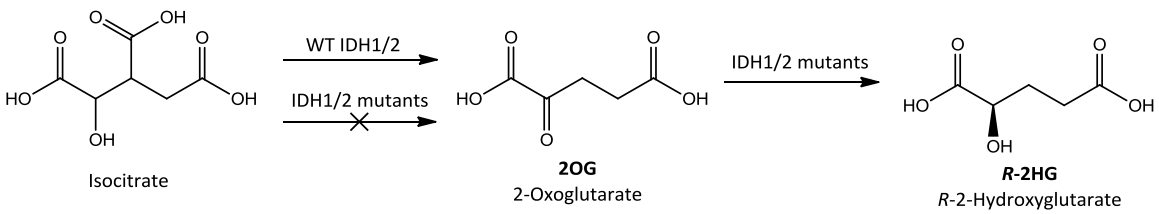


Figure 6-6: Reaction scheme showing the conversion of isocitrate to 2OG by wildtype isocitrate dehydrogenase (IDH), but not mutant IDH1/2, and the gain-of-function conversion of 2OG to *R*-2HG by IDH mutants.

Given the high concentrations of *R*-2HG accumulating in human cells, and the structural similarity of 2HG to 2OG, the inhibition of ABH2 by both *R*- and *S*-2HG was determined (Figure 6-7), and compared to the NOG IC_{50}. The IC_{50} for *S*-2HG (IC_{50} = 150 ± 20 μM) is an order of magnitude greater than that for NOG, with *R*-2HG (IC_{50} = 424 ± 77 μM) being the weakest inhibitor, suggesting that the stereochemistry of the hydroxyl group is important for inhibition.

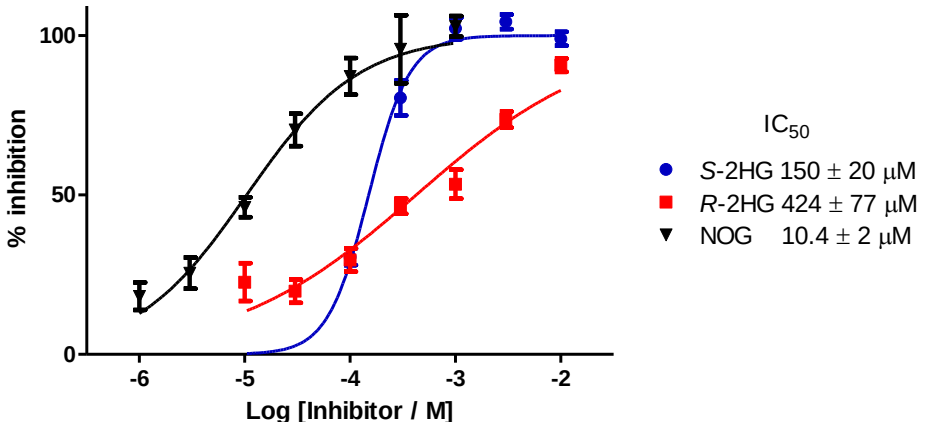


Figure 6-7: Inhibition of ABH2 by *R*-/*S*-2HG and *N*-oxalylglycine determined using an FDH-coupled assay with an 18mer 1meA oligonucleotide substrate (10 μM 2OG). Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean. IC_{50} values were calculated using log(inhibitor) compared with normalised response (variable slope), using GraphPad Prism.

To further understand the effect that accumulation of this oncometabolite may have in cells the inhibition of a number of other human 2OG oxygenases was determined by members of the CJS laboratory (Table 6-2)[362]. Overall, the *R*- form is generally less potent, with the *R*-2HG IC_{50} values for several enzymes being at least an order of magnitude greater than that for the *S*-enantiomer. However, inhibition of the JMJD2 family of N^{ε}-methyl lysine histone demethylases was found to be comparable for both enantiomers, and for JMJD2A both 2HG enantiomers show a similar potency to NOG.

PHD2 is very weakly inhibited *in vitro* by both 2HG enantiomers, which was confirmed in cells. Crystal structures of FIH and JMJD2A with 2HG show bidentate binding of both enantiomers at the active site metal, possibly competing with 2OG for binding, although kinetic studies of JMJD2A suggested that inhibition may be mixed-mode.

Enzyme IC_{50} ± SEM /μM	*R*-2HG	*S*-2HG	NOG
FIH	1500 ± 400	189 ± 34	46 ± 8
PHD2	7300 ± 3300	419 ± 150	0.8 ± 0.2
JMJD2A	24 ± 2	26 ± 3	17 ± 2
JMJD2C	79 ± 7	97 ± 24	14 ± 3
FBXL11	106 ± 22	48 ± 15	252 ± 50
ABH2	424 ±77	150 ± 20	10 ± 2
BBOX1	13200 ± 1100	142 ± 30	19 ± 2

Table 6-2: Inhibition of human 2OG oxygenases by NOG and *R*- and *S*-2HG. Data were obtained using the standard assays and substrates for each enzyme. Data taken from[362].

As a number of 2OG oxygenases are inhibited by 2HG this is likely to contribute to oncogenesis in the presence of elevated *R*-2HG, and possibly also some of the symptoms of 2-hydroxyglutaric aciduria. As inhibition by 2HG was found to be greatest for the JMJD histone demethylases it is likely that the major effect of 2HG accumulation will be on chromatin modification, possibly altering gene expression patterns. Inhibition of PHD2 is weak, however inhibition of FIH may affect HIF transcriptional targets. As moderate inhibition of ABH2 was observed the DNA damage response in cells accumulating high levels of *R*-2HG is also likely to be impaired, which may further contribute to oncogenesis.

6.3 2OG mimics

6.3.1 Inhibition by FDH-coupled assay

Initially, several compounds determined to be compatible with the FDH assay were screened at 100 μM for inhibition of AlkB (Figure 6-8). The compounds tested are 2OG mimics, including *N*-oxalyl amino acids, Krebs cycle intermediates, and a selection of synthetic 2OG analogues.

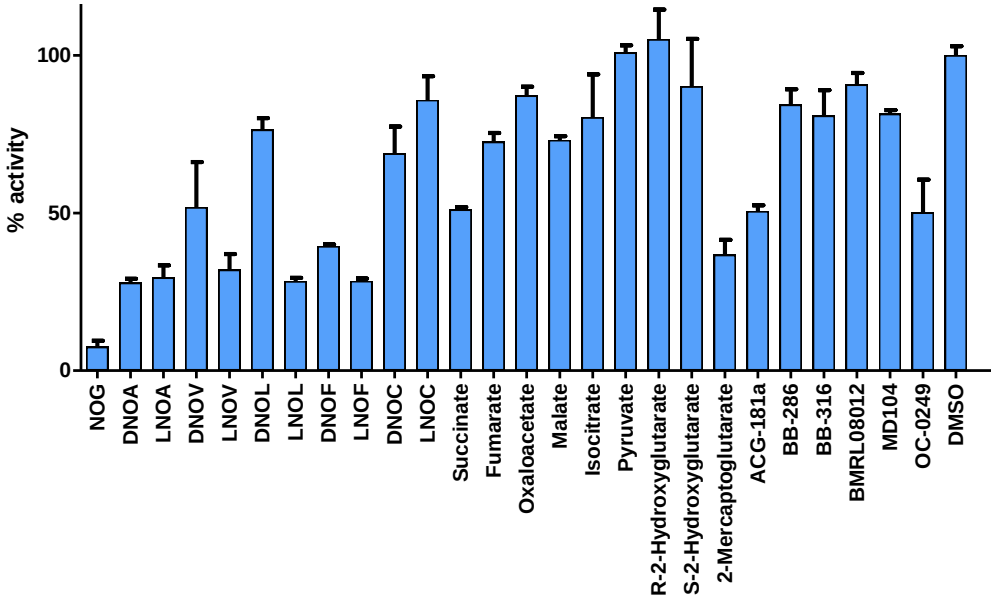


Figure 6-8: Initial reaction rate of AlkB with 1meA 18mer oligonucleotide after 15 min preincubation with 100 µM compounds, followed using the FDH-coupled assay. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean. Data are normalised to activity of FDH in the absence of small molecule (100 % activity), and activity in the absence of oligonucleotide (0 % activity).

In this assay, NOG appears to be the most potent inhibitor, with 100 µM NOG reducing the activity of AlkB to ~10 %. Several *N*-oxalyl amino acids, particularly those with L-stereochemistry, appear to inhibit AlkB, with DNOA, LNOA, LNOV, LNOL and LNOF causing a ~75 % drop in activity. *R*- and *S*-2HG show minimal inhibition of AlkB, as was observed previously for a 2HG racemic mixture[325] suggesting that, similarly to PHD2, some 2OG oxygenases are poorly inhibited by this oncometabolite[362]. In the previous 2OG turnover study, 2-mercaptoglutarate was found to be the most potent compound tested, with an IC_{50} value of 120 µM[325], and here this reduces the activity of AlkB to ~40 %. However, excluding the *N*-oxalyl amino acids (which will be discussed further later) none of the compounds tested appear to show significant potency with AlkB. OC0249 is a 4-carboxy-2-2'bipyridyl compound that has been found to inhibit the histone demethylase JMJD2E (IC_{50} = 270 nM)[386], and here appears to show some inhibition of AlkB, whereas ACG-181a, which also shows some AlkB inhibition, is based on the NOG scaffold.

6.3.2 Differential Scanning Fluorimetry

To screen more broadly for possible AlkB and ABH2 inhibitors, a large library of 2OG analogues was used. These compounds have been compiled from inhibition studies of a number of 2OG oxygenases from different species, which catalyse the full spectrum of reactions of this enzyme family.

Differential scanning fluorimetry (DSF) of AlkB and ABH2 was carried out previously (Section 5.3.3), with both proteins producing well defined melting curves. DSF is an effective method for rapidly identifying and validating selective enzyme inhibitors, and has been applied in screening of kinase inhibitors[387]. The binding assay (DSF) was chosen as a primary screen against AlkB over the turnover assays, due to its throughput and reduced requirements of reagents. Therefore T_m shifts produced on addition of small molecules were determined, as an initial quantifiable screening method for the study of protein binding and stabilisation. To enable detection above the limit of resolution a high concentration of small molecule was used, 100-fold that of the protein concentration.

A broad range of T_m shift values were obtained for the two proteins, with T_m shifts of ± 10 °C observed (Table 6-3). Some compounds, particularly with ABH2, cause significant destabilisation, as observed as the large negative T_m shifts. In general, the T_m shift profiles of AlkB and ABH2 appear quite different, although some compounds either stabilised (e.g. IS-52, OC-0345) or destabilised (e.g. 2,5-DHB, LBE-526) both AlkB and ABH2, suggesting that some compounds affect both 1meA DNA demethylases similarly, but that selectivity between the two enzymes is likely to be possible.

Assessment of the T_m shifts for the compounds previously investigated for AlkB inhibition does not appear to show significant differences, as the majority of the compounds have T_m shifts in the region of + 1 °C, although OC-0345 has a slightly higher T_m shift of 2.03 °C, and appears slightly more potent than other compounds tested. However, as these DSF and inhibition studies have not produced a broad range

of results for the compounds tested it is difficult to identify any significant correlation between the values obtained.

T_m shift /°C	AlkB	ABH2
DMSO	0.00	0.00
succinate	1.60	4.17
fumarate	2.18	4.72
oxaloacetate	0.60	-0.09
malate	1.18	2.08
citrate	-0.74	0.40
isocitrate	0.93	1.06
pyruvate	0.78	-0.38
R-2-HG	0.34	1.19
S-2-HG	1.37	2.30
2,3-PDCA	0.88	0.32
2,4-PDCA	5.49	6.15
2,5-PDCA	2.29	5.31
2,6-PDCA	1.66	0.23
3,4-PDCA	0.32	0.35
3,5-PDCA	1.25	0.77
5-Carboxy-8-hydroxy quinoline	2.22	n/a
Ciclopirox Olamine	5.81	0.85
Daminozide	1.79	n/a
2,4-DHB	0.59	3.03
2,5-DHB	-5.71	-3.74
3,4-DHB	-0.97	0.35
Gallate	n/a	n/a
HQNO	0.55	-6.10
Hydralazine	0.87	1.67
3-Hydroxy mandelate	0.01	-0.45
N-Hydroxy phthalimide	0.93	0.56
Malathion	0.45	-2.11
Malaoxon	0.75	-4.21
Minoxidil	0.45	-0.19
Prohexadione	0.74	-0.43
Quercetin	-2.71	1.90
SAHA	0.06	0.17
TSA	0.37	0.51
ACG-109a	-0.53	0.77
ACG-114c	0.77	1.43
ACG-137a	0.17	2.83
ACG-138a	-0.32	2.02
ACG-144a	0.64	1.94
ACG-148a	4.90	3.56
ACG-180a	1.19	2.39

T_m shift /°C	AlkB	ABH2
ACG-181a	1.55	-2.19
ACG-194b	2.83	0.34
ACG-223a	3.33	n/a
ACG-226a	0.58	2.29
ACG-227a	3.86	3.83
ACG-228a	0.16	1.92
ACG-237a	-0.80	1.95
ACG-253a	2.41	-3.01
ACG-255d	2.00	3.89
ACG-263	8.39	1.46
BB-103	6.46	1.35
BB-110H	1.51	2.56
BB-169	10.66	0.58
BB-204	0.10	0.48
BB-212	1.16	-1.64
BB-286	0.58	0.08
BB-287	1.27	2.28
BB-294	0.26	-0.21
BB-307	-0.87	-0.48
BB-308	0.22	n/a
BB-315	1.10	2.47
BB-316	1.18	-0.82
BB-318	0.67	n/a
BMRL08012	1.30	3.56
BMRL08014	2.35	1.13
BMRL08050	2.28	2.37
CO-A1-6d	1.21	-7.23
CO-A2-5a	0.63	-0.35
CO-A7-6h	5.88	-2.31
CO-B10-7a	1.29	-6.39
CO-C3-7f	4.52	-1.54
CO-C9-8d	-0.07	-0.06
CO-C10-10a	1.36	-10.19
CO-D4-8m	0.68	0.62
CO-D8-10k	2.42	-1.88
CO-D9-12	1.07	-1.01
EW-B17	0.36	-0.58
EW-T4	0.88	-1.02
EW-TNB	0.86	-1.66
EW-29	0.43	1.24

T_m shift /°C	AlkB	ABH2
EW-30	-0.08	1.11
EW-36	3.18	0.02
EW-37	n/a	-1.73
FG2216	7.96	1.34
IS-3	3.58	2.39
IS-21	0.90	n/a
IS-37	0.07	2.96
IS-38	-0.20	1.84
IS-44	0.43	-0.95
IS-45	1.35	-0.33
IS-52	6.73	4.91
IS-54	-0.25	4.43
IS-84	0.07	-0.53
JM-87	2.60	5.12
JM-102	2.00	0.56
JM-103	1.29	0.69
JM-104	1.70	4.26
JM-106	2.26	0.59
JM-109	1.69	1.25
JM-110	4.39	6.12
JM-124	2.62	1.13
LBE060	0.38	1.41
LBE096	1.49	-1.80
LBE-153	0.65	-1.03
LBE-314	1.17	0.01
LBE-329	0.72	-0.86
LBE-332-6-3	0.03	1.06
LBE-526	n/a	-13.38
MD-104	1.51	0.01
MD-228	2.25	1.07
MD-236	-0.05	0.48
OC-0206	1.12	0.95
OC-0248	1.57	2.00
OC-0249	2.03	1.50
OC-0336	1.07	0.69
OC-0340	0.94	-0.77
OC-0344	-0.07	-0.27
OC-0345	10.60	4.85
OC-0353	2.66	-2.20
OC-0358	1.84	-1.64

Table 6-3: T_m shifts obtained by DSF for ΔN11 AlkB and ΔN55 ABH2 with 2OG analogue library compounds. Reagent-mixes included 200 µM compound, 2 µM protein and 50 µM Mn(II), with 5 % DMSO (v/v).

These results are a useful survey of different scaffolds and functional groups that may be used as templates for derivatisation and optimisation of inhibitors. They also show that this technique produces a good range of T_m shift data, so it will be of use in screening newly developed compounds. Furthermore, the data obtained here will be used within the laboratory to enable comparisons to other 2OG oxygenases, which will also be screened against this library. The intention of these studies is to identify scaffolds that will show good selectivity for AlkB or other DNA modifying enzymes over other 2OG oxygenases, therefore an alternate approach was taken to identify novel inhibitors.

6.4 First generation inhibitors

6.4.1 Dynamic Combinatorial Mass Spectrometry

To attempt to develop novel small molecule inhibitors for AlkB a dynamic combinatorial mass spectrometry (DCMS) approach was employed by Dr Esther Woon. This methodology has been successfully applied in inhibitor design for JMJD2 histone demethylases[373], and carbonic anhydrase II[388], and enables rapid investigation of a range of functionalised compounds that may bind to the protein of interest, without the requirement for synthesis of many complex species.

Care must be taken when interpreting binding results obtained using mass spectrometry (MS), as the binding strength of small molecules in the MS "gas phase" can be predominantly due to van der Waals surface interactions, whereas solution binding is affected by solvation and hydrophobic interactions between the protein and small molecule[287,389-390]. Despite these potential concerns, many studies have found good correlation between MS binding and inhibition in solution[373,389].

Initially, the binding of *N*-oxalyl amino acids to AlkB was screened using non-denaturing electrospray ionisation (ESI) mass spectrometry. Binding of small molecules to a protein is identified by appearance of an additional peak in the mass spectrum at an increased mass, corresponding to the mass of the protein added to that of the small

molecule. Strong binding of the compound to the protein means that the peak in the mass spectrum is almost entirely shifted to the increased mass. To obtain appreciable and defined protein peaks in the mass spectra a 5-fold excess of Fe(II) was added to protein samples prior to mixing with small molecules.

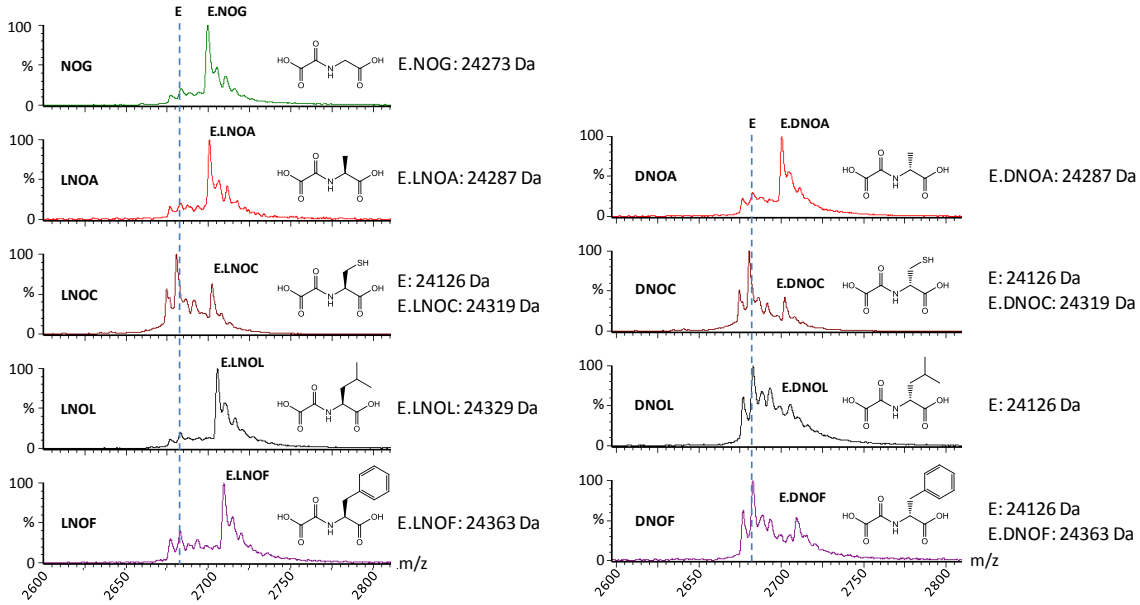


Figure 6-9: Binding of *N*-oxalyl amino acids (15 µM) to AlkB (15 µM, with 75 µM Fe(II)), analysed by non-denaturing ESI mass spectrometry. "E" and vertical dashed line highlights enzyme-only peaks, "E.NOX" labels highlight peaks corresponding to enzyme/compound complexes.

Strong binding of NOG, LNOA, DNOA, LNOL and LNOF to AlkB was observed by ESI-MS (Figure 6-9). It appears that when the C_α substituents are small there is sufficient space in the AlkB active site to allow binding of compounds with the either the L- or D-stereochemistry. However when the side chains are larger, there is only room to bind L-enantiomers. Generally, the binding to AlkB of *N*-oxalyl amino acids with L-stereochemistry appears to be preferred, as was implied by earlier inhibition studies (Figure 6-8). This stereochemical preference is interesting as FIH[368] and members of the JMJD2 histone demethylase subfamily show a preference for binding of, and inhibition by, compounds of the opposite D- stereochemistry[373]. Therefore L-substituted *N*-oxalyl amino acids are likely to be a good starting point for development of potent and selective AlkB inhibitors.

Both LNOC and DNOC bind weakly to AlkB by MS (Figure 6-9); however the NOC compounds are very useful in dynamic combinatorial chemistry (DCC) methods due to

the presence of the thiol side-chain, with which the facile formation of disulphide bonds is possible. As the L- stereochemistry is preferred for binding to AlkB, LNOC can be used as a "support ligand". The support ligand can bind in the active site of AlkB, most likely in the same orientation as the co-substrate 2OG, leaving the thiol group free to react with other thiols in solution. This may be exploited in a dynamic combinatorial reaction, by addition of a dynamic combinatorial library (DCL) comprising many structurally diverse thiols. These thiols may reversibly react with the support ligand, forming an equilibrium mixture of disulphides, driven using the active site as a template. Any disulphide compounds that can fit tightly into the active site of the protein formed preferentially. This process is enthalpically driven, and will lead to a shift of the position of equilibrium towards the formation of tight binders. A cartoon depiction of this general dynamic combinatorial process is presented in Figure 6-10.

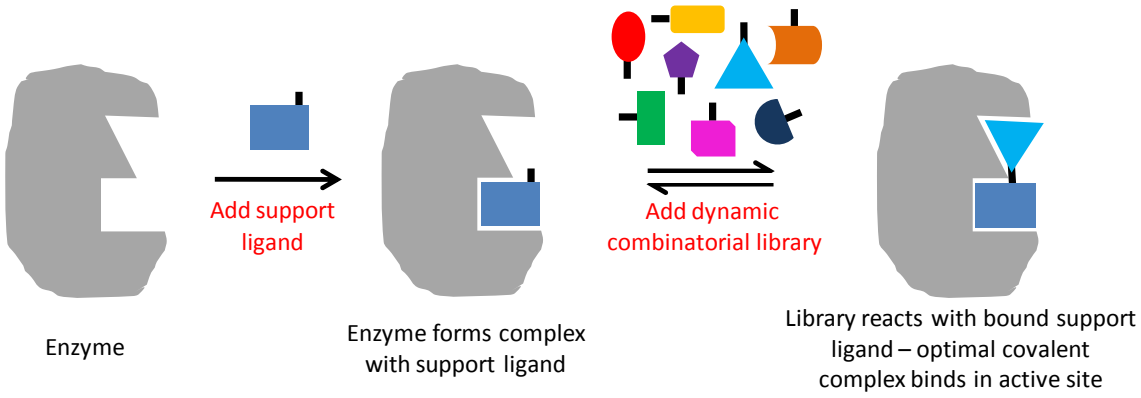


Figure 6-10: Enzyme template driven dynamic combinatorial chemistry (DCC). Support ligand binds to the enzyme, then the dynamic combinatorial library (which has reactive functionality with support ligand) is added. Formation of a preferential disulphide species is driven by templating by the enzyme active site.

The LNOC-AlkB complex was incubated with a thiol library and followed over time by MS. After 4 h, preferential formation of two complexes was observed, with the binding species identified as 2- and 3-hydroxythiophenol, and 3-nitrothiophenol. To allow further investigation of these compounds, which on binding to LNOC appear to fit well in the active site of AlkB, their stable carbon analogues were synthesised by Dr Esther Woon, (referred to as LNOC 2OH, LNOC 3OH and LNOC $3NO_2$, see Figure 6-11), which were observed to bind strongly to AlkB (Figure 6-11), with competition

experiments ranking the binding affinities in the order LNOC $3NO_2$ > LNOC 3OH > LNOC 2OH.

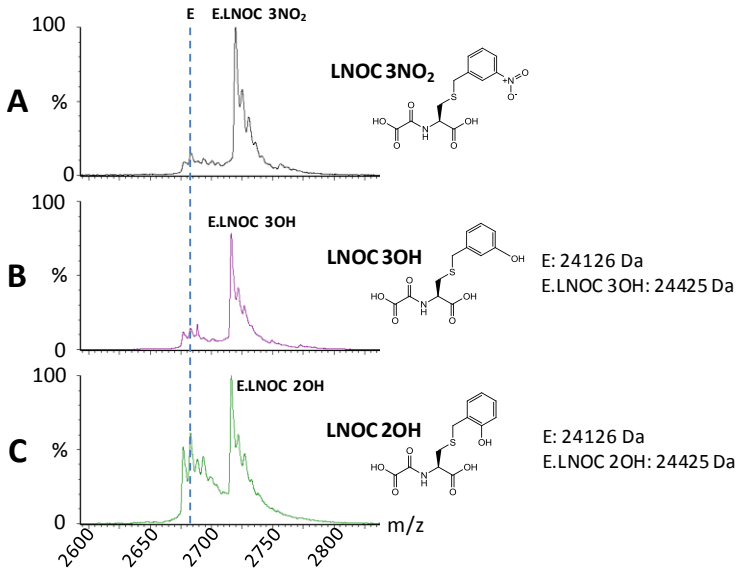


Figure 6-11: Non- denaturing mass spectrometry showing binding of (A) LNOC $3NO_2$, (B) LNOC 3OH and (C) LNOC 2OH carbon analogue compounds (all 15 µM) to AlkB (15 µM, with 75 µM Fe(II)). "E" and vertical dashed line highlights enzyme-only peaks, "E.LNOC" labels highlight peaks corresponding to enzyme/compound complexes.

6.4.2 Differential Scanning Fluorimetry

6.4.2.1 Disulphides

To verify results obtained by DCMS for species potentially binding to AlkB, complementary DSF analyses were carried out. Initially, an attempt was made to determine whether disulfides could be formed in a similar manner to the DCMS methodology, and T_m shifts obtained. AlkB was incubated either with or without the support ligands LNOC or DNOC, then a reduced selection of individual thiols were added, and T_m shifts obtained.

None of the individual thiols incubated with AlkB in the absence of the support ligands stabilise AlkB (data not shown), confirming that binding of the added thiols in DCMS is likely to occur via binding to the free thiol of the support ligand, rather than to a thiol in the AlkB active site. However, T_m shifts for the thiols added in the presence of the support ligand do not appear to correlate with those obtained by DCMS, as the 2-/3-hydroxythiophenol and 3-nitrothiophenol coupling partners observed to bind to AlkB/LNOC by MS here produce significant destabilisation (T_m shift: -10 – -20 °C).

This is likely due to the continual increase in temperature through the course of the DSF experiment, during which the disulphides are not stable. Although NOC appears to bind at the active site (destabilisation relative to the no-compound control is here similar to that observed later, Table 6-4) the added partner thiol may initially bind, but as the temperature increases, the lower bond strength of the S-S bond is unlikely to be able to withstand heating. The presence of a high concentration of thiols in solution (total 400 μM, compared to 2 μM protein) means that attack by other thiols, particularly at higher temperatures, is likely. This is less likely to be observed in MS, where the protein peaks are under analysis, and the greater sensitivity of the MS technique means binding may be observed when the thiols are present at a 1:1 ratio with AlkB. This therefore shows that MS is a valuable and important tool for the initial rapid investigation of possible functionalisation of the scaffold compound, whereas DSF is not appropriate for these studies.

6.4.2.2 *N*-Oxalyl amino acids and derivatives

DSF was subsequently used to screen the *N*-oxalyl amino acid compounds previously studied by MS, against both AlkB and ABH2 (Table 6-4). T_m shift data were also obtained following preincubation with NOG, as this may stabilise the protein if the *N*-oxalyl compounds bind at a site other than the metal centre.

It appears that compounds which bind well to AlkB by MS (Figure 6-9) also give good thermal stabilisation, with greater T_m shifts generally observed for L- compounds than for their D- counterparts, except for NOY, which also did not appear to bind well to AlkB by MS. The DNOY scaffold was used for development of specific JMJD inhibitors, suggesting a degree of selectivity between these 2OG oxygenase subfamilies[373]. ABH2 does not appear to show a significant preference for either stereochemistry, although in most cases slightly greater stabilisation appears to be observed with the L- isomers. Interestingly, the optimal T_m shifts, and therefore greatest stabilisation, are observed for ABH2 with DNOC (T_m shift 4.79 °C) and LNOC (T_m shift 4.00 °C). These compounds destabilise AlkB and, together with the relatively low

T_m shift for ABH2 with NOG (0.46 °C/1.77 °C with NOG preincubation), compared to other *N*-oxalyl amino acids, and the greater IC_{50} value for ABH2 compared to AlkB (10.4 μM vs. 1.57 μM) again suggests that the two proteins have relatively different binding characteristics at the metal binding site. In most cases a relative destabilisation is observed for T_m shifts obtained following preincubation with NOG, possibly as NOG preferentially binds to the protein's metal centre, restricting the binding of the compounds added later.

T_m shift /°C	ΔN11 AlkB	ΔN11 AlkB + NOG	ΔN55 ABH2	ΔN55 ABH2 + NOG
NOG	2.89	2.18	0.46	1.77
DNOA	2.92	1.55	2.83	0.99
LNOA	3.82	2.08	2.19	0.23
DNOV	0.52	0.26	0.39	-0.09
LNOV	1.62	0.81	0.81	-0.42
DNOL	0.57	0.43	0.37	-0.59
LNOL	4.95	3.16	1.04	-2.90
DNOF	0.96	0.47	1.36	0.19
LNOF	4.27	2.40	1.87	-0.77
DNOY	0.81	0.62	1.36	1.15
LNOY	0.62	0.41	-0.11	-0.80
DNOC	-1.20	0.12	4.79	2.13
LNOC	-1.24	-0.72	4.00	1.01
LNOC 2OH	1.28	0.78	-1.20	-1.79
LNOC 3OH	1.79	1.29	0.16	-2.05
LNOC $3NO_2$	3.97	3.41	2.20	1.38

Table 6-4: T_m shifts obtained by DSF for addition of 200 μM compounds to mixtures containing 2 μM protein and 50 μM Mn(II), with 5 % DMSO (v/v). 20 μM NOG was added where indicated.

Good stabilisation of AlkB was observed for the carbon analogues of the LNOC derivatives identified by DCMS, with the T_m shifts correlating with the MS binding preference. The optimal compound with AlkB, LNOC $3NO_2$, also stabilised ABH2, although to a lesser extent.

6.4.3 Inhibition

Having identified compounds that appear to bind well to AlkB, and which thermally stabilise the protein structure, it was important to investigate possible AlkB inhibition. Initially, the compounds were screened in the FDH-coupled assay, at a concentration of 100 μM, to determine whether any inhibition is observed. Results were compared to

activity in the presence of either the NOC support ligands, or NOG, a benchmark for good inhibition.

Inhibition of AlkB by each of the LNOC derivatives was observed (Figure 6-12), although to a lesser extent than that by NOG. The relative order of inhibition by the three LNOC derivatives is slightly different from that observed by DSF and MS, as LNOC 3OH appears to be more potent than LNOC $3NO_2$. However this single-concentration inhibition test is less accurate than use of a range of inhibitor concentrations; in order to more accurately compare the compounds the IC_{50} values should be obtained.

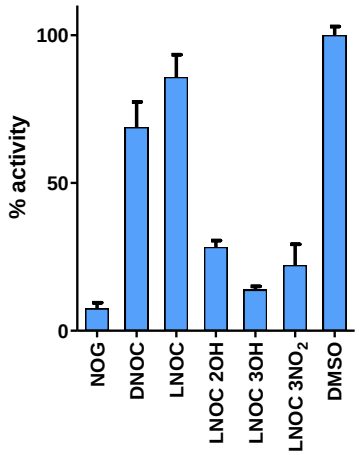


Figure 6-12: Test of inhibition of the reaction of ΔN11 AlkB with an 18mer 1meA oligonucleotide substrate by 100 μM compounds (15 min preincubation) in an FDH-coupled assay, with data presented as residual activity. Data points represent the mean of triplicate experiments, with error bars showing the standard error of the mean.

One of the previous AlkB inhibition studies previously published by this laboratory was carried out in collaboration with the Krylov group at York University, Toronto, Canada. The detection method for this assay is capillary electrophoresis, which allows sensitive detection of very small reaction quantities. The methylated DNA substrate (sequence: $CC(3meC)C_{12}$) is labelled with fluorescein at the 3' end, allowing detection by laser induced fluorescence. The demethylated DNA product and unreacted starting material are separated using an appropriate capillary electrophoresis method, and the spectral peaks integrated to determine the extent of demethylation.

An advantage of this assay over the FDH-coupled assay is that reduced IC_{50} values may be obtained. The FDH-coupled assay requires the use of 2 μM enzyme to achieve good

signal above the assay background, thus placing a relatively high limit on the IC_{50} value that may be obtained. In contrast the enzyme concentration used for the capillary electrophoresis method is much lower at 5 nM, enabling far lower IC_{50} values to be obtained. As the intention of these studies is the development of compounds more potent than NOG (IC_{50} = 1.57 μM in the FDH-coupled assay), the ability to confidently determine lower IC_{50} values would be beneficial.

Additionally the possible effect of inhibition of FDH should also be considered, as 50 % FDH inhibition is observed in the presence of 5 mM LNOC $3NO_2$ (Figure 6-2). It is likely that further derivatives of LNOC $3NO_2$ developed in subsequent iterations of the optimisation process may also inhibit FDH, precluding their analysis with AlkB in this FDH-coupled assay. As direct comparison of inhibition between different assays can be difficult it was thus decided to use the capillary electrophoresis assay for IC_{50} determination for all of the compounds investigated.

Capillary electrophoresis inhibition assays were carried out by Dr Lana Krylova (York University, Toronto). In this assay no inhibition of AlkB by either NOC support ligand, was observed (IC_{50} > 1 mM in each case, Table 6-5), supporting results from MS and DSF binding studies. Inhibition of AlkB also correlated with MS and DSF results, with LNOC $3NO_2$, the previously identified optimal compound having a low micromolar IC_{50} value (5.22 μM). This IC_{50} value is also lower than that obtained for NOG in this assay (32 μM), showing that the DCMS methodology has been able to identify a potent AlkB inhibitor, which would hopefully be more selective than NOG.

Compound	IC_{50} /μM
NOG	32.0
DNOC	> 1000
LNOC	> 1000
LNOC 2OH	50.4
LNOC 3OH	48.0
LNOC $3NO_2$	5.22

Table 6-5: Inhibition of AlkB by first generation compounds. Assays were carried out using a capillary electrophoresis method with a 15mer 3meC-containing oligonucleotide.

6.4.4 LNOC 3NO$_2$ crystal structure

To date, crystal structures of AlkB have been published with AlkB in complex with the co-product succinate[192], and DNA[194], but not with inhibitory small molecules. Therefore it was unknown how any inhibitor compounds may bind at the active site. As LNOC 3NO$_2$ inhibits AlkB, and gives a good positive T_m shift, increasing the likelihood of obtaining a crystal structure[298], crystallisation attempts were initiated, using the pET-24a(+) ΔN11 AlkB construct, as used in previous studies[192,196]. AlkB crystallisation was carried out by Jerome Ma and Dr Michael McDonough, and the structure was solved to 1.6 Å by Dr Michael McDonough using the molecular replacement technique.

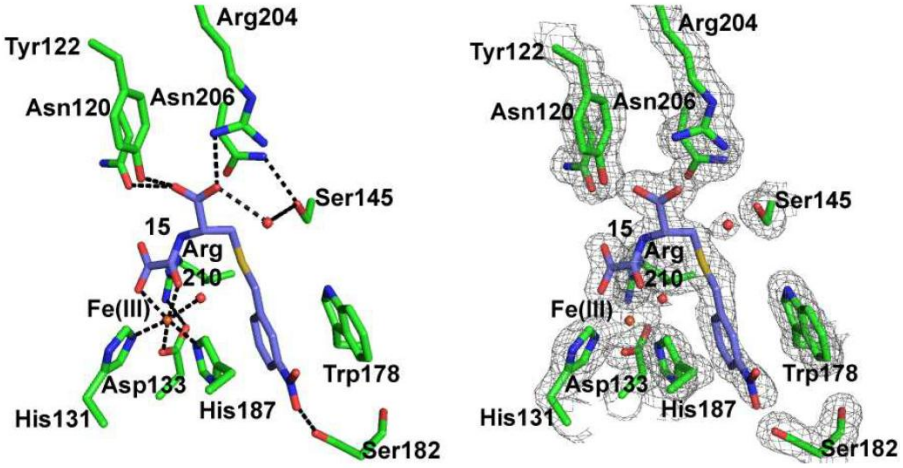


Figure 6-13: Active site view of crystal structure of ΔN11 AlkB bound to LNOC 3NO$_2$ (purple sticks). Left: grey mesh represents experimental 2F$_o$-F$_c$ electron density map contoured to 1.0σ. PDB ID: 3T4H. Figure prepared by Wei Shen Aik.

The obtained crystal structure (Figure 6-13, PDB ID: 3T4H) shows that the AlkB metal centre is coordinated by residues H131, D133, and H187, and a water molecule, as previously observed[192]. The oxalyl group of LNOC 3NO$_2$ occupies the remaining metal coordination sites, binding in a bidentate manner. The further two carboxylic acid oxygen atoms LNOC 3NO$_2$ are positioned to form hydrogen bonds to the sidechains of Y122 and N120, or R204 and S145. The phenyl group of LNOC 3NO$_2$ is present in a relatively hydrophobic pocket within the active site. Nonetheless, the 3-nitro group is able to make an additional hydrogen bond to S182, an interaction which would be

weakened by the presence of a 3-hydroxyl group (as in LNOC 3OH), and is unlikely to be possible if the hydroxyl is present at the 2-position (as for LNOC 2OH), rationalising the increased IC_{50} values for these related compounds. Comparison to a structure of AlkB in complex with 2OG (PDB ID: 313Q[200]) shows that the majority of the active site residues are present in very similar orientations in both cases, excluding the side chain of W178. In the structure presented here the orientation of the indole group is slightly rotated, likely enhancing a π-stacking interaction between the indole rings, the phenyl ring of LNOC $3NO_2$, and the imidazole group of H187.

Complementary crystallisation of the ΔN55 ABH2 protein was attempted by Dr Rashed Chowdhury, but unfortunately no crystals appropriate for crystallography were obtained. This may be due to the presence of the positively charged *N*-terminal His-tag, which likely alters the overall surface properties of the protein, interfering with successful crystallisation.

6.5 Second generation inhibitors

Following successful identification of compounds binding and inhibiting AlkB, two approaches were taken to try to identify more potent and hopefully more selective compounds, based on the crystal structure of AlkB in complex with LNOC $3NO_2$.

6.5.1 Pfizer compound docking studies

An initial attempt to identify further AlkB inhibitors was carried out in collaboration with scientists at Pfizer (Sandwich, UK). One way in which selectivity could be achieved for the AlkB class of 2OG oxygenases would be targeting of the nucleic acid substrate binding site.

The crystal structure of AlkB in complex with Fe(II) and LNOC $3NO_2$ was used for docking studies (carried out by Samantha Hughes, Pfizer) of compounds from Pfizer fragment libraries. 1meA and 3meT nucleosides were used as probes to identify nucleoside mimics. These docking analyses identified 93 compounds (referred to as PF1-93). For an initial investigation of these compounds, T_m shifts produced on addition of compounds were determined. As a 2OG mimic was present in the structure into

which the compounds were docked, and the PF compounds are likely to bind in the substrate binding cleft, T_m shifts were determined with and without NOG. NOG should bind to AlkB with the same binding mode as 2OG, stabilising the structure in the correct conformation to allow possible binding of the docked compounds. T_m shifts were also determined for ABH2 to assess potential selectivity.

T_m shift /°C	AlkB	AlkB + NOG	ABH2	ABH2 + NOG
Reference	0.00	0.00	0.00	0.00
PF1	-0.58	-0.70	-1.06	-1.64
PF2	-0.97	-0.65	-0.97	-1.61
PF3	0.14	-0.47	-0.36	-1.29
PF4	-7.53	-0.54	-5.97	-2.25
PF5	-6.63	-1.79	-8.65	-2.96
PF6	-2.97	-1.50	-3.95	-2.37
PF7	-0.53	-0.61	-1.53	-1.86
PF8	-1.13	-0.68	-0.46	-0.15
PF9	-1.25	-0.68	-2.06	-1.10
PF10	-0.92	-0.66	-1.62	-1.15
PF11	-7.28	-0.57	-10.30	-2.12
PF12	-0.56	-0.47	-0.45	-0.53
PF13	-1.31	-0.80	-1.57	-2.33
PF14	-0.68	-0.46	-0.07	0.34
PF15	-1.29	-0.83	-2.44	-1.85
PF16	-2.00	-0.97	-3.96	-2.61
PF17	-2.04	-1.07	-3.93	-2.62
PF18	-1.22	-0.91	0.17	-1.49
PF19	-6.25	-1.88	-6.54	-3.25
PF20	-1.40	-0.58	-2.64	-1.99
PF21	0.11	-0.40	-2.66	-0.74
PF22	0.30	-0.67	-2.39	-1.02
PF23	-0.02	-0.63	-1.99	-1.09
PF24	0.02	-0.37	-1.74	-0.34
PF25	0.22	-0.64	-2.59	-1.52
PF26	-7.17	-1.51	-8.93	0.24
PF27	-0.17	-0.89	-4.03	-1.77
PF28	-2.94	-1.47	-7.79	-2.54
PF29	-8.40	-0.94	-9.68	-2.47
PF30	-0.08	-0.27	-2.94	-0.42
PF31	-0.16	-0.54	-4.07	0.76
PF32	-0.22	-1.05	-2.26	-0.74
PF33	-0.22	-0.06	-3.95	-1.48
PF34	-1.65	-0.89	-2.87	-2.69
PF35	-5.60	0.30	-10.36	-0.84
PF36	-0.38	0.04	-2.27	-1.19
PF37	-0.52	0.12	-2.37	-0.45
PF38	-0.08	0.03	-2.32	-1.73
PF39	-1.18	-0.37	-4.52	-3.12
PF40	-1.67	-0.34	-6.14	-0.08
PF41	-0.80	-0.55	-4.49	-2.67
PF42	-1.12	-0.24	-3.11	-1.76
PF43	-3.69	-0.89	-6.00	-2.81
PF44	-1.80	-0.76	-3.59	-2.62
PF45	-7.97	-0.27	-2.37	-1.86
PF46	-0.34	-0.01	-2.67	-0.97

T_m shift /°C	AlkB	AlkB + NOG	ABH2	ABH2 + NOG
PF47	-0.11	-0.08	-3.78	-1.62
PF48	-0.64	-0.04	-1.85	-0.34
PF50	-0.70	-0.49	-3.93	-1.94
PF51	-0.04	0.30	-1.27	-1.01
PF52	-0.41	0.07	-3.53	-4.13
PF53	-0.98	-1.24	-4.15	-2.01
PF54	-1.16	-0.50	-4.57	-2.22
PF55	-0.47	-0.71	-4.19	-1.31
PF56	-0.26	-0.17	-4.61	-1.21
PF57	-0.83	-0.53	-4.41	-0.81
PF58	-5.67	-1.10	-4.65	-2.70
PF59	-0.89	-0.19	-3.23	-1.15
PF60	-4.05	-1.38	-4.96	-2.60
PF61	-1.13	-0.41	-2.12	-1.10
PF62	-1.22	-0.50	-3.14	-1.79
PF63	-0.97	-0.37	-5.09	-0.78
PF64	-0.09	-0.48	-3.53	-2.17
PF65	-0.32	-0.25	-3.51	-1.43
PF66	-6.93	-0.33	-8.16	-1.79
PF67	0.05	0.22	-4.44	0.39
PF68	-0.11	0.07	-3.03	-1.20
PF69	-0.91	-0.17	-4.01	-1.51
PF70	-0.63	-0.36	-3.72	-1.66
PF71	-1.79	-0.76	-0.05	-0.84
PF72	-0.67	-0.31	-2.94	-1.74
PF73	-0.34	-0.37	-2.21	-1.37
PF74	-0.79	-0.82	-2.97	-1.08
PF75	-0.30	0.05	-1.83	-1.25
PF76	0.04	0.45	-0.95	-0.91
PF77	-0.14	0.18	-0.77	-0.58
PF78	-0.33	-0.10	-1.24	-1.01
PF79	0.31	-0.13	-0.91	-0.69
PF80	-0.02	-0.37	-0.81	-0.86
PF81	-0.09	0.07	-1.38	-0.85
PF82	-0.21	0.02	-1.24	-0.88
PF83	-0.38	0.30	-0.75	-0.96
PF84	-0.02	0.18	-1.16	-0.35
PF85	-0.32	-0.03	-1.94	-0.95
PF86	-0.43	-0.11	-1.03	-1.14
PF87	-0.20	-0.02	-1.27	-1.19
PF88	-0.49	-0.09	-1.35	-0.88
PF89	-0.47	-0.16	-1.49	-0.89
PF90	-1.38	0.19	N/A	N/A
PF91	-0.59	-0.24	-1.19	-0.93
PF92	-0.50	0.03	-1.85	-1.17
PF93	0.16	0.47	-1.54	-0.57

Table 6-6: T_m shifts obtained by DSF for addition of 200 µM compounds to mixtures containing 2 µM protein and 50 µM Mn(II), with 5 % DMSO (v/v). 20 µM NOG was added where indicated.

Unfortunately no significant stabilisations were observed for either AlkB or ABH2 on addition of the PF compounds (Table 6-6), with only one compound giving a T_m shift > 0.5 °C (ABH2 + PF31: 0.76 °C in the presence of NOG). Many of the compounds tested in fact cause significant destabilisation of the proteins, with destabilisations > 7 °C not uncommon. In the presence of NOG these destabilisations are generally significantly decreased, in some cases leading to minor protein stabilisation. However, this is likely due to stabilisation caused by NOG binding at the protein metal centre rather than a specific stabilisation by the PF compounds. As the compounds are intended to bind at discrete sites it would be expected that addition of the compounds would have an additive effect, therefore any T_m shifts observed on binding would be large. However, on comparison to the magnitude of the T_m shifts observed with the LNOC derivatives (in the region of +3/4 °C) the T_m shift values obtained here were not taken to be significant, therefore these compounds were not pursued further.

6.5.2 Structure based synthesis

In a complementary approach, further optimisation of the compounds identified by DCMS was carried out through synthesis of analogues, based on the initially identified LNOC scaffold. By modelling of structures in the active site of AlkB, compounds were designed and synthesised to hopefully allow exploitation of the deep, predominantly hydrophobic, nucleotide binding pocket, while retaining compound binding at the metal centre (Table 6-7). To exploit the hydrophobic pocket in which the phenyl ring of LNOC $3NO_2$ was observed to bind, this substituted phenyl was substituted by a bulky naphthalene side-chain (1-napthalene, MD314; 2-napthalene, MD316). Additionally, it appears that there is space around the oxalyl group for structural elaboration. Therefore two compound series were synthesised, replacing the oxalyl group with substituted pyridyl (e.g. MD244) or quinolene (e.g. MD310) rings. Synthesis of these compounds was carried out by Marina Demetriades (MD compounds) and Dr Esther Woon (EW compounds).

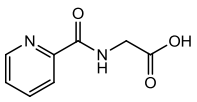
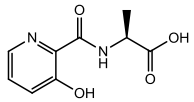
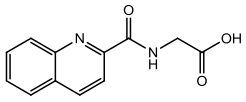
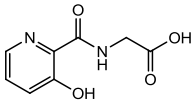
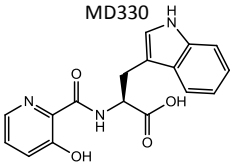
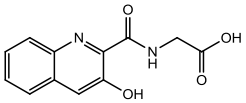
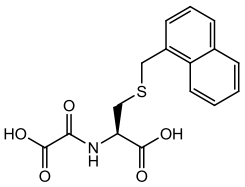
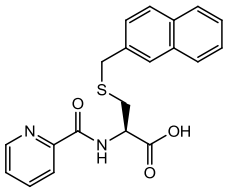
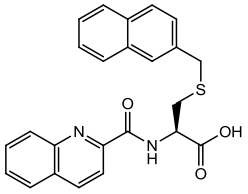
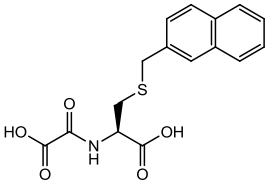
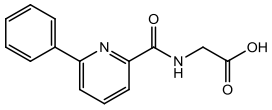
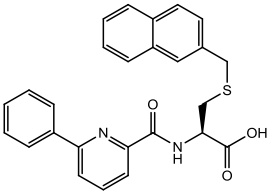
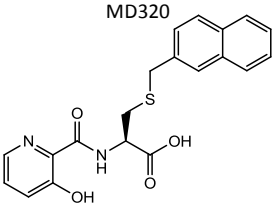
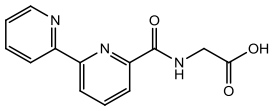
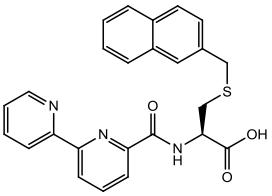
MD244 	MD328 	MD310 
MD246 	MD330 	EW38 
MD314 	MD318 	MD326 
MD316 	MD312 	MD324 
MD320 	MD308 	MD322 

Table 6-7: Structures of synthesised AlkB second generation inhibitors.

6.5.3 DSF and inhibition

The synthesised compounds were analysed by DSF with AlkB and ABH2. A selection of these compounds was tested for inhibition of FDH; unfortunately a number of the compounds showed significant FDH inhibition (>70 % at 5 mM compound) therefore the inhibition of AlkB was investigated using the capillary electrophoresis assay (Table 6-8).

All of the compounds investigated (excluding MD310) stabilise AlkB, with several compounds conferring significant stabilisation (MD320 = 10.57 °C, MD328 = 9.96 °C, MD246 = 8.30 °C, MD316 = 7.82 °C). In contrast, most of the compounds appear to destabilise ABH2, suggesting that a degree of selectivity is possible between the two proteins. However, good stabilisation of ABH2 is observed with MD328, suggesting

that the presence of the small methyl substituent facilitates binding. Many of the compounds are also more potent inhibitors of AlkB than NOG (IC_{50} = 32 μM).

The structures of two of the compounds giving the largest T_m shifts (MD246 and MD316) each contain a portion of the structure of MD320, which gives the greatest stabilisation by DSF. These compounds also show good inhibition of AlkB, and MD316 (IC_{50} = 0.50 μM) and MD246 (IC_{50} = 3.39 μM) are more potent than the previously identified LNOC $3NO_2$. The 2-napthalene group appears to stabilise AlkB, and compounds with this substituent are more potent than those either lacking a substituent at this position (e.g. MD322, IC_{50} = 16.7 μM vs. MD308, IC_{50} = 54.1 μM) or with the alternate indole group (MD320, IC_{50} = 7.92 μM vs. MD330, IC_{50} = 14.1 μM).

	T_m shift /°C		IC_{50} /μM	T_m shift /°C	
	ΔN11 AlkB	**ΔN11 AlkB + NOG**	**ΔN11 AlkB**	**ΔN55 ABH2**	**ΔN55 ABH2 + NOG**
MD244	0.70	0.29	> 1000	-2.18	-1.23
MD328	9.96	8.03	9.5	3.08	1.80
MD312	1.08	0.89	54.1	-2.11	-2.41
MD246	8.30	6.72	3.39	-0.89	-0.40
MD330	3.65	2.18	14.1	1.07	0.39
EW38	5.75	4.48	42.5	1.67	0.15
MD314	3.83	3.23	5.35	0.70	-1.74
MD318	3.09	2.00	17.8	-5.89	-5.37
MD326	1.98	0.78	22.3	N/A	-5.58
MD316	7.82	7.35	0.5	-1.21	-3.56
MD310	-0.58	-0.22	165	-2.72	-2.28
MD324	1.65	0.06	14.7	N/A	N/A
MD320	10.57	9.12	7.92	0.31	-0.98
MD308	0.32	0.15	51.3	-4.48	-6.91
MD322	0.37	-0.60	16.7	-6.25	-5.63

Table 6-8: T_m shifts obtained by DSF for addition of 200 μM compounds to mixtures containing 2 μM protein and 50 μM Mn(II), with 5 % DMSO (v/v). 20 μM NOG was added where indicated. N/A: DSF curve did not follow a standard two-state transition, therefore no T_m shift could be determined. Inhibition assays for AlkB were carried out using a capillary electrophoresis method with a 15mer 3meC oligonucleotide.

Three of the most potent compounds identified, LNOC $3NO_2$, MD314, and MD316, were tested against the prolyl hydroxylase PHD2 (by Andrew Chan) and a representative histone demethylase, PHF8[391] (by Louise Walport) in standard assays. All three compounds have an IC_{50} > 1 mM, showing significant selectivity (200-2000-fold) of these compounds for AlkB.

6.5.4 *Comparing inhibition and T_m shift*

Comparison of the rankings of the DSF and inhibition results for the first and second generation inhibitors investigated (Table 6-9) using the Pearson product-moment correlation coefficient (Equation 6-1), gives ρ = 0.85 showing that there is a strong positive correlation between the two data sets. This suggests that the T_m shift is in general a good predictor of the inhibition potency of the compounds, therefore is valuable for use as an initial rapid screening method to identify possible inhibitor compounds or scaffolds.

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

Equation 6-1: Equation for calculation of the Pearson product-moment correlation coefficient (ρ). x_i and y_i are the rank orders of the x and y variables respectively; $\bar{x}$ and $\bar{y}$ are means of the x and y variables respectively.

Compound	IC_{50} /μM	IC_{50} rank	T_m shift /°C	T_m shift rank
MD316	0.50	1	7.82	4
MD246	3.39	2	8.30	3
LNOC $3NO_2$	5.22	3	3.97	6
MD314	5.35	4	3.83	7
MD320	7.92	5	10.57	1
MD328	9.50	6	9.96	2
MD330	14.1	7	3.65	8
MD324	14.7	8	1.65	13
MD322	16.7	9	0.37	16
MD318	17.8	10	3.09	9
MD326	22.3	11	1.98	11
NOG	32.0	12	2.89	10
EW38	42.5	13	5.75	5
LNOC 3OH	48.0	14	1.79	12
LNOC 2OH	50.4	15	1.28	14
MD308	51.3	16	0.32	18
MD312	54.1	17	1.08	15
MD310	165.0	18	-0.58	19
MD244	> 1000	20	0.70	17
LNOC	> 1000	20	-1.20	20
DNOC	> 1000	20	-1.24	21

Table 6-9: Comparison of inhibition and melting temperature rank orders.

6.5.5 *Crystal structures*

To identify the binding mode of the compounds with the greatest T_m shifts and highest potency, crystallisation with AlkB was attempted. Crystallisation and subsequent structure solutions were carried out by Wei Shen Aik and Dr Michael McDonough. The

crystal structures were solved by molecular replacement to a resolution of 2.0 Å with MD246 (PDB ID: 3T4V, Figure 6-14A), and to 1.7 Å with MD316 (PDB ID: 3T3Y, Figure 6-14B).

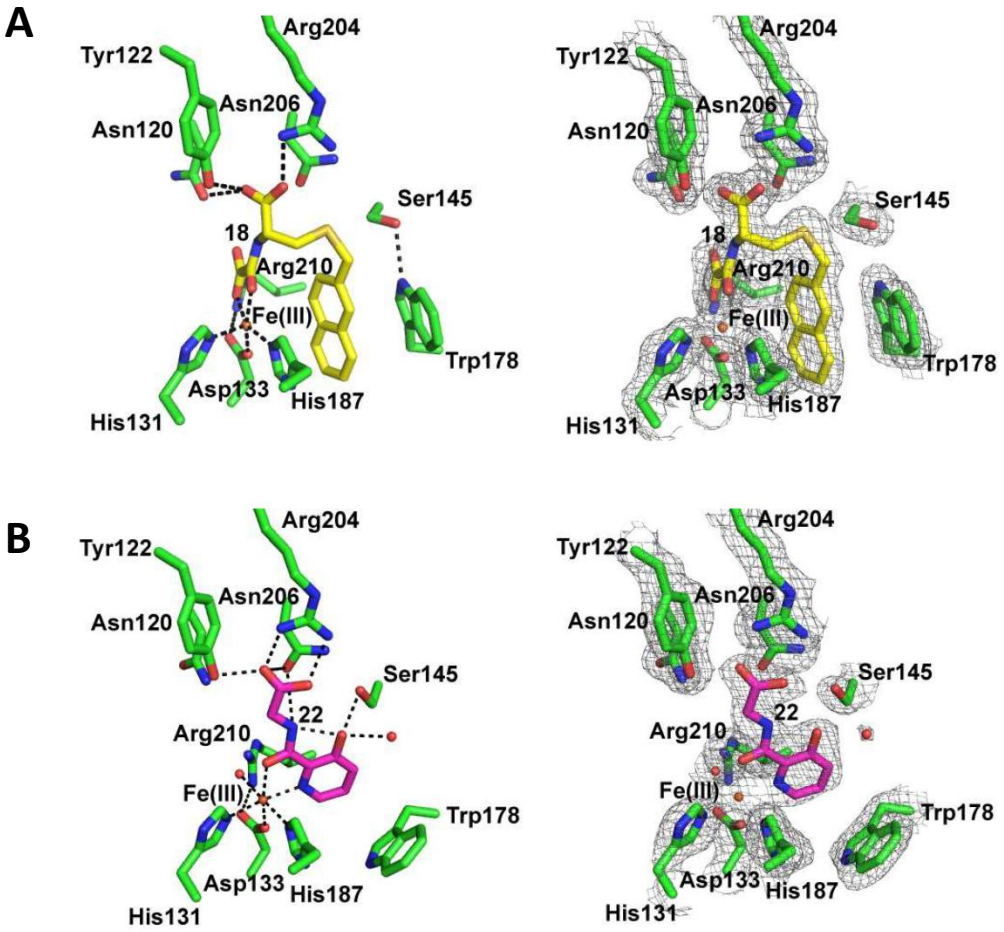


Figure 6-14: Active site views of crystal structures of ΔN11 AlkB bound to (A) MD316 (yellow sticks); PDB ID: 3T4V, (B) MD246 (magenta sticks); PDB ID: 3T3Y. Left: grey mesh represents experimental $2F_o$-F_c electron density map contoured to 1.0σ. Figures prepared by Wei Shen Aik.

The crystal structure of AlkB with the most potent compound, MD316 (IC_{50} = 0.50 μM), shows a very similar binding mode to that observed for LNOC $3NO_2$ previously (Section 6.4.4), with all hydrogen binding interactions (barring that of the 3-nitro group absent here) being conserved. In order to accommodate the bulkier naphthalene group the cysteinyl linker adopts a slightly different orientation to place the naphthalene in the same orientation as that for the phenyl ring of LNOC $3NO_2$, enabling similar π-stacking interactions to those observed previously. It appears that this

orientation may not be as favourable for the 1-napthalene ring of MD314, as this compound has reduced potency (IC_{50} = 5.35 μM).

In contrast, the crystal structure of AlkB with the smaller MD246 (IC_{50} = 3.39 μM) shows a considerably different binding mode (Figure 6-14B). Whereas it may be predicted that the pyridine ring of MD246 may coordinate in the same manner as the oxalyl group of the compounds previously crystallised, it instead binds in the reverse direction, with the pyridyl group projecting into the hydrophobic pocket occupied by the aromatic rings of the other compounds. The metal centre is therefore coordinated by the pyridyl nitrogen and the adjacent carbonyl oxygen, with the amide nitrogen forming hydrogen bonds to N206, and the carboxylic acid capable of hydrogen binding to Y122, R204 and N206. The 3-hydroxyl group can hydrogen bond to S145, an interaction which is also possible for MD320 and EW38, but not for the analogues lacking this substitution (MD244, MD318 and MD310), rationalising the greater potencies of the 3-hydroxyl-containing compounds. A considerable rotation around the Cα-Cβ bond of W178, placing the indole ring in a considerably altered position, allows a π-stacking interaction with MD246. This alternate orientation at the active site may explain the slightly reduced potency of MD320, which also has a pyridyl group, despite its high stabilisation by DSF, as there is likely to be less space available in the active site for binding of MD320 in this orientation. Although crystallisation has also been attempted for MD320 and EW38, compounds observed to bind well by DSF, no suitable crystals have yet been obtained.

6.6 Analysis in E.coli

6.6.1 E. coli *strains*

To further investigate the effect of the synthesised compounds on AlkB, *in vivo E. coli* studies were attempted. For this work, an *E. coli* knockout strain was obtained from the Keio collection[392] in which AlkB is deleted (JW2200). The wildtype strain (BW25113) from which the deletion strains are derived was used for comparison of growth.

6.6.2 Growth curve analysis

A previous high-throughput study of *E. coli* growth conditions identified the methylating reagent MMS as the most damaging compound to JW2200 *E. coli*[393]. In this previous study bacterial growth was analysed by comparison of colony areas for WT and knockout strains at a fixed timepoint[394-395]. A similar approach was used in the original investigations of the MMS sensitivity of AlkB mutants, through plating of cultures onto agar plates containing different MMS quantities, and assessing bacterial survival by counting colonies, typically after 16 h[155-156]. However it would be useful to follow the growth of bacteria over time, as different forms of damage may affect the growth of bacteria in different ways e.g. an extended lag-phase, or a growth arrest in the logarithmic phase. Therefore an optical density based plate-reader assay was used, developed from an initial protocol provided by Dr Gail Preston (Department of Plant Sciences, Oxford). The initial protocol involved inoculation from a bacterial lawn, but was developed to use a liquid starter culture, facilitating establishment of cultures of identical optical densities, and producing more consistent growth curves.

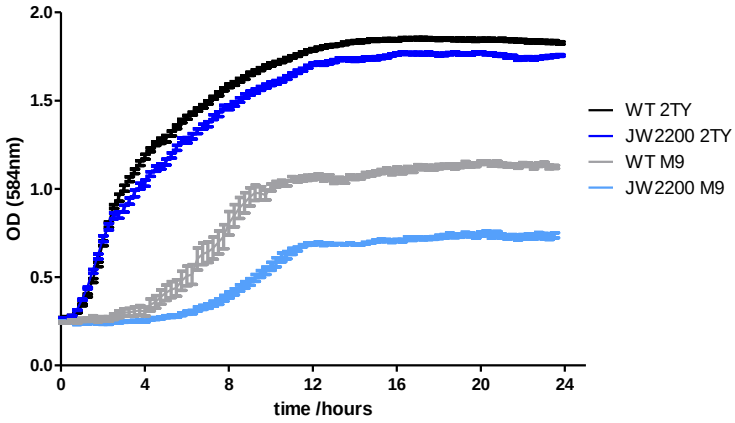


Figure 6-15: Bacterial growth curves following growth of wildtype BW25113 (WT) and AlkB deletion (JW2200) *E. coli* in either rich (2TY) or minimal (M9) media.

Initially, the growth of the JW2200 AlkB deletion strain was compared to that of the wildtype strain in both rich (2TY) and minimal (M9) media (Figure 6-15). The growth of the two strains was found to be very similar in rich media, but in minimal media the lag phase for the JW2200 strain was longer than for the wildtype bacteria, with growth apparently halted at a lower optical density. This suggests that the deletion strain may

grow in a manner similar to that of the wildtype strain when reagents are plentiful, but that under conditions of stress the AlkB deletion strain does not grow as efficiently, and presumably is less able to deal with the additional burden of unrepaired DNA damage. These results are consistent with those obtained in an initial study of the Keio collection strains[392]. This study found that in rich media after 22 h growth JW2200 has an optical density 94 % that of WT (here 95 %), whereas in minimal media JW2200 has an OD 67 % that of the wildtype strain after 24 h (here 66 %).

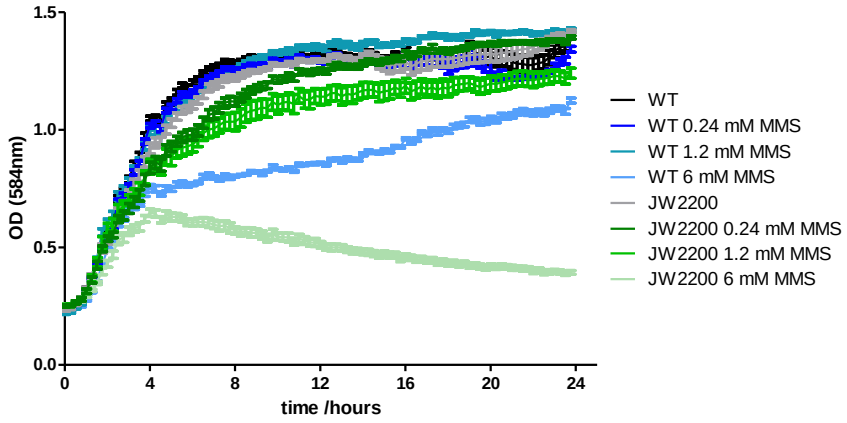


Figure 6-16: Bacterial growth curves following growth of wildtype BW25113 (WT) and AlkB deletion (JW2200) *E. coli* in 2TY rich media. Cultures grown in media containing indicated concentrations of MMS.

Subsequently, the addition of the S_N2 methylating agent MMS to bacterial growth media was investigated, by addition of a range of MMS concentrations to growth media (0.05 % (v/v) MMS (equivalent to 6 mM) is commonly used in the literature[154]). In 2TY media wildtype *E. coli* were able to withstand the addition of up to 1.2 mM MMS, but their growth was reduced in the presence of 6 mM MMS, with cells unable to reach the maximal optical density achieved by cells treated with reduced MMS. The growth of the JW2200 strain in rich media is slightly reduced even at the lowest concentration of MMS tested (240 μM), and growth was significantly reduced in the presence of 6mM MMS. MMS is likely to have the greatest effect during the exponential growth phase, during which most DNA synthesis occurs and DNA exists in a single stranded form, revealing the sites of preferential modification by S_N2 reagents (Chapter 1, Figure 1-15).

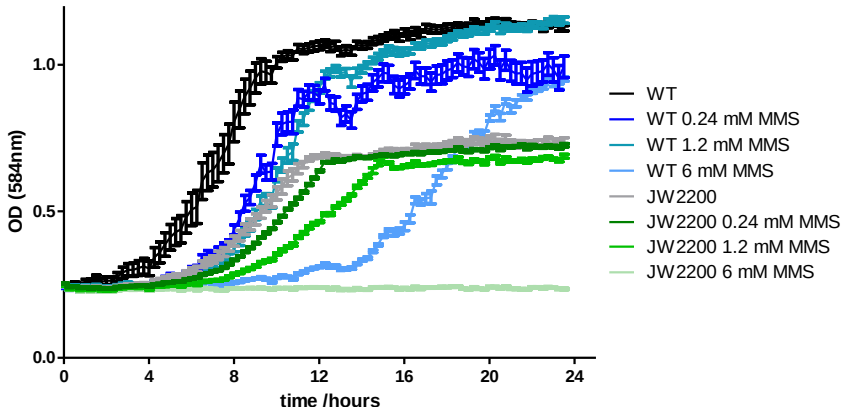


Figure 6-17: Bacterial growth curves following growth of wildtype BW25113 (WT) and AlkB deletion (JW2200) *E. coli* in M9 minimal media. Cultures grown in media containing indicated concentrations of MMS.

In minimal media a greater effect was observed on addition of MMS for both the wildtype and knockout strains (Figure 6-17). Here, the major effect on the bacterial growth when MMS was added was an extension of the initial lag phase, but JW2200 cells treated with 6 mM MMS did not show any growth.

6.6.3 Compound treatment

Monitoring cell growth was used to investigate the effect of addition of AlkB inhibitors. To attempt to determine whether any growth alterations were due to the inhibition of AlkB rather than cross-inhibition of other *E. coli* targets, the cells can be treated with MMS. Repair of DNA methylation would be hampered if AlkB is inhibited. Therefore wildtype cells in which AlkB is inhibited by an added compound would be likely to have a growth pattern similar to that for both treated and untreated JW2200. As growth of JW2200 is poor in minimal media, particularly on addition of MMS, it was decided to use 2TY media with the addition of 3 mM MMS, which should be sufficient to reduce cell growth, without leading to cell death. Compounds were prepared in DMSO, and bacterial growth was unaffected by the addition of 1% (v/v) DMSO.

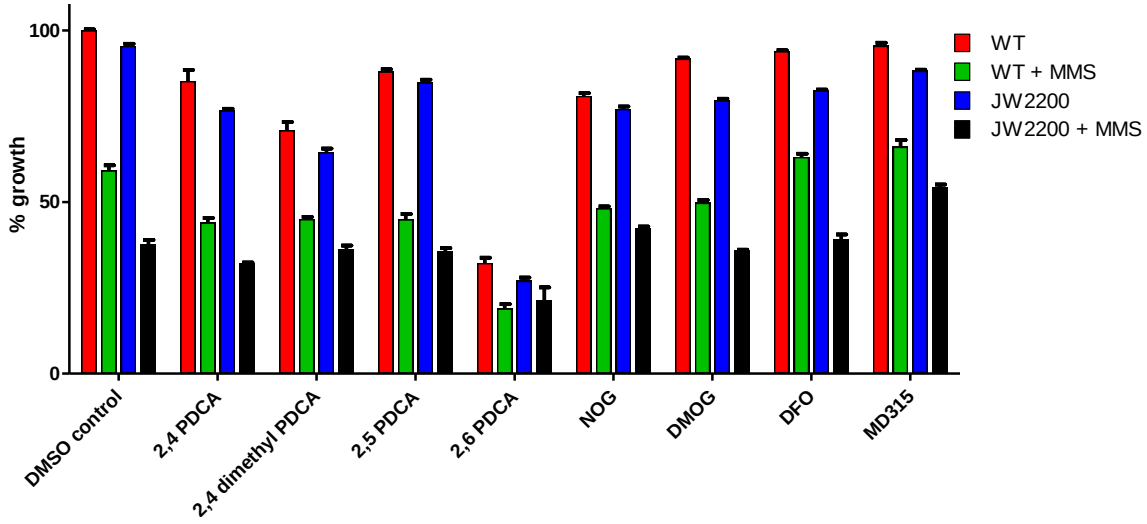


Figure 6-18: Relative optical density of wildtype (WT) and AlkB deletion (JW2200) cultures in 2TY media, 1 % DMSO (v/v) after 12 h growth. Cultures contain 3 mM MMS, and/or 10 mM small molecule (MD315 at 1 mM due to poor solubility) as indicated.

A range of compounds were tested, including pyridine dicarboxylic acid (e.g. 2,4-pyridine dicarboxylic acid; 2,4-PDCA) isomers, the siderophore deferoxamine (DFO), and the 2OG analogue NOG, and its dimethyl ester DMOG. Compounds were added to the cell cultures, which were then preincubated at 37 °C for 30 min to allow the cells take up the compounds, and recover from the transfer into fresh media. Following this, MMS was added to a final concentration of 3 mM.

To allow comparison between compounds added, the optical density of the cultures at 12 h was normalised to the optical density of WT cells to which no compound was added (Figure 6-18). It appears that the compounds are successfully taken up by the cells, as altered growth rates are observed on their addition. 2,6-PDCA appears to inhibit growth of cells under all conditions, therefore it likely interferes in another cellular process.

To facilitate the entry of compounds through the cell membrane of mammalian cells, methyl esters are commonly used, therefore 2,4-PDCA and NOG were tested alongside their dimethyl esters, although there was little obvious correlation between the presence of the methyl groups and the effect on the growth rate. Use of some of the selective compounds developed in this chapter was attempted, but the solubility in aqueous solutions was very poor, restricting the maximum compound concentration that could

be used. Addition of 1 mM MD315 (the methyl ester of MD316, the most potent compound identified here) did not produce a significant effect on the growth of either wildtype or JW2200 *E. coli*. Overall no clear effect on the growth of the *E. coli* strains was observed on addition of compounds, and as solubility of the AlkB-specific compounds is low it is difficult to determine whether AlkB is directly targeted.

6.7 Discussion

This chapter describes the development of selective and potent inhibitors of AlkB. A number of approaches were taken to identify potential inhibitors, including DSF, docking studies, DCMS and structure based design.

Initially, an FDH-coupled assay was used to determine IC_{50} values of AlkB and ABH2 with the generic 2OG oxygenase inhibitor NOG. Additionally, inhibition of ABH2 by the recently identified oncometabolite 2-hydroxyglutarate was determined, allowing comparison to inhibition data obtained for a suite of 2OG oxygenases, aiding studies of the effect of elevated 2HG on cells.

Following a broad screen of many 2OG mimic compounds by DSF a more focused approach was taken, using DCMS. Importantly, it was identified that AlkB has a strong preference for compounds with the L- stereochemistry at the α-carbon of the *N*-oxalyl amino acid scaffold investigated, the opposite preference to that observed previously for JMJD2A and FIH. DCMS identified the lead compound LNOC $3NO_2$ as a strong AlkB binder, and a more potent inhibitor than NOG. Unfortunately, due to significant inhibition of FDH by several of the developed inhibitors, the FDH assay could not be used to determine AlkB inhibition, so an alternate capillary electrophoresis method was employed.

An initial crystal structure of AlkB in complex with LNOC $3NO_2$ enabled development of a series of potent AlkB inhibitors, which were found to be selective for AlkB over representatives of two other 2OG oxygenase subfamilies. Contrary to expectations, further crystal structures of AlkB in complex with the two most potent inhibitors,

MD246 and MD316, revealed two alternate binding modes for the compounds at the metal centre.

Good correlation was observed between the T_m shift and inhibition data obtained, which validates the use of the DSF technique as a rapid screening method to identify compounds that are likely to inhibit the protein investigated.

This work is the first report of selective and potent inhibitors of a DNA modifying 2OG oxygenase. This therefore suggests that selective inhibition of the human DNA modifying 2OG oxygenases, including ABH1-8, FTO and TET1-3, is likely to be possible. In several cases this inhibition may be therapeutically beneficial, as several ABHs, particularly ABH2[169] and ABH8[220], are implicated in cancer, and FTO is linked to obesity[228].

Chapter 7 - Future Studies

Results obtained and observations made during the course of this work have suggested many future lines of investigation which would continue and complement the work described here:

7.1 FTO

A priority for the cellular study of FTO would be establishment of a cell line stably expressing tagged FTO. In the absence of an antibody suitable for successful specific anti-FTO IP reactions this would allow efficient pulldown reactions (using the fusion tag) at endogenous protein levels to enable ChIP assays.

In vitro studies would benefit from development of a robust LCMS assay, which could be used to directly follow the demethylation of the oligonucleotide substrates. This would require optimisation of the reaction conditions, and the LC method, and would allow relatively low-throughput testing of possible inhibitors. If used in conjunction with T_m shift data this could enable development of FTO inhibitors, which could be used further in cell-based assays to complement knock-out and knock-down studies, and help to understand the physiology of FTO. In addition, further crystallisation attempts should be made for the truncated ΔN31 mFTO protein. This would also be assisted by the development of compounds that bind strongly to FTO, to produce stabilisation during the crystallisation process.

7.2 ABH1 and ABH7

As the substrates for ABH1 and ABH7 are currently unknown or unclear, a priority for this work would be identification of the true substrates. To attempt to achieve this a wider range of nucleic acid substrates, particularly RNA, substrates should be tested. Additionally, crystallisation of the proteins would be very useful. As the truncated ΔN17 ABH7 construct purifies well and is stable in solution it may be particularly amenable to crystallisation. Knowledge of the three dimensional structure of the protein could assist in identifying any substrate binding regions, and any distinctive features of the active sites. Crystallisation of ABH1 would also be particularly interesting as it may

identify any additional domains or possible sites which could be responsible for the postulated DNA lyase activity of the protein.

Further attempts should be made to extract functional mitochondria from cultured cells, to allow the study of endogenous ABH1 and ABH7. This would enable further activity studies on native proteins, as the recombinantly expressed proteins may lack modifications required for successful activity. Proteomic studies should be repeated using the commercial anti-ABH1 antibody in the pulldown reaction, as this antibody was found to be successful in the test IP reaction. This should hopefully enable identification of whether any proteins interact specifically with ABH1, rather than with the fusion tag, which would assist with the identification of the ABH1 substrate.

Furthermore, a commercial ABH7 antibody has recently become available. Therefore this should be tested for suitability in the range of techniques used here (e.g. recognition of ABH7 in western blots, immunohistochemistry, and IP reactions). ChIP assays could also be carried out using these antibodies, to investigate any nucleic acid species with which the proteins may interact.

7.3 AlkB

As potent selective inhibitors of AlkB have been developed, the DCMS approach successfully used here can be expanded for the development of inhibitors of ABH2 and ABH3. As revealed by T_m shift analyses it appears that the preference of ABH2 is likely to be different to that of AlkB, and inspection of the published crystal structures shows that a further alternate preference is likely for ABH3. Therefore this could yield inhibitors that are selective for each of the AlkB family members, and hopefully also selective for the DNA methylases over other 2OG oxygenases.

In order to carry out these studies, crystallisation of the truncated ABH2 and ABH3 should be optimised, and improved assays for determining the activity of ABH3 are required.

Chapter 8 - Materials and Methods

Chemicals were supplied by Sigma-Aldrich unless otherwise stated. Water used was from a Millipore Elix® 10 system, purified further through a 0.22 μm filter at the outlet by a Millipore Milli-Q system.

Microbiological techniques:

Standard sterile practices were followed throughout, with procedures carried out in a Biological Safety Hood (Heraeus KJ12). All equipment and media used were sterilised prior to use by autoclaving (Thermo Life Sciences MAT 490 LEI) at 121°C for 20 mins. Solutions of heat-sensitive substances were sterilized using 0.20 μm filters. Tryptone, yeast extract and agar (bacteriological) for use in culture media were obtained from Oxoid.

pH measurement:

A Hanna Instruments HI 9321 microprocessor pH meter fitted with either a 5 mm or 1.5 mm diameter electrode was used for pH measurement. Prior to each use the probe was calibrated in the range pH 4-7 or pH 7-10, appropriate to the required pH, using standard calibration solutions (Fisher Scientific) then stored in 4 M KCl solution between uses. The pH of solutions was adjusted using stock solutions of NaOH or HCl as necessary.

Waterbaths:

A Grant W14 waterbath was used for transformations at 42 ˚C, and a Grant Y14 waterbath was used for incubations at 37 ˚C.

Bacterial growth media:

Selective antibiotics (Melford Laboratories Ltd) were added post-autoclaving, kanamycin to 30 μg ml^{-1}, ampicillin to 100 μg ml^{-1} and chloramphenicol to 30 μg ml^{-1}.

LB agar – (per litre) 10 g Tryptone, 10 g NaCl, 5 g Yeast extract. To obtain a solid medium, 20 g agar was added to media before autoclaving.

2TY – (per litre) 16 g Tryptone, 5 g NaCl, 10 g Yeast extract, pH 7.0. Before autoclaving, the medium is adjusted to pH 7.0. 2TY medium used for large-scale growths was prepared by Mr H. M. Jubier (Chemistry Research Laboratory, Oxford)

SOC – (per litre) 980 ml 2TY, 10 ml 1 M $MgCl_2$, 10 ml 2 M glucose.

Vectors:

cDNA encoding the mFTO sequence was obtained as an I.M.A.G.E. clone (#4237261, I.M.A.G.E. Consortium), which contained 1506 base pairs coding for a 502 amino acid protein with a predicted mass of approximately 57 kDa. This was subcloned as a NheI and EcoRI fragment by Dr Thomas Gerken into the pET-28a(+) (which contains an *N*-terminal 20 residue His_6 tag) vector (Novagen), creating mFTOa[165]. The *N*-terminal fusion peptide containing the thrombin cleavage site was deleted creating the mFTO construct.

The coding sequence of human FTO, sequence optimized for expression in *E. coli* and with a C-terminal TAA stop codon, was commercially synthesised (GeneArt) then subcloned into the NdeI & BamHI sites of pET-28a(+), by Dr C. Webby.

The sequence encoding AlkB was cloned as an NdeI and BamHI fragment into pET-24a(+) and pET-28a(+) by Dr R. Welford[325].

ABH2 was amplified from human cDNA and cloned into the NdeI and NotI sites of pET-28a(+) by Dr Kirsty Hewitson

PCR:

Oligonucleotide primers were synthesised to order by Sigma-Genosys, obtained as lyophilised samples then resuspended in sterile water as 100 μM stocks. Sequences of all primers used in this work are presented in Appendix B. Polymerase chain reactions (PCRs) were performed in thin walled 0.5 ml Eppendorf tube

DNA amplification

PCR was performed to amplify DNA sequences for cloning into alternative vectors.

Reaction mixture:

10 × reaction buffer	5 µl
Plasmid template	100 ng
Oligonucleotide primers	0.5 µM
dNTP mix (2.5 mM each dNTP)	0.2 mM
DMSO	0/5/10 %
PfuTurbo® DNA polymerase (2.5 U/µl)	1 µl
MilliQ H_2O	to final volume of 50 µl

Function	Temperature	Time	Number of cycles
Denaturation of template DNA	95 °C	2 min	1
Denaturation	95 °C	30 s	7
Annealing	5 °C below $T_{an}1$	30 s	
Extension	72 °C	1 min/kb	
Denaturation	95 °C	30 s	25
Annealing	5 °C below $T_{an}2$	30 s	
Extension	72 °C	1 min/kb	
Final extension	72 °C	10 min	1

Table 8-1: PCR cycling conditions used for DNA amplification. $T_{an}1$: Annealing temperature for fraction of primer binding to original parent template. $T_{an}2$: Annealing temperature for binding of full primer to newly synthesised daughter sequence.

Site-directed Mutagenesis

All single-site site directed mutageneses (point mutations or deletions) were carried out using the Stratagene QuikChange® site-directed mutagenesis kit. Briefly, primers incorporating the desired mutation were used to amplify both vector and insert (thermal cycling as in Table 8-2). Template DNA, which is extracted from a bacterial clone and therefore methylated, was digested with methylation specific DpnI. To increase the proportion of bacterial colonies containing the newly synthesised mutated DNA rather than the unmodified parental template DpnI digestion was allowed to proceed at 37 °C for 2 h rather than the recommended 1 h. The mutated DNA was then transformed into XL10-Gold competent cells for growth and subsequent plasmid DNA extraction. Typically, >80 % of colonies selected contained the desired mutation.

Reaction mixture:

10 × reaction buffer	5 µl
dsDNA template	25 ng
Oligonucleotide primers	125 ng
dNTP mix (2.5 mM each dNTP)	5 µl
PfuTurbo® DNA polymerase (2.5 U/µl)	1 µl
MilliQ H_2O	to final volume of 50 µl

Function	Temperature	Time	Number of cycles
Denaturation of template DNA	95 °C	2 min	1
Denaturation	95 °C	30 s	16 (point mutation) 18 (deletion)
Annealing	50 °C	1 min	
Extension	72 °C	1 min /kb	
Final extension	72 °C	10 min	1

Table 8-2: PCR cycling conditions used for site directed mutageneses.

Multi site-directed mutagenesis

Multiple site-directed mutations were introduced simultaneously using the Stratagene QuikChange® Multi site-directed mutagenesis kit (thermal cycling carried out as detailed in Table 8-3).

Reaction mixture:

10 × Multi reaction buffer	2.5 µl
dsDNA template	100 ng
Oligonucleotide primers	100 ng each
dNTP mix	1 µl
QuikSolution	0.5 µl
Quikchange Multi enzyme blend (2.5 U/µl)	1 µl
MilliQ H_2O	to final volume of 25 µl

Function	Temperature	Time	Number of cycles
Denaturation of template DNA	95 °C	1 min	1
Denaturation	95 °C	1 min	30
Annealing	55 °C	1 min	
Extension	65 °C	2 min /kb	

Table 8-3: PCR cycling conditions used for Multi site directed mutageneses.

Restriction digestion:

DNA digests were carried out using the appropriate restriction enzymes (NEB) in the recommended NEB buffers. Reactions were carried out in volumes of 10-60 µl as

required, containing 2-10 U enzyme, with BSA added if necessary. Reactions were incubated at 37°C for 2 hours unless otherwise stated.

Preparative:

Gel extracted PCR product	30 µl
Restriction enzyme A	1 µl
Restriction enzyme B	1 µl
10 x NEB buffer	4 µl
10 x BSA	4 µl
H_2O	to final volume of 40 µl

Purified vector	2 µl
Restriction enzyme A	1 µl
Restriction enzyme B	1 µl
10 x NEB buffer	2 µl
10 x BSA	2 µl
H_2O	to final volume of 20 µl

Analytical digest:

Miniprepped DNA	5 µl
Restriction enzyme A	0.5 µl
Restriction enzyme B	0.5 µl
10 x NEB buffer	1 µl
10 x BSA	1 µl
H_2O	to final volume of 15 µl

Ligation:

Insert and vector were digested with appropriate restriction enzymes creating complementary overhanging ends. The vector was dephosphorylated at the 5' end using alkaline phosphatase (CIP; NEB) to reduce self-ligation. The vector restriction digest was heated to 80°C for 20 minutes, inactivating the restriction enzymes, cooled to room temperature, then 1 µl alkaline phosphatase added and incubated for 2 hours at 37°C. The digestion products were run on agarose gels and purified by extraction. Ligation reactions were incubated overnight at 16°C, before transformation of 7.5 µl ligation mix into XL1 Blue competent cells.

Reaction mixture:

T4 DNA ligase	1 µl
10x T4 ligase buffer	1.5 µl
Doubly digested insert	3 parts
Doubly digested vector	1 parts
H_2O	to final volume of 15 µl

Restriction digests were carried out to verify that the plasmids contained the desired insert, and digest results were visualised by agarose gel electrophoresis.

Subcloning:

Subcloning of constructs between vectors with compatible restriction enzyme sites was achieved by restriction digest of purified plasmid DNA templates for both vector and insert.

DNA gel electrophoresis:

Agarose gel electrophoresis was used for DNA visualisation. 1 % agarose gels were formed by boiling 0.4 g agarose (Bioline) in 40 ml TAE buffer until fully dissolved, using a SuperWave 750 microwave (Tecnolec), with regular mixing throughout. This was allowed to cool to 60°C, before addition of 4 µl SYBR Safe™ DNA gel stain (Invitrogen). Gel were cast in an H2-SET gel tray (Anachem) to a depth of 0.7 cm and a comb inserted at one end of the gel to enable sample loading, then allowed to cool at room temperature for the gel to set. Gels were run on a Bio-Rad system at a constant potential of 100 V for approximately 45 min, then bands visualised by illumination of the gel with UV light using a Gel Logic 200 imaging system (Kodak) and a 535 nm emission filter. Samples were run against a 2-log DNA Ladder 0.1-10 kb (Invitrogen) allowing molecular weight determination.

50 x TAE (running buffer solution) – 2 M Tris, 1 M acetic acid, 0.5 M EDTA, pH 8.0.

5 x sample loading buffer – 70% glycerol, 30% 1xTAE, 0.25% bromophenol blue.

Gel extraction:

Bands were excised from agarose gels using a clean scalpel, weighed, and DNA extracted using a Qiagen Gel Extraction kit, according to the manufacturer's protocol. DNA was eluted in 30 µl H_2O.

DNA purification:

A Qiagen Miniprep DNA purification kit was used according to the manufacturer's instructions for small scale DNA purification. DNA was eluted in 30 µl H_2O.

Qiagen Maxiprep DNA purification kits were used for large scale purification of transfection-grade DNA.

DNA quantification:

Concentration was determined through measurement of light absorbance at 260 nm using a NanoDrop ND-1000 UV/Vis Spectrophotometer.

DNA sequencing:

DNA sequencing was performed by Source BioScience, Oxford.

Competent cell strains:

E. coli BL21 (DE3) was used for protein expression[395], *E. coli* XL10 Gold for transformation of mutageneses and ligations, and *E. coli* XL1-Blue for general cloning applications (all Stratagene). Genotypes of these strains are given in Table 8-4.

Strain	Genotype	Supplier
BL21 (DE3)	F^- dcm *ompT hsd* $S(r^-m^-)$ *gal* λ(DE3) (chromosomal copy of the T7 RNA polymerase)	Stratagene
XL1 Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F′ *proAB lacI* q*Z* Δ*M15* Tn*10* (Tet^r)]	Stratagene
XL10 Gold	Tet^r D(mcrA)183 D(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F′ proAB lacIqZDM15 Tn10 (Tet^r) Amy Cam^r]	Stratagene
W3110 (Keio collection BW25113)	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, rph-1, Δ(rhaD-rhaB)568, hsdR514	Keio collection
Keio collection JW2200	F-, Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ-, ΔalkB765::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514	Keio collection

Table 8-4: Genotypes of *E. coli* strains used in this study.

Competent cell preparation:

Aliquots of competent cells (XL1-Blue and BL21 (DE3)) were prepared from commercially obtained stocks (Stratagene), in antibiotic-free media. The appropriate strain was streaked onto an LBA plate, and colonies grown overnight. A single colony was used to inoculate 100 ml sterile 2TY medium, which was incubated in a shaker incubator (220 rpm) at 37°C overnight. 800 µl starter culture was used to inoculate 400 ml sterile growth medium, which was grown with shaking at 37°C to an OD_{600nm} of 0.6-0.8. The culture was pelleted in pre-chilled centrifuge tubes, by centrifugation at 750xg, 4°C for 5 mins. Pellets were gently resuspended in 200 ml ice cold 100 mM $CaCl_2$, 20% glycerol, and the suspension incubated on ice for 20 minutes. Cells were re-pelleted by centrifugation at 750×g, 4°C for 5 mins and gently resuspended in 8 ml

100 mM $CaCl_2$, 20 % glycerol. Cells were stored as 100 µl aliquots in freshly sterilised 0.5 ml Eppendorf tubes, then immediately stored at -80 °C.

Transformations:

E. coli BL21 (DE3) competent cells were thawed on ice then 0.5 µl plasmid DNA added to 30 µl competent cells in a sterile 15 ml Falcon tube. Cells were chilled on ice for 30 mins before being subjected toheat shock at 42 °C for 45 s and returning to ice for 2 mins. 1 ml SOC medium was added to cells, which were then incubated in a 37 °C shaker incubator for 1 hour. 100 µl of the transformation mixture was plated onto an agar plate containing the appropriate selective antibiotic and incubated overnight at 37 °C in a Gallenkamp Duostat incubator.

XL1 Blue and XL10 Gold were both transformed according to the suppliers protocol (as for BL21 (DE3), but heat shock 30 s). Cells were harvested by centrifugation (1500 rpm), resuspended in 100 µl fresh 2TY medium, then plated.

Starter culture preparation and growth:

2TY medium (10 ml or 100 ml) was inoculated with a single colony from an agar plate, then incubated overnight in a New Brunswick Scientific G25 environmental shaker at 37 °C and 250 rpm.

Glycerol stocks:

A cell pellet obtained from an overnight starter culture was resuspended in 400 µl 2TY media. 400 µl glycerol stock solution (65 % glycerol, 0.1 M $MgSO_4$, 25 mM Tris, pH 7.5) was added, and the mixture stored at -80 °C.

Large scale growths:

12 x 2 l flasks containing 600 ml 2TY media were inoculated with 7 ml starter culture. Flasks were incubated at 37°C at 220 rpm in a New Brunswick Scientific G25 environmental shaker and once an $OD_{600\ nm}$ of at least 0.6 was reached (measured using a 1 ml sample of bacterial culture compared to a 1 ml reference solution of initial growth medium in a Novaspec II spectrophotometer, Pharmacia), the appropriate concentration of Isopropyl-β-D-1-thiogalactopyranoside (IPTG, Melford Laboratories

Ltd) (0.1-0.5 mM) was added to induce protein expression. Cultures were further incubated under the optimal conditions determined by expression trials, then cells harvested by centrifugation in a JLA 9.100 rotor in a Beckman Avanti J25 centrifuge at 9 krpm and 4 °C for 8 mins. The resulting cell pellet was weighed and stored at -80 °C.

Protein expression trials:

Protein	Optimal expression conditions		
	Temperature /°C	Time /hours	IPTG /mM
mFTOa	18	4	0.5
mFTO	18	4	0.5
mFTO NCBI	Poor expression		
1-E408 mFTO	Very poor expression		
1-K387 mFTO	Poor expression		
1-D329 mFTO	Very poor expression		
1-A321 mFTO	Very poor expression		
CmFTOa	15	16	0.5
CmFTO	15	16	0.5
CmFTO NCBI	15	16	0.5
I367F mFTO	insoluble		
I367F CmFTO	insoluble		
I367A mFTO	15	16	0.2
I367A CmFTO	15	4	0.2
I367A mFTO NCBI	Very poor expression		
I367A CmFTO NCBI	Very poor expression		
hFTOa	15	4	0.5
hFTO	15	4	0.5
ChFTO	15	4	0.2
pET24a AlkB	28	4	0.2
pET24a AlkB ΔN11	15	16	0.2
pET28a AlkB ΔN11	28	16	0.2
ABH1	28	16	0.5
ABH2	28	4	0.2
ABH2 ΔN55	18	16	0.2
ABH7	Insoluble		
ABH7 ΔN14	37	4	0.5

Table 8-5: Optimal protein expression conditions used.

100 ml starter cultures were prepared as described above until induction. Flasks were then incubated at 15 °C, 28 °C and 37 °C and 220 rpm for 3 hours and 50 ml of each culture harvested in an S-4180 rotor in a Beckman Allegra™ 21R bench top centrifuge at 4 krpm and 4 °C for 10 mins. Resulting cell pellets were stored at -20 °C with remaining cultures left at the appropriate temperature overnight before harvesting.

Cell pellets were resuspended in 5 ml 50 mM Tris pH 7.5, then lysed by sonication on ice with two 30 s pulses separated by a 30 s pause using an MSE Soniprep 150 sonicator

at an amplitude of 10 μm. 1 ml of the cell suspension was transferred to a sterile 1.5 ml Eppendorf tube and spun for 1 min at 13 krpm in a Jouan A-14 microcentrifuge. Soluble protein was removed and the remaining insoluble precipitate resuspended in 1 ml 50 mM Tris pH 7.5. The concentration of protein in each sample was determined using a NanoDrop ND-1000 UV/Vis Spectrophotometer. 10 μl of the protein solutions were analysed by SDS-PAGE revealing conditions producing greatest soluble protein. Negative (transformed with wildtype protein plasmid, without induction) and positive (cells expressing wildtype or parent constructs under optimal expression conditions) controls were run alongside mutants for comparison. Expression conditions determined for the proteins used are summarised in Table 8-5.

Large-scale cell lysate preparation:

Cell pellets obtained from large scale growths were resuspended in approximately 5 times their mass of the appropriate buffer for the column to be used, with addition of 100 μl 1 M $MgCl_2$, a Roche cOmplete EDTA-free protease inhibitor cocktail tablet and < 1 mg DNase (bovine pancreas, grade II, Roche). This mixture was stirred at 4°C until homogeneous. Cells were lysed on ice by sonication, with a Sonics VibraCell VCX-500 and 13 mm probe at 60 % intensity, for 4 x 45 s separated by 45 s breaks. Insoluble cell debris was removed by centrifugation in a JA-25.50 rotor at 23 krpm for 20 mins, each at 4 °C. An additional initial centrifugation step was carried out in JLA-16.250 rotor in a Beckman Avanti J25 centrifuge at 14500 rpm for 20 mins in the purification of AlkB proteins. Supernatants were filtered to 0.20 μm before purification.

Protein purification:

Purification was carried out at 4 °C using ÄKTA-FPLC™ systems (GE Healthcare, P920 pump system, UPC900 UV detector, Frac900 fraction collector, controlled by Amersham Pharmacia Unicorn Software V5.1). Prior to use all buffers were freshly prepared with Milli-Q water and filtered with a 0.20 μm filter.

Cation exchange (SP Sepharose fast flow 50)

S Sepharose Buffer A: 0.1 M MES, pH 5.8

S Sepharose Buffer B: 0.1 M MES, 1 M NaCl, pH 5.8

Purification was performed with a 50 ml settled bed volume S ≈epharose (GE Healthcare) column, equilibrated with 2CV Milli-Q water, 2CV 1 M NaCl and 2CV S buffer A. Cells were lysed in S buffer A, the filtered supernatent loaded onto the column at 4 ml min^{-1}, then the column washed with S buffer A until the UV trace returned to its initial level. To elute protein a gradient of S buffer B was run over 8CV, and 5 ml fractions collected. SDS-PAGE and UV trace analysis allowed determination of the fractions containing the required protein, which were combined and concentrated to 2 ml using a 10 kDa MWCO concentrator in an S-4180 rotor in a Beckman Allegra™ 21R bench top centrifuge at 4 krpm and 4 °C. Following use the column was cleaned by washing with 2CV each of water, 1 M NaOH, water, 1 M NaCl, 1 M AcOH and water.

His bind column

Buffer	NaCl /mM	Tris HCl /mM	Imidazole /mM	EDTA /mM	β-mercaptoethanol /mM	$NiSO_4$ /mM	pH
Binding	500	20	10	-	1	-	7.5
Wash	500	20	40	-	1	-	7.5
Elution	500	20	500	-	1	-	7.5
Strip	300	20	-	100		-	7.5
Charge	-	-	-	-		50	-

Table 8-6: Nickel IMAC purification buffer compositions.

Nickel immobilised metal affinity chromatography (IMAC) purification was carried out using a 5 ml HisTrap FF column (GE Healthcare), which was manually charged and equilibrated using 5CV water, 5CV Charge buffer, 5CV Binding buffer (buffer composition in Table 8-6). After loading the cell lysate, the column was washed with several CV Binding buffer until the 280 nm UV trace fell to approximately zero. The column was then washed with 10-15CV Wash buffer, again until a stable 280 nm baseline was achieved. The protein of interest was eluted in a gradient of 8CV Elution buffer and collected as 5 ml fractions. Fractions were analysed by SDS-PAGE to determine those containing the greatest amount of pure protein, which were pooled and

diluted 1:1 in gel filtration buffer, before concentration to ~2 ml using a 10 kDa MWCO ultrafiltration device (Amicon). Following use the column was manually stripped of Ni(II) using 5CV Strip buffer, then cleaned and stored according to the manufacturers recommendations.

Gel filtration column

Gel filtration buffer: 100 mM tris HCl, 100 mM NaCl, 1 mM DTT, pH 7.5.

Previously purified and concentrated protein was loaded onto a 320 ml gel filtration Superdex 75 column (Pharmacia) containing S75 gel filtration resin (Amersham Biosciences). Loading was via a 2 ml injection loop at 0.5 ml min^{-1}, then subsequently run at 2 ml min^{-1}. The column was equilibrated by washing with 1.1CV Milli-Q water, and 1.1CV gel filtration buffer. Prior to use the loop was washed with two volumes each of Milli-Q water and gel filtration buffer. Protein was eluted with 1CV of gel filtration buffer, with 5 ml fractions collected when the UV trace exceeded a threshold absorption limit. SDS-PAGE was used to determine the fractions to be combined then concentrated. The resulting protein was exchanged into the appropriate buffer for storage by further concentration. The concentration of the purified protein was determined then the concentrated protein stored at -80°C in appropriate aliquots. Following use, the column was washed with Milli-Q water.

Protein concentration determination:

Concentrated, desalted protein was spun in a microcentrifuge at 13 krpm for 1 min to remove precipitate or dust, then transferred to a fresh 0.5 ml tube. Concentration was determined by measurement of the absorbance at 280 nm using a NanoDrop ND-1000 UV/Vis Spectrophotometer. $A = \varepsilon c l$, so if ε is known, c can be determined. 2 µl of the appropriate buffer was used as a blank, and 2 µl protein used for measurements. Protein extinction coefficients and molecular weights were estimated using the ExPASy ProtParam tool[396].

Polyacrylamide gel electrophoresis:

Denaturing SDS-PAGE (polyacrylamide gel electrophoresis) was conducted as described[397], using SeeBlue® Plus 2 (Invitrogen) as a molecular weight marker. Samples were initially heated at 100 °C for 5 mins prior to loading. Electrophoresis was run for 50 mins at 200 V and 20 °C on a Bio-Rad Mini-PROTEAN II system, then gels stained for 15 mins, and placed in SDS-PAGE destain for > 30 min, with shaking.

Native-PAGE materials were prepared as for SDS-PAGE, with SDS and β-mercaptoethanol replaced by water. Native gels were run at 4 °C and 80 V, using appropriate protein standards of known molecular weight and oligomerisation state as markers for comparison.

Images of gels were recorded (scanned using a CanoScan LiDE 25 scanner) and some gels were preserved with a gel drying kit (Promega).

Separating gel 15 % acrylamide: (per gel) 1.25 ml acrylamide, 1.25 ml 2 x separating gel buffer (750 mM Tris-HCl, 0.2 % SDS, pH 8.8), 5 µl TEMED, 15 µl 25 % APS.

Stacking gel 4.5 % acrylamide: (per gel) 700 µl Milli-Q, 300 µl acrylamide, 1.0 ml 2 x stacking gel buffer (350 mM Tris-HCl, 0.2 % SDS, pH 6.8), 5 µl TEMED, 15 µl 25 % APS. (TEMED and APS were added to the mixture immediately prior to pouring of both gels).

10 x running buffer: 250 mM Tris, 1.92 M glycine, 10% SDS, pH 8.8.

2 x sample loading buffer: 10 % 0.5 M Tris-HCl; pH 6.8, 0.5 % bromophenol blue, 2 % SDS, 12 % Glycerol, 0.5 % β-mercaptoethanol, 53 % Milli-Q water.

Stain: 0.5 % Coomassie Brilliant Blue, 50 % methanol, 10 % glacial acetic acid.

Destain: 10 % methanol, 10 % glacial acetic acid.

Alternatively the Invitrogen NuPAGE® system was used for analysis of IP protein samples to be prepared for mass spectrometry. NuPAGE® Bis-Tris 4-12 % gels were run in NuPAGE® MOPS SDS-PAGE running buffer, with NuPAGE® antioxidant. Gels were stained using the Colloidal Blue Staining Kit (Invitrogen), following the manufacturers protocol.

Protein buffer exchange:

Micro BioSpin6 Chromatography columns (Bio-Rad) were used following the supplier's procedure, including 4 washes with new buffer before application of the sample. Alternatively, larger quantities of protein were more rapidly exchanged into an alternate buffer using a NICK column (Amersham Biosciences). Protein solutions were concentrated to 3 ml then run through the equilibrated columns according to the manufacturer's instructions, for exchange into the new buffer.

Analytical gel filtration:

Standard	Molecular mass /kDa	Amount /µg	
		S200	S75
Blue Dextran	2000	200	-
Apoferritin (Equine spleen)	443	200	-
Alcohol Dehydrogenase (*S. cerevisiae*)	150	200	-
Albumin (Bovine serum)	66	200	200
Albumin (Chicken Egg)	42.7	-	500
Carbonic Anhydrase (Bovine erythrocytes)	29	200	200
Cytochrome C (Equine heart)	12.4	200	200
Aprotinin (Bovine lung)	6.5	-	200

Table 8-7: Molecular weight standards used for analytical gel filtration.

A Superdex 75 10/300 GL column or Superdex 200 10/300 GL column (GE Healthcare) were equilibrated with 1CV Protein Storage Buffer. 250 µg of the protein of interest was loaded onto the column via a 100 µl injection loop. Eluted volume was collected as 1 ml fractions and the UV trace monitored. Fractions containing protein were analysed by SDS-PAGE to confirm protein identity. Protein standards used (Table 8-7) were premixed before injection, then analysed by the same method, creating calibration curves.

Circular Dichroism:

Circular dichroism (CD) measurements were taken with an Applied Photophysics Chirascan machine, using 0.25 mg/ml protein samples in 300 µl CD buffer (50 mM KCl, 20 mM KPO_4, pH 7.0). In all cases, blank samples of CD buffer were run initially and subtracted from other spectra to allowing direct protein comparison. Scans were

carried out at the appropriate temperature over 190-260 nm, with measurements taken every 0.5 nm, repeated four times and averaged. Data were smoothed using a window size of 2. The supplied CDNN software (Dr. Gerald Böhm, Institut für Biotechnologie, http://www.photophysics.com/cdnn.php, Martin-Luther Universität Halle-Wittenberg, Germany) was used for curve deconvolution into secondary structure elements.

Oligonucleotides:

Oligonucleotides used in enzymatic assays were commercially synthesised by ATDBio Ltd, or Sigma Genosys (details in Table 8-8).

Name	Sequence	Mass /Da	Supplier
T	GCTAGGTCCCGTAGTGCG	5532.0	Sigma Genosys
3meT	GC(**3meT**)AGGTCCCGTAGTGCG	5545.3	ATDBio Ltd
cmpT	CGCACTACGGGACCTAGC	5469.5	ATDBio Ltd
C	GCCAGGTCCCGTAGTGCG	5557.0	Sigma Genosys
3meC	GC(**3meC**)AGGTCCCGTAGTGCG	5530.5	ATDBio Ltd
5meC	GC(**5meC**)AGGTCCCGTAGTGCG	5530.5	ATDBio Ltd
A	GCAAGGTCCCGTAGTGCG	5541.0	Sigma Genosys
1meA	GC(**1meA**)AGGTCCCGTAGTGCG	5554.5	ATDBio Ltd
G	GCGAGGTCCCGTAGTGCG	5517.0	Sigma Genosys
1meG	GC(**1meG**)AGGTCCCGTAGTGCG	5570.3	ATDBio Ltd
3meC-FAM	TT(**3meC**)TTTTTTTTTTT-fluorescein	5066.0	ATDBio Ltd
C-FAM	TTCTTTTTTTTTTT-fluorescein	5052.0	ATDBio Ltd

Table 8-8: Oligonucleotides used in demethylation assays and binding experiments.

Activity assays:

To determine the extent of reaction of 2OG dependent oxygenases, a 2OG turnover assay has been developed[291]. This follows the reaction through turnover of [1-^{14}C] labelled 2OG, releasing $^{14}CO_2$ gas (Fig. 3). Standard conditions employed were 4 μM enzyme, 100 μM substrate, 4 mM ascorbic acid, 300 μM 2OG (4 % [1-^{14}C]-2OG (Perkin Elmer)), 0.3 mg/ml catalase, 1 mM DTT and 50 μM $(NH_4)_2Fe(SO_4)_2.7H_2O$ (dissolved first to 250 mM with 20 mM HCl, then 2.5 mM with Milli-Q water). Assay mixtures were made up to 100 μl with 50 mM Tris pH 7.5.

Reagents were pipetted as three separate spots at the base of a 5 ml tube: enzyme, substrate, and all other reagents prepared as a master mix. A 0.5 ml Eppendorf tube, with its lid removed and containing 200 μl hyamine hydroxide (Perkin Elmer), was

placed into the tube, which was sealed with a bung and transferred to a 37°C shaker incubator for typically 10-30 mins. Reactions were quenched by injection of 200 µl methanol and incubated on ice for 20 mins to allow absorption of $^{14}CO_2$ by hyamine hydroxide. Following this the exterior of the Eppendorfs were wiped, transferred to scintillation vials, 5 ml scintillant (Opti-phase "Safe", Fisher) was then added and ^{14}C activity measured in a Beckman LS6500 counter. Assays were carried out in at least duplicate. Typically wildtype proteins were used as a positive control, and uncoupled turnover assessed without substrate. In each assay counts were recorded for the master mix alone. This showed total counts for the known amount of 2OG added, allowing calculation of the number of counts per nmol 2OG used, and then of the number of nmol 2OG turned over in the course of the reaction.

FDH assay:

Assays were optimised to determine optimal reagent conditions and reaction volumes, as described in the appropriate results chapter. Reactions carried out in 100 µl total volume used a buffer of 50 mM HEPES pH 7.5, whereas those in < 100 µl total volume used a buffer of 50 mM HEPES pH 7.5, 0.01 % Tween 20.

Inhibitors were dissolved in DMSO at various concentrations, and added to the assay mixture such that the final DMSO concentration was 5 %. Inhibitors were tested for FDH inhibition in reactions containing 0.025 U FDH, 500 µM NAD, 5 mM inhibitor (typically) and 10 mM HCHO in 25 µl final volume. FDH, NAD and inhibitor were preincubated for 15 min at room temperature, prior to initiation by addition of HCHO, then incubated at 25 °C while fluorescence was recorded (355 nm excitation, 460 nm emission) at 30 s intervals over 15 min.

FDH (0.001 U/µl), NAD^+ (500 µM), ascorbate, Fe(II), enzyme (2 µM), substrate, 2OG and inhibitor were incubated together at 25 °C and monitored as above. All other reagents were used as solutions in buffer, with the exception of Fe(II) solutions, which were made using $(NH_4)_2Fe(SO_4)_2$ dissolved in 20 mM HCl to make 400 mM stock solutions, which were then diluted to the appropriate concentration using MilliQ water.

Assays with preincubation of enzyme and inhibitor were carried out using the same reagent and enzyme concentrations as above with enzyme, FDH, and inhibitor pre-incubated together for 15 min at room temperature prior to addition of the remaining reagents. Kinetic constants were determined using the same FDH-coupled assay, without the addition of inhibitors. All data were processed with GraphPad Prism 5.0™.

Differential Scanning Fluorimetry (DSF):

DSF was performed using a MiniOpticon™ Real-Time PCR Detection System (Bio-Rad). SYPRO orange (Invitrogen) dye was used to monitor unfolding of proteins occurring through thermal denaturation, due to its unspecific binding to hydrophobic residues. The corresponding changes in fluorescence were monitored as a function of time. The method used was developed from that presented in [398].

For measurements of the effect of additives on protein stability a protein concentration of 2 μM was used. Metals were dissolved as 100 mM stocks in 20 mM HCl, then further diluted in MilliQ water and added to protein at a final concentration of 50 μM. Compounds tested were prepared in 100% DMSO and added to a final concentration of 200 μM such that the final concentration of DMSO was 5 %. NOG was prepared in buffer, and added as required to a final concentration of 20 μM.

Protein	2 μM
Compound	200 μM
Mn^{2+}	50 μM
Sypro orange	1 x

To 50 μl with 50 mM HEPES pH 7.5

FAM (492 nm) and ROX (610 nm) filters were used for excitation and emission respectively. Readings of fluorescence were taken every 1 °C in the range 25-95 °C, with the temperature increased linearly by 1 °C min^{-1}. The software provided was used to perform global minimum subtraction. The inflection point, representing T_m, was calculated by fitting the Boltzmann equation (Equation 8-1) to the sigmoidal curves obtained; data were processed using GraphPad Prism 5.0™.

$$y = LL + \frac{UL - LL}{1 + e^{\frac{T_m - x}{a}}}$$

Equation 8-1: Boltzmann equation. T_m (melting temperature): Temperature at which fluorescence intensity is half maximal, interpreted as 50 % of the protein in solution being unfolded. LL (lower limit): value of minimum intensity, UL (upper limit): value of maximum intensity, α: curve gradient within T_m.

The T_m shift caused by the addition of small molecules/fragments was determined by subtraction of the "reference" T_m (protein incubated with metal and 5 % DMSO) from the T_m obtained in the presence of the compound. Conditions were tested in duplicate, with standard deviations typically < 1 °C.

Crystallisation:

Crystals of AlkB in complex with LNOC $3NO_2$, MD316, and MD246 were grown in sitting drops at 293K using vapour diffusion. The ratio of protein to reservoir solution for the AlkB:LNOC $3NO_2$ and AlkB:MD316 co-crystallisation was 2:1 (300 nL total volume) and for AlkB:MD246 was 1:1 (200 nL total volume). The AlkB protein solution contained 10.1 mg/ml protein, 50 mM HEPES pH 7.5, 2.2 mM ammonium iron (II) sulfate, and 5.7 mM LNOC $3NO_2$ or 1 mM MD316. The protein solution for the AlkB:MD246 complex contained the same but with 0.44 mM ammonium iron (II) sulfate and 1mM MD246. The reservoir solution for AlkB:LNOC $3NO_2$ contained 0.2 M NaCl, 0.1 M HEPES pH 7.5 and 25 % w/v polyethylene glycol (PEG) 3350; for AlkB:MD316, 0.2 M ammonium sulfate, 0.1 M tris-hydrochloride pH 8.5 and 25 % w/v PEG 3350; for AlkB:MD246, 0.1 M bis-tris pH 6.5 and 25 % w/v PEG 3350. The crystals were cryoprotected using well solution diluted with 25 % v/v glycerol, then flash cooled in an Oxford Cryosystems nitrogen gas stream. Data were collected from single crystals at 100 K using a Rigaku FR-E+ Superbright diffractometer equipped with a copper rotating anode, Osmic HF optics, and a Saturn 944+ CCD detector. The data were indexed, integrated, and scaled using HKL2000[399] and the structure was determined by molecular replacement using the AutoMR (PHASER)[400] subroutine in PHENIX[401] using 2FDJ (PDB ID) as a search model for AlkB:LNOC $3NO_2$ and 3T4H

(PDB ID) as a search model for AlkB:MD316 and AlkB:MD246. Iterative rounds of model building and refinement using COOT[402] and PHENIX[401] were performed until the decreasing R and R_{free} no longer converged. The final R-factors for the models were: AlkB: LNOC $3NO_2$, R = 15.9 % and R_{free}=18.7 %; AlkB: MD316, R = 16.1 % and R_{free} = 20.2 %; AlkB:MD246, R = 18.2 % and R_{free} = 22.1%.

***E. coli* Growth Curves:**

Growth of the BW25113 (wildtype) and JW2200 (AlkB deletion) *E. coli* strains was analysed following absorbance of the cultures, using a POLARStar Optima (BMG Labtech) plate reader.

Bacterial growth was carried out in either 2TY rich media or M9 minimal media (42 mM Na_2HPO_4, 22 mM KH_2PO_4, 18 mM NH_4Cl, 8 mM NaCl, 2 μM $CaCl_2$; heat sensitive compounds (500 μM $MgSO_4$, 0.2 % glucose (v/v), 2.5 μM vitamin B1) were filtered to 0.20 μM and added post autoclaving)[154,403].

Compounds were typically prepared as 1 M stocks in DMSO, then added to media at a final concentration of 10 mM (1 % DMSO (v/v) final). Bacterial cultures at OD_{600} 0.05 in media containing compounds to be tested were prepared from stationary liquid (2TY media) cultures. Cells to be grown in M9 media were harvested from 2TY stationary cultures by centrifugation, then washed twice in M9 media to remove rich media components. Assays involving addition of compounds and MMS were set up with cells preincubated with compound at 37 °C for 30 min prior to addition of MMS.

150 μl culture was added to wells of flat bottomed clear 96-well plates. Low evaporation lids were sealed on the plates using parafilm. To prevent evaporation effects the outside rows of wells were not analysed, but instead filled with blank media. Tests of each condition were carried out in biological duplicates, each analysed in triplicate. Plates were incubated at 37 °C during growth. Absorbance was measured at 584 nm, with 20 flashes per well. Readings were recorded every 15 min for 24 h, with orbital shaking (600 rpm, diameter 2 mm) of the plates for 5 min prior to each reading. Data were analysed graphically using GraphPad Prism.

Antibodies:

A polyclonal rabbit anti-recombinant mFTO antibody was commercially prepared by Eurogentec. The antibody was raised to recombinant mFTO (prepared by Dr Thomas Gerken) then affinity purified using further recombinant mFTO. Other antibodies used are detailed in Table 8-9.

Type	Supplier	Immunofluorescence	Western blot
Rabbit polyclonal anti-FTO	Eurogentec Custom made	1 in 1000	1 in 2000
Mouse monoclonal anti-FTO	Abcam ab54294	-	1 in 1000
Rabbit polyclonal anti-FTO	Novus Biologicals NB110-60935	-	1 in 1000
Mouse monoclonal anti-ABH1	Sigma A8103	1 in 200	1 in 500
Mouse monoclonal anti-His	Sigma H1029	-	1 in 1000
Mouse monoclonal anti-FLAG	Sigma F1804	1 in 1000	1 in 1000
Mouse monoclonal mixture anti-GFP	Roche 11-814-460-001	-	1 in 1000
Rat monoclonal	Chromotek 3H9	-	1 in 2000
Mouse monoclonal anti-Histone H3	Abcam ab10799	-	1 in 1000
Mouse monoclonal anti-β actin, HRP coupled	Abcam ab49900	-	1 in 20000
Goat anti-mouse HRP	Sigma A9917	-	1 in 1000
Goat anti-rabbit HRP	BioRad 172-1019	-	1 in 1000
Goat anti-rat HRP	Santa Cruz SC-2032	-	1 in 1000
Sheep anti-mouse Cy3	Jackson Immunoresearch 515-165-062	1 in 250	-
Goat anti-rabbit Cy3	GE Healthcare PA43004	1 in 500	-

Table 8-9: Antibodies used in this work.

Cell Culture:

Procedures were carried out in a Class II Biological Safety Cabinet (Nuaire Nu 437-400E). *H. sapiens* HeLa or HEK293T cells were obtained from Dr Rob Klose (Department of Biochemistry, Oxford) and grown in Dulbecco's Modified Eagle Medium (DMEM, Lonza) supplemented with glutamine, 100 U ml^{-1} penicillin,

100 μg ml^{-1} streptamycin and 10 % fetal bovine serum in a Binder BD 53 incubator with a humidified 5 % CO_2 atmosphere at 37 °C.

Vessel	Surface area /cm^2	No. cells at 100% confluence	Media volume /ml	DNA amount /μg	Transfection reagent /μl		OptiMEM /μl
					Lipofectamine	PEI	
12 well plate	3.8	3.8×10^5	1	2	2	3	150
6 well plate	9.5	9.5×10^5	2	4	4	6	250
25 cm^2 flask	25	2.5×10^6	5	12	n/a	18	450
100 mm diameter Petri dish	55	5.5×10^6	12.5	24	n/a	36	870
150 mm diameter Petri dish	148	1.48×10^7	30	50	n/a	90	2250

Table 8-10: Cell culture conditions

Cell Passaging

Semi adherent 293T cells were gently washed with PBS, then removed from the surface by resuspension in PBS by repeated pipetting. Cells in suspension were added to new plates containing fresh prewarmed medium, then allowed to grow.

Strongly adherent HeLa cells were washed in PBS, then 0.2 ml/ml trypsin (Lonza) added, and the plate incubated at 37 °C until cells were released from the plate surface. The reaction was stopped by addition of 10x fresh medium, which contains the trypsin inhibitors Mg^{2+} and Ca^{2+}. Cells were resuspended fully by pipetting to break up cell clumps, then the appropriate volume of cells was added to new plates containing fresh prewarmed medium.

Protein expression

Cells to be harvested for protein expression were prepared to 50 % confluency (as judged by a 4x/0.10 objective on a Motic AE20 (Ted Pella) microscope) then transfected if necessary using 1 mg/ml poly(ethylenimine) (PEI) pH 7.2 as a transfection reagent. Cell growth and transfection reagent conditions are shown in Table 8-10. 2 hours prior to transfection media was removed and exchanged for fresh media to initiate cell division. OptiMEM® was added to the DNA to be transfected, followed by

addition of PEI. The mixture was incubated for 15 min at room temperature before adding dropwise to the plated cells. PEI can be toxic to cells so media was exchanged 3 hours post-transfection. Cells were harvested 24 hours post-transfection in PBS either using a rubber cell scraper, or by pipetting to release the cells from the plate surface. The cell suspension was centrifuged at 1000 g, the supernatant removed carefully and the pellet either kept in ice for immediate use, or frozen at – 80 °C.

Localisation and Immunofluorescence

Cells were plated onto glass cover slips one day prior to transient transfection to a confluency of ~25 % as judged by a 4x/0.10 objective on a Motic AE20 (Ted Pella) microscope. Cells used for microscopy were transfected, if required, using the Lipofectamine™ 2000 transfection reagent (Invitrogen). Media was exchanged 2 hours prior to transfection to initiate a cycle of cell division. Lipofectamine™ and the DNA to be transfected were separately incubated with OptiMEM® serum free medium (Invitrogen) for 5 mins at room temperature, before mixing to allow liposome formation over a further 20 min incubation. The DNA mixture was added dropwise to the cells, which were then left to grow for 4 hours, allowing the DNA to be taken up. As Lipofectamine™ contains surfactants which can be toxic, media was exchanged 4 hours after transfection. A full cell cycle (20-24 hours for HeLa cells) was allowed to occur to maximise expression of the constructs.

24 hours after transfection, live cells were washed once in PBS then fixed with 3.75 % formaldehyde in PBS with shaking for 15 minutes. The formaldehyde was removed, and the cells washed twice in PBS.

GPF expressing cells were permeabilised for 2 minutes with ice cold methanol. Cells were stained for 5 minutes with 20 μg/ml 4',6-diamidino-2-phenylindole dilactate (DAPI dilactate), then mounted on glass slides using Vectashield hardset mounting medium (VECTOR laboratories).

Cells to be stained with antibodies were permeabilised in 1 % Triton-X100 for 10 minutes with shaking, then washed twice in PBS. Coverslips were incubated in

blocking buffer (10 % FBS, 0.2 % Tween 20 in PBS) for 1 hour with shaking, then incubated for 1 hour each with primary and secondary antibodies, made to the appropriate dilution in blocking buffer. After antibody incubations cells were washed 3 times in blocking buffer for 5 mins each with shaking. Coverslips were than stained with DAPI and mounted as above.

Cells prepared on slides were visualised by oil immersion, using a Zeiss Axioobserver epifluorescence microscope with a 40 x objective, with the parameters in Table 8-11.

Fluorophore	Excitation wavelength	λ(max) Absorption	λ(max) Emission
EGFP	525 nm	507 nm	488 nm
Cy3	543 nm	550 nm	570 nm
DAPI	445 nm	358 nm	461 nm

Table 8-11: Fluorescence microscopy excitation and emission parameters.

Western Blot:

Protein samples were separated on a denaturing SDS-PAGE gel, alongside either a SeeBlue2® (Invitrogen) or a PageRuler™ Prestained (Fermentas) protein molecular weight ladder. The gel was then blotted in Transfer Buffer (2.5 mM Tris, 192 mM glycine, 20 % MeOH) using a Mini Trans-Blot® Electrophoretic Transfer Cell (Bio-Rad) system, onto Hybond ECL membrane (GE Healthcare), run on ice at 200 mA for 2-3 hours. Protein transfer was verified by staining the membrane with Ponceau S dye, and was then destained in MilliQ water. Membranes were blocked in Blocking Buffer (5 % milk in 1 x Tris Buffered Saline 0.1 % Tween (TBST; 25 mM Tris, 137 mM NaCl, 2.68 mM KCl, 0.1 % Tween 20; pH 8.8)) on a rotating platform either overnight at 4 ^{o}C or for 1 hour at room temperature. Membranes were probed with primary and secondary antibodies, prepared to the appropriate dilution in 1 ml Blocking Buffer, for 1 hour each at room temperature. After each antibody incubation the membrane was washed extensively in 1 x TBST Wash Buffer (3 x 10 min incubations with shaking). To detect signals ECL™ Detection Reagents (GE Healthcare) were used, with 1 ml each of both reagents 1 and 2 per gel blot. Blots were incubated with detection reagent for 1 min,

then immediately exposed to Amersham™ Hyperfilm ECL film (GE Healthcare) and developed using a Compact X4 (Xograph) film developer.

Cell lysis and Immunoprecipitation:

Various cell lysis and immunoprecipitation methods were attempted using cultured HeLa and 293T cells. The general method involved the following steps, with precise details given in the relevant results chapters.

Cells were thawed at room temperature or 37 °C, then resuspended in 25 µl (large scale 1 ml) lysis buffer (typically 10 mM tris pH 7.5, 0.5 % NP40 (Roche), 300 mM NaCl) and incubated on ice for 30 min. If required, cells were sonicated for 3 x 10 s using a MSE Soniprep 150 sonicator at an amplitude of 10 µm. Insoluble debris was pelleted at 14 kg for 1 min. Beads used for IP were GFP-TRAP® (Chromotek), Protein G Sepharose 4 Fast Flow (GE Healthcare) or Anti-FLAG® M2 affinity gel (Sigma A2220). 20 µl (large scale 100 µl) beads were equilibrated by 3 washes with 5 x bead volume lysis buffer lacking NP40, followed each time by centrifugation at 2 kg and 4 °C for 2 min.

Cell lysate was applied to beads and rotated end-over-end at 4 °C for 1 hour. Protein bound beads were washed 3 times with 5 x bead volume lysis buffer lacking NP40, followed each time by centrifugation at 2 kg and 4 °C for 2 min. For a final wash beads were transferred to a fresh tube to remove any proteins adsorbed to the plastic tube surface. Proteins were either eluted by addition of SDS-PAGE loading buffer, incubation at 65 °C, or by addition of 3xFLAG peptide.

Extraction of nuclear proteins from mouse tissue samples was carried out using the CelLytic™ NuClear® kit (Sigma). 100 mg muscle tissue was used, and protein extracted, quantified by Bradford assay and verified by western blot. Fractions of protein samples at each stage of the protocol were retained for analysis by western blot.

In-Gel Trypsin Digestion:

Following complete destaining of the SDS-PAGE gel in MilliQ water stained gel slices to be analysed were excised, cut into ~1 mm^2 pieces and transferred to a sterile 1.5 ml

tube. Gel pieces were washed twice, with shaking, in 200 µl wash solution (50 % MeOH, 5 % AcOH). Liquid was removed, then the gel pieces dried through addition of 200 µl MeCN. Cysteine residues were reduced to thiols, and selectively alkylated with iodoacetamide to reduce their nucleophilicity. Gel pieces were swelled in 30 µl 10 mM DTT (in NH_4HCO_3) and incubated for 30 mins at room temperature. Excess liquid was removed and 30 µl 100 mM iodoacetamide (in NH_4HCO_3) added, and incubated at room temperature for 30 mins. All liquid was removed, then gel pieces were dehydrated by addition of 200 µl MeCN and incubation with shaking for 5 min. Gel pieces were again rehydrated in 200 µl 100 mM NH_4HCO_3, then excess liquid was removed, and gel pieces dried again with addition of 200 µl MeCN. Remaining liquid was removed. Gel pieces were rehydrated in 30 µl trypsin (20 µg ml^{-1} in 50 mM NH_4HCO_3) and protein digestion was allowed to proceed for 16 hours at 37 °C.

Following incubation, the supernatant contained trypsin and the digested peptides. These peptides were in equilibrium with those still remaining in the gel, which must be extracted. 30 µl additional 100 mM NH_4HCO_3 was added to the digest and the mixture incubated with shaking for 10 min. The supernatant was transferred to a clean tube and 30 µl extraction buffer 1 (50 % MeCN, 5 % HCOOH) was added to the gel pieces and incubated for 10 min at room temperature with shaking. The supernatant was retained and added to the initial extracted supernatant. 30 µl extraction buffer 2 (80 % MeCN, 5 % HCOOH) was added to gel pieces incubated for 10 min at room temperature with shaking. All supernatants were pooled and dried by vacuum centrifugation, carried out at 55 °C. The resulting peptides were redissolved in ~20 µl 2 % MeCN, 0.1% formic acid and transferred to mass spectrometry vials prior to mass spectrometric analysis.

LC-MS/MS analysis

For the analysis of in-gel digested protein material, liquid chromatography was performed using an Ultimate 3000 nano-HPLC system (Dionex, Sunnyvale, CA, USA) comprising a WPS-3000 micro auto sampler, a FLM-3000 flow manager and column compartment, a UVD-3000 UV detector, an LPG-3600 dual-gradient micro-pump, and

an SRD-3600 solvent rack controlled by Hystar (Bruker Daltonics, Billerica, MA, USA) and DCMS link 2.0 software (Dionex). Samples were concentrated on a trapping column (Dionex), 300 µm internal diameter (i.d.), 0.1 cm) at a flow rate of 20 µl min^{-1}. For separation with a C18 Pepmap column (75 µm i.d., 15 cm, Dionex), a flow rate of 250 nl min^{-1} was used as generated by a cap-flow splitter cartridge (1/1000). Peptides were eluted by the application of a ~ 30 min multi-step gradient (Table 8-12) using solvents A (98 % water, 2 % acetonitrile, 0.1 % formic acid) and B (80 % acetonitrile, 20 % water, 0.1 % formic acid).

Run time /min	Solvent composition /% B
0	2
3	10
18	25
30	50
30.5	90

Table 8-12: Liquid chromatography gradient used for analysis of in-gel digested protein material.

The liquid chromatography was interfaced directly with a 3D high capacity ion trap mass spectrometer (amaZon; Bruker Daltonics) utilizing 10 µm i.d. distal coated SilicaTips (New Objective, Woburn, MA, USA) and nano-ESI mode. SPS parameter settings on the ion trap were tuned for a target mass of 850 m/z, compound stability 100 % and a smart ICC target of 250000. MS/MS analysis was initiated on a contact closure signal triggered by HyStar software (version 3.2). Up to four precursor ions were selected per cycle with active exclusion (0.5 min) in collision-induced dissociation (CID) mode. CID fragmentation was achieved using helium gas and a 30 %-200 % collision energy sweep with amplitude 1.0 (ions are ejected from the trap as soon as they fragment).

Data processing and database searching:

Raw LC-MS/MS data were processed, and MASCOT compatible files created, using DataAnalysis 4.0 software (Bruker Daltonics) with the following parameters: compounds (autoMS) threshold 10000, number of compounds 500, retention time windows 0.5 min for C18 (30 min gradient). Database searches were performed using

the MASCOT software (version 2.3) and the SwissProt database (v2011.02.11, number of entries 525 207) with the following parameters: 1+, 2+ and 3+ ions, peptide tolerance 0.3 Da, ^{13}C = 2, fragment tolerance 0.6 Da, missed cleavages: 2, instrument type: ESI-TRAP.

Non-denaturing Mass Spectrometry:

ESI-MS analyses were carried out by Dr Jasmin Mecinović (FTO), Dr Esther Woon, or Marina Demetriades (AlkB) on a Waters Micromass Q-ToF spectrometer and an Advion Biosciences NanoMate™ HD Robot chip-based nano-electrospray device. Protein samples were desalted into 15 mM ammonium acetate, pH 7.5. 15 μM protein was mixed with 75 μM Fe(II), and 15 μM inhibitor as required, and incubated for 30 min at room temperature prior to ESI-MS analysis.

Data were acquired with a chip voltage of 1.70 kV and a delivery pressure 0.25 psi. The sample cone voltage was typically 80 V with a source temperature of 40 °C. Calibration and sample acquisition were performed in positive ion mode in the range of 500-5000 m/z. The pressure at the interface between the atmospheric source and the high vacuum region was fixed at 6.60 mbar. External instrument calibration was achieved using sodium iodide. Data were processed with MassLynx 4.0 (Waters). FTO samples were additionally run on a Waters Synapt™ HDMS™, sprayed in 25 mM ammonium acetate, pH 7.0, at 90 V.

References

1. Clifton, I.J., *et al.* Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J Inorg Biochem* **100**, 644-669 (2006).
2. Loenarz, C. & Schofield, C.J. Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* **36**, 7-18 (2011).
3. Flashman, E. & Schofield, C.J. The most versatile of all reactive intermediates? *Nat Chem Biol* **3**, 86-87 (2007).
4. Roach, P.L., *et al.* Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature* **375**, 700-704 (1995).
5. Hausinger, R.P. FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol* **39**, 21-68 (2004).
6. Hegg, E.L. & Que, L., Jr. The 2-His-1-carboxylate facial triad--an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem* **250**, 625-629 (1997).
7. Mehn, M.P., Fujisawa, K., Hegg, E.L. & Que, L., Jr. Oxygen activation by nonheme iron(II) complexes: alpha-keto carboxylate versus carboxylate. *J Am Chem Soc* **125**, 7828-7842 (2003).
8. Welford, R.W., *et al.* Incorporation of oxygen into the succinate co-product of iron(II) and 2-oxoglutarate dependent oxygenases from bacteria, plants and humans. *FEBS Lett* **579**, 5170-5174 (2005).
9. Hutton, J.J., Jr., Kaplan, A. & Udenfriend, S. Conversion of the amino acid sequence gly-pro-pro in protein to gly-pro-hyp by collagen proline hydroxylase. *Arch Biochem Biophys* **121**, 384-391 (1967).
10. Risteli, J., Tryggvason, K. & Kivirikko, K.I. Prolyl 3-hydroxylase: partial characterization of the enzyme from rat kidney cortex. *Eur J Biochem* **73**, 485-492 (1977).
11. Prockop, D.J., Weinstein, E. & Mulveny, T. Hydroxylation of lysine in a polypeptide precursor of collagen. *Biochem Biophys Res Commun* **22**, 124-128 (1966).
12. Jenkins, C.L., Bretscher, L.E., Guzei, I.A. & Raines, R.T. Effect of 3-hydroxyproline residues on collagen stability. *J Am Chem Soc* **125**, 6422-6427 (2003).
13. Schofield, C.J. & Ratcliffe, P.J. Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Biol* **5**, 343-354 (2004).
14. Mahon, P.C., Hirota, K. & Semenza, G.L. FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. *Genes Dev* **15**, 2675-2686 (2001).
15. Epstein, A.C.R., *et al.* C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107**, 43-54 (2001).
16. Bruick, R.K. & McKnight, S.L. A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294**, 1337-1340 (2001).
17. McNeill, L.A., *et al.* Hypoxia-inducible factor asparaginyl hydroxylase (FIH-1) catalyses hydroxylation at the β-carbon of asparagine-803. *Biochem J* **367**, 571-575 (2002).
18. Lando, D., Peet, D.J., Whelan, D.A., Gorman, J.J. & Whitelaw, M.L. Asparagine hydroxylation of the HIF transactivation domain: A hypoxic switch. *Science* **295**, 858-861 (2002).
19. Masson, N., Willam, C., Maxwell, P.H., Pugh, C.W. & Ratcliffe, P.J. Independent function of two destruction domains in hypoxia-inducible factor-α chains activated by prolyl hydroxylation. *EMBO Journal* **20**, 5197-5206 (2001).
20. Jaakkola, P., *et al.* Targeting of HIF-α to the von Hippel-Lindau ubiquitylation complex by O_2-regulated prolyl hydroxylation. *Science* **292**, 468-472 (2001).
21. Huang, L.E., Gu, J., Schau, M. & Bunn, H.F. Regulation of hypoxia-inducible factor 1α is mediated by an O_2-dependent degradation domain via the ubiquitin-proteasome pathway. *Proc Natl Acad Sci U S A* **95**, 7987-7992 (1998).

22. Hon, W.-C., *et al.* Structural basis for the recognition of hydroxyproline in HIF-1a by pVHL. *Nature* **417**, 975-978 (2002).
23. Maxwell, P.H., *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271-275 (1999).
24. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
25. Tsukada, Y., *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811-816 (2006).
26. Cloos, P.A., *et al.* The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* **442**, 307-311 (2006).
27. Horton, J.R., *et al.* Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases. *Nat Struct Mol Biol* **17**, 38-43 (2010).
28. Webby, C.J., *et al.* Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90-93 (2009).
29. Mantri, M., *et al.* Crystal structure of the 2-oxoglutarate- and Fe(II)-dependent lysyl hydroxylase JMJD6. *J Mol Biol* **401**, 211-222 (2010).
30. Hong, X., *et al.* Interaction of JMJD6 with single-stranded RNA. *Proc Natl Acad Sci U S A* **107**, 14568-14572 (2010).
31. Vaz, F.M. & Wanders, R.J. Carnitine biosynthesis in mammals. *Biochem J* **361**, 417-429 (2002).
32. Leung, I.K., *et al.* Structural and mechanistic studies on gamma-butyrobetaine hydroxylase. *Chem Biol* **17**, 1316-1324 (2010).
33. Sjakste, N., Gutcaits, A. & Kalvinsh, I. Mildronate: an antiischemic drug for neurological indications. *CNS Drug Rev* **11**, 151-168 (2005).
34. Kershaw, N.J., *et al.* Studies on phytanoyl-CoA 2-hydroxylase and synthesis of phytanoyl-coenzyme A. *Bioorg Med Chem Lett* **11**, 2545-2548 (2001).
35. Wierzbicki, A.S., Lloyd, M.D., Schofield, C.J., Feher, M.D. & Gibberd, F.B. Refsum's disease: a peroxisomal disorder affecting phytanic acid alpha-oxidation. *J Neurochem* **80**, 727-735 (2002).
36. McDonough, M.A., *et al.* Structure of human phytanoyl-CoA 2-hydroxylase identifies molecular mechanisms of Refsum disease. *J Biol Chem* **280**, 41101-41110 (2005).
37. Abbott, M.T., Schandl, E.K., Lee, R.F., Parker, T.S. & Midgett, R.J. Cofactor requirements of thymine 7-hydroxylase. *Biochim Biophys Acta* **132**, 525-528 (1967).
38. Liu, C.K., Hsu, C.A. & Abbott, M.T. Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. *Arch Biochem Biophys* **159**, 180-187 (1973).
39. Thornburg, L.D., Lai, M.T., Wishnok, J.S. & Stubbe, J. A non-heme iron protein with heme tendencies: an investigation of the substrate specificity of thymine hydroxylase. *Biochemistry* **32**, 14023-14033 (1993).
40. Cultrone, A., *et al.* Convergent evolution of hydroxylation mechanisms in the fungal kingdom: molybdenum cofactor-independent hydroxylation of xanthine via alpha-ketoglutarate-dependent dioxygenases. *Mol Microbiol* **57**, 276-290 (2005).
41. Gommers-Ampt, J.H., *et al.* β-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan *T. brucei*. *Cell* **75**, 1129-1136 (1993).
42. Yu, Z., *et al.* The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res* **35**, 2107-2115 (2007).
43. MedlinePlus. Sleeping Sickness. (National Institutes of Health, 2011).
44. Horn, D. & Barry, J.D. The central roles of telomeres and subtelomeres in antigenic variation in African trypanosomes. *Chromosome Res* **13**, 525-533 (2005).
45. Barry, J.D. & McCulloch, R. Antigenic variation in trypanosomes: enhanced phenotypic variation in a eukaryotic parasite. *Adv Parasitol* **49**, 1-70 (2001).
46. van Leeuwen, F., *et al.* Localization of the modified base J in telomeric VSG gene expression sites of *Trypanosoma brucei*. *Genes Dev* **11**, 3232-3241 (1997).
47. van Leeuwen, F., Kieft, R., Cross, M. & Borst, P. Biosynthesis and function of the modified DNA base beta-D-glucosyl-hydroxymethyluracil in *Trypanosoma brucei*. *Mol Cell Biol* **18**, 5643-5651 (1998).

48. Iyer, L.M., Abhiman, S., de Souza, R.F. & Aravind, L. Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. *Nucleic Acids Res* **38**, 5261-5279 (2010).
49. Noma, A., *et al.* Expanding role of the jumonji C domain as an RNA hydroxylase. *J Biol Chem* **285**, 34503-34507 (2010).
50. Delhommeau, F., *et al.* Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-2301 (2009).
51. Tahiliani, M., *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-935 (2009).
52. Ito, S., *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129-1133 (2010).
53. Ito, S., *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* (2011).
54. Ko, M., *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839-843 (2010).
55. van der Ploeg, J.R., *et al.* Identification of sulfate starvation-regulated genes in *Escherichia coli*: a gene cluster involved in the utilization of taurine as a sulfur source. *J Bacteriol* **178**, 5438-5446 (1996).
56. Eichhorn, E., van der Ploeg, J.R., Kertesz, M.A. & Leisinger, T. Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*. *J Biol Chem* **272**, 23031-23036 (1997).
57. Marschall, C., *et al.* Molecular analysis of the regulation of csiD, a carbon starvation-inducible gene in *Escherichia coli* that is exclusively dependent on sigma s and requires activation by cAMP-CRP. *J Mol Biol* **276**, 339-353 (1998).
58. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).
59. Hollaender, A. & Claus, W.D. The Bactericidal Effect of Ultraviolet Radiation on *Escherichia Coli* in Liquid Suspensions. *J Gen Physiol* **19**, 753-765 (1936).
60. Drew, H.R., *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc Natl Acad Sci U S A* **78**, 2179-2183 (1981).
61. Itoh, Y., Chiba, S., Sekine, S. & Yokoyama, S. Crystal structure of human selenocysteine tRNA. *Nucleic Acids Res* **37**, 6259-6268 (2009).
62. Lodish, H., *et al. Molecular Cell Biology, 4th Edition*, (W. H. Freeman & Co., 2000).
63. Barta, A., Dorner, S. & Polacek, N. Mechanism of ribosomal peptide bond formation. *Science* **291**, 203 (2001).
64. Wu, L. & Belasco, J.G. Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Mol Cell* **29**, 1-7 (2008).
65. Bernstein, J.A., Khodursky, A.B., Lin, P.H., Lin-Chao, S. & Cohen, S.N. Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays. *Proc Natl Acad Sci U S A* **99**, 9697-9702 (2002).
66. Hargrove, J.L. & Schmidt, F.H. The role of mRNA and protein stability in gene expression. *Faseb J* **3**, 2360-2370 (1989).
67. Cheng, X. & Roberts, R.J. AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res* **29**, 3784-3795 (2001).
68. Marinus, M.G. & Casadesus, J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* **33**, 488-503 (2009).
69. Ratel, D., Ravanat, J.L., Berger, F. & Wion, D. N6-methyladenine: the other methylated base of DNA. *Bioessays* **28**, 309-315 (2006).
70. Wion, D. & Casadesus, J. N6-methyladenine: an epigenetic signal for DNA-protein interactions. *Nat Rev Microbiol* **4**, 183-192 (2006).
71. Cokus, S.J., *et al.* Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215-219 (2008).

72. Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nat Rev Genet* **10**, 295-304 (2009).
73. Pomraning, K.R., Smith, K.M. & Freitag, M. Genome-wide high throughput analysis of DNA methylation in eukaryotes. *Methods* **47**, 142-150 (2009).
74. Wolffe, A.P. & Matzke, M.A. Epigenetics: regulation through repression. *Science* **286**, 481-486 (1999).
75. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-254 (2003).
76. Lister, R., *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322 (2009).
77. Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat Rev Mol Cell Biol* **11**, 607-620 (2010).
78. Pogribny, I.P. & Beland, F.A. DNA hypomethylation in the origin and pathogenesis of human diseases. *Cell Mol Life Sci* **66**, 2249-2261 (2009).
79. Penn, N.W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J* **126**, 781-790 (1972).
80. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929-930 (2009).
81. Li, W. & Liu, M. Distribution of 5-hydroxymethylcytosine in different human tissues. *J Nucleic Acids* **2011**, 870726 (2011).
82. Cantara, W.A., *et al.* The RNA Modification Database, RNAMDB: 2011 update. *Nucleic Acids Res* **39**, D195-201 (2011).
83. Furuichi, Y. & Shatkin, A.J. Viral and cellular mRNA capping: past and prospects. *Adv Virus Res* **55**, 135-184 (2000).
84. Daffis, S., *et al.* 2'-O methylation of the viral mRNA cap evades host restriction by IFIT family members. *Nature* **468**, 452-456 (2010).
85. Kirino, Y. & Suzuki, T. Human mitochondrial diseases associated with tRNA wobble modification deficiency. *RNA Biol* **2**, 41-44 (2005).
86. Noma, A., Kirino, Y., Ikeuchi, Y. & Suzuki, T. Biosynthesis of wybutosine, a hyper-modified nucleoside in eukaryotic phenylalanine tRNA. *Embo J* **25**, 2142-2154 (2006).
87. Anderson, J., Phan, L. & Hinnebusch, A.G. The Gcd10p/Gcd14p complex is the essential two-subunit tRNA(1-methyladenosine) methyltransferase of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **97**, 5173-5178 (2000).
88. Helm, M., Giege, R. & Florentz, C. A Watson-Crick base-pair-disrupting methyl group (m1A9) is sufficient for cloverleaf folding of human mitochondrial tRNALys. *Biochemistry* **38**, 13338-13346 (1999).
89. Takeuchi, Y., Tazawa, I. & Inoue, Y. Intramolecular stacking association of three dinucleoside monophosphates containing naturally-occurring 1-methyladenosine residue(s): m1ApA, Apm1A and m1Apm1A. *Bull. Chem. Soc. Jpn* **55**, 3598-3602 (1982).
90. Yang, H., Zhan, Y., Fenn, D., Chi, L.M. & Lam, S.L. Effect of 1-methyladenine on double-helical DNA structures. *FEBS Lett* **582**, 1629-1633 (2008).
91. Sierzputowska-Gracz, H., Gopal, H.D. & Agris, P.F. Comparative structural analysis of 1-methyladenosine, 7-methylguanosine, ethenoadenosine and their protonated salts IV: 1H, 13C, and 15N NMR studies at natural isotope abundance. *Nucleic Acids Res* **14**, 7783-7801 (1986).
92. Rydberg, B. & Lindahl, T. Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *Embo J* **1**, 211-216 (1982).
93. Lindahl, T. & Nyberg, B. Rate of depurination of native deoxyribonucleic acid. *Biochemistry* **11**, 3610-3618 (1972).
94. Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715 (1993).

95. Nair, J., Barbin, A., Velic, I. & Bartsch, H. Etheno DNA-base adducts from endogenous reactive species. *Mutat Res* **424**, 59-69 (1999).
96. Palejwala, V.A., Rzepka, R.W., Simha, D. & Humayun, M.Z. Quantitative multiplex sequence analysis of mutational hot spots. Frequency and specificity of mutations induced by a site-specific ethenocytosine in M13 viral DNA. *Biochemistry* **32**, 4105-4111 (1993).
97. Wyatt, M.D. & Pittman, D.L. Methylating agents and DNA repair responses: Methylated bases and sources of strand breaks. *Chem Res Toxicol* **19**, 1580-1594 (2006).
98. Hang, B. Formation and repair of tobacco carcinogen-derived bulky DNA adducts. *J Nucleic Acids* **2010**, 709521 (2010).
99. Sedgwick, B. Repairing DNA-methylation damage. *Nat Rev Mol Cell Biol* **5**, 148-157 (2004).
100. Pegg, A.E. Methylation of the O6 position of guanine in DNA is the most likely initiating event in carcinogenesis by methylating agents. *Cancer Invest* **2**, 223-231 (1984).
101. Osborne, M.R. & Phillips, D.H. Preparation of a methylated DNA standard, and its stability on storage. *Chem Res Toxicol* **13**, 257-261 (2000).
102. Loechler, E.L., Green, C.L. & Essigmann, J.M. In vivo mutagenesis by O6-methylguanine built into a unique site in a viral genome. *Proc Natl Acad Sci U S A* **81**, 6271-6275 (1984).
103. Hurley, L.H. DNA and its associated processes as targets for cancer therapy. *Nat Rev Cancer* **2**, 188-200 (2002).
104. Altshuler, K.B., Hodes, C.S. & Essigmann, J.M. Intrachromosomal probes for mutagenesis by alkylated DNA bases replicated in mammalian cells: a comparison of the mutagenicities of O4-methylthymine and O6-methylguanine in cells with different DNA repair backgrounds. *Chem Res Toxicol* **9**, 980-987 (1996).
105. Doublie, S., Tabor, S., Long, A.M., Richardson, C.C. & Ellenberger, T. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* **391**, 251-258 (1998).
106. Delaney, J.C. & Essigmann, J.M. Mutagenesis, genotoxicity, and repair of 1-methyladenine, 3-alkylcytosines, 1-methylguanine, and 3-methylthymine in alkB *Escherichia coli*. *Proc Natl Acad Sci U S A* **101**, 14051-14056 (2004).
107. Shooter, K.V., Howse, R., Shah, S.A. & Lawley, P.D. The molecular basis for biological inactivation of nucleic acids. The action of methylating agents on the ribonucleic acid-containing bacteriophage R17. *Biochem J* **137**, 303-312 (1974).
108. Jackson, S.P. Sensing and repairing DNA double-strand breaks. *Carcinogenesis* **23**, 687-696 (2002).
109. Fugmann, S.D., Lee, A.I., Shockett, P.E., Villey, I.J. & Schatz, D.G. The RAG proteins and V(D)J recombination: complexes, ends, and transposition. *Annu Rev Immunol* **18**, 495-527 (2000).
110. Kaina, B. Mechanisms and consequences of methylating agent-induced SCEs and chromosomal aberrations: a long road traveled and still a far way to go. *Cytogenet Genome Res* **104**, 77-86 (2004).
111. Engelward, B.P., *et al.* A chemical and genetic approach together define the biological consequences of 3-methyladenine lesions in the mammalian genome. *J Biol Chem* **273**, 5412-5418 (1998).
112. Darst, R.P., Pardo, C.E., Ai, L., Brown, K.D. & Kladde, M.P. Bisulfite sequencing of DNA. *Curr Protoc Mol Biol* **Chapter 7**, Unit 7 9 1-17 (2010).
113. Middleton, M.R. & Margison, G.P. Improvement of chemotherapy efficacy by inactivation of a DNA-repair pathway. *Lancet Oncol* **4**, 37-44 (2003).
114. Memisoglu, A. & Samson, L. Base excision repair in yeast and mammals. *Mutat Res* **451**, 39-51 (2000).
115. Fortini, P., *et al.* The base excision repair: mechanisms and its relevance for cancer susceptibility. *Biochimie* **85**, 1053-1071 (2003).

116. Lindahl, T., Sedgwick, B., Sekiguchi, M. & Nakabeppu, Y. Regulation and expression of the adaptive response to alkylating agents. *Annu Rev Biochem* **57**, 133-157 (1988).
117. Asaeda, A., *et al.* Substrate specificity of human methylpurine DNA *N*-glycosylase. *Biochemistry* **39**, 1959-1965 (2000).
118. Dizdaroglu, M. Substrate specificities and excision kinetics of DNA glycosylases involved in base-excision repair of oxidative DNA damage. *Mutat Res* **531**, 109-126 (2003).
119. Scharer, O.D. Chemistry and biology of DNA repair. *Angew Chem Int Ed Engl* **42**, 2946-2974 (2003).
120. Costa, R.M., Chigancas, V., Galhardo Rda, S., Carvalho, H. & Menck, C.F. The eukaryotic nucleotide excision repair pathway. *Biochimie* **85**, 1083-1099 (2003).
121. Schofield, M.J. & Hsieh, P. DNA mismatch repair: molecular mechanisms and biological function. *Annu Rev Microbiol* **57**, 579-608 (2003).
122. Kunkel, T.A., Pavlov, Y.I. & Bebenek, K. Functions of human DNA polymerases eta, kappa and iota suggested by their properties, including fidelity with undamaged DNA templates. *DNA Repair (Amst)* **2**, 135-149 (2003).
123. Prakash, S. & Prakash, L. Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. *Genes Dev* **16**, 1872-1883 (2002).
124. Helleday, T. Pathways for mitotic homologous recombination in mammalian cells. *Mutat Res* **532**, 103-115 (2003).
125. Lees-Miller, S.P. & Meek, K. Repair of DNA double strand breaks by non-homologous end joining. *Biochimie* **85**, 1161-1173 (2003).
126. Sancar, G.B. Enzymatic photoreactivation: 50 years and counting. *Mutat Res* **451**, 25-37 (2000).
127. Sedgwick, B., Bates, P.A., Paik, J., Jacobs, S.C. & Lindahl, T. Repair of alkylated DNA: recent advances. *DNA Repair (Amst)* **6**, 429-442 (2007).
128. Nakabeppu, Y., Kajitani, K., Sakamoto, K., Yamaguchi, H. & Tsuchimoto, D. MTH1, an oxidized purine nucleoside triphosphatase, prevents the cytotoxicity and neurotoxicity of oxidized purine nucleotides. *DNA Repair (Amst)* **5**, 761-772 (2006).
129. Mo, J.Y., Maki, H. & Sekiguchi, M. Hydrolytic elimination of a mutagenic nucleotide, 8-oxodGTP, by human 18-kilodalton protein: sanitization of nucleotide pool. *Proc Natl Acad Sci U S A* **89**, 11021-11025 (1992).
130. Anderson, S., *et al.* Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457-465 (1981).
131. Shen, J., *et al.* Oxygen consumption rates and oxygen concentration in molt-4 cells and their mtDNA depleted (rho0) mutants. *Biophys J* **84**, 1291-1298 (2003).
132. Kujoth, G.C., *et al.* Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging. *Science* **309**, 481-484 (2005).
133. Stuart, J.A., Mayard, S., Hashiguchi, K., Souza-Pinto, N.C. & Bohr, V.A. Localization of mitochondrial DNA base excision repair to an inner membrane-associated particulate fraction. *Nucleic Acids Res* **33**, 3722-3732 (2005).
134. Myers, K.A., Saffhill, R. & O'Connor, P.J. Repair of alkylated purines in the hepatic DNA of mitochondria and nuclei in the rat. *Carcinogenesis* **9**, 285-292 (1988).
135. St-Pierre, J., Buckingham, J.A., Roebuck, S.J. & Brand, M.D. Topology of superoxide production from different sites in the mitochondrial electron transport chain. *J Biol Chem* **277**, 44784-44790 (2002).
136. Nishioka, K., *et al.* Expression and differential intracellular localization of two major forms of human 8-oxoguanine DNA glycosylase encoded by alternatively spliced OGG1 mRNAs. *Mol Biol Cell* **10**, 1637-1652 (1999).
137. Miller, F.J., Rosenfeldt, F.L., Zhang, C., Linnane, A.W. & Nagley, P. Precise determination of mitochondrial DNA copy number in human skeletal and cardiac muscle by a PCR-based assay: lack of change of copy number with age. *Nucleic Acids Res* **31**, e61 (2003).
138. Yoshida, Y., *et al.* Human mitochondrial transcription factor A binds preferentially to oxidatively damaged DNA. *Biochem Biophys Res Commun* **295**, 945-951 (2002).

139. Fukuoh, A. & Kang, D. Methods for assessing binding of mitochondrial transcription factor A (TFAM) to DNA. *Methods Mol Biol* **554**, 87-101 (2009).
140. Sedgwick, B. & Lindahl, T. Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* **21**, 8886-8894 (2002).
141. Kondo, H., *et al.* Structure and expression of the alkB gene of *Escherichia coli* related to the repair of alkylated DNA. *J Biol Chem* **261**, 15772-15777 (1986).
142. Potter, P.M., Kleibl, K., Cawkwell, L. & Margison, G.P. Expression of the ogt gene in wild-type and ada mutants of *E. coli*. *Nucleic Acids Res* **17**, 8047-8060 (1989).
143. Demple, B., *et al.* Active site and complete sequence of the suicidal methyltransferase that counters alkylation mutagenesis. *Proc Natl Acad Sci U S A* **82**, 2688-2692 (1985).
144. Landini, P. & Volkert, M.R. Regulatory responses of the adaptive response to alkylation damage: a simple regulon with complex regulatory features. *J Bacteriol* **182**, 6543-6549 (2000).
145. Saget, B.M. & Walker, G.C. The Ada protein acts as both a positive and a negative modulator of *Escherichia coli*'s response to methylating agents. *Proc Natl Acad Sci U S A* **91**, 9730-9734 (1994).
146. Gasparutto, D., Dherin, C., Boiteux, S. & Cadet, J. Excision of 8-methylguanine site-specifically incorporated into oligonucleotide substrates by the AlkA protein of *Escherichia coli*. *DNA Repair (Amst)* **1**, 437-447 (2002).
147. Labahn, J., *et al.* Structural basis for the excision repair of alkylation-damaged DNA. *Cell* **86**, 321-329 (1996).
148. Berdal, K.G., Johansen, R.F. & Seeberg, E. Release of normal bases from intact DNA by a native DNA repair enzyme. *Embo J* **17**, 363-367 (1998).
149. Rippa, V., *et al.* Specific DNA binding and regulation of its own expression by the AidB protein in *Escherichia coli*. *J Bacteriol* **192**, 6136-6142 (2010).
150. Rippa, V., *et al.* Preferential DNA damage prevention by the *E. coli* AidB gene: A new mechanism for the protection of specific genes. *DNA Repair (Amst)* (2011).
151. Bowles, T., Metz, A.H., O'Quin, J., Wawrzak, Z. & Eichman, B.F. Structure and DNA binding of alkylation response protein AidB. *Proc Natl Acad Sci U S A* **105**, 15299-15304 (2008).
152. Rohankhedkar, M.S., Mulrooney, S.B., Wedemeyer, W.J. & Hausinger, R.P. The AidB component of the *Escherichia coli* adaptive response to alkylating agents is a flavin-containing, DNA-binding protein. *J Bacteriol* **188**, 223-230 (2006).
153. Mulrooney, S.B., Howard, M.J. & Hausinger, R.P. The *Escherichia coli* alkylation response protein AidB is a redox partner of flavodoxin and binds RNA and acyl carrier protein. *Arch Biochem Biophys* (2011).
154. Kataoka, H., Yamamoto, Y. & Sekiguchi, M. A new gene (*alkB*) of *Escherichia coli* that controls sensitivity to methyl methane sulfonate. *J Bacteriol* **153**, 1301-1307 (1983).
155. Kataoka, H. & Sekiguchi, M. Molecular cloning and characterization of the alkB gene of Escherichia coli. *Mol Gen Genet* **198**, 263-269 (1985).
156. Dinglay, S., Trewick, S.C., Lindahl, T. & Sedgwick, B. Defective processing of methylated single-stranded DNA by *E. coli* AlkB mutants. *Genes Dev* **14**, 2097-2105 (2000).
157. Aravind, L. & Koonin, E.V. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol* **2**, research0007.0001-0007.0008 (2001).
158. Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174-178 (2002).
159. Falnes, P.O., Johansen, R.F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in *Escherichia coli*. *Nature* **419**, 178-182 (2002).
160. van den Born, E., *et al.* Bioinformatics and functional analysis define four distinct groups of AlkB DNA-dioxygenases in bacteria. *Nucleic Acids Res* **37**, 7124-7136 (2009).

161. Leiros, I., *et al.* Structural basis for enzymatic excision of N1-methyladenine and N3-methylcytosine from DNA. *Embo J* **26**, 2206-2217 (2007).
162. Wei, Y.F., Carter, K.C., Wang, R.P. & Shell, B.K. 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 (1996).
163. Duncan, T., *et al.* Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A* **99**, 16660-16665 (2002).
164. Kurowski, M.A., Bhagwat, A.S., Papaj, G. & Bujnicki, J.M. Phylogenomic identification of five new human homologs of the DNA repair enzyme AlkB. *BMC Genomics* **4**, 48 (2003).
165. Gerken, T., *et al.* The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469-1472 (2007).
166. Sanchez-Pulido, L. & Andrade-Navarro, M.A. The FTO (fat mass and obesity associated) gene codes for a novel member of the non-heme dioxygenase superfamily. *BMC Biochem* **8**, 23 (2007).
167. Aas, P.A., *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**, 859-863 (2003).
168. Gilljam, K.M., *et al.* Identification of a novel, widespread, and functionally important PCNA-binding motif. *J Cell Biol* **186**, 645-654 (2009).
169. Cetica, V., *et al.* Pediatric brain tumors: mutations of two dioxygenases (hABH2 and hABH3) that directly repair alkylation damage. *J Neurooncol* **94**, 195-201 (2009).
170. Lee, S.Y., *et al.* TP53 regulates human AlkB homologue 2 expression in glioma resistance to Photofrin-mediated photodynamic therapy. *Br J Cancer* **103**, 362-369 (2010).
171. Wu, S.S., *et al.* Down-regulation of ALKBH2 increases cisplatin sensitivity in H1299 lung cancer cells. *Acta Pharmacol Sin* (2011).
172. Konishi, N., *et al.* High expression of a new marker PCA-1 in human prostate carcinoma. *Clin Cancer Res* **11**, 5090-5097 (2005).
173. Lee, D.H., *et al.* Repair of methylation damage in DNA and RNA by mammalian AlkB homologues. *J Biol Chem* **280**, 39448-39459 (2005).
174. Ringvoll, J., *et al.* Repair deficient mice reveal mABH2 as the primary oxidative demethylase for repairing 1meA and 3meC lesions in DNA. *Embo J* **25**, 2189-2198 (2006).
175. Falnes, P.O., Bjoras, M., Aas, P.A., Sundheim, O. & Seeberg, E. Substrate specificities of bacterial and human AlkB proteins. *Nucleic Acids Res* **32**, 3456-3461 (2004).
176. Koivisto, P., Robins, P., Lindahl, T. & Sedgwick, B. Demethylation of 3-methylthymine in DNA by bacterial and human DNA dioxygenases. *J Biol Chem* **279**, 40470-40474 (2004).
177. Falnes, P.O. Repair of 3-methylthymine and 1-methylguanine lesions by bacterial and human AlkB proteins. *Nucleic Acids Res* **32**, 6260-6267 (2004).
178. Mishina, Y., Yang, C.G. & He, C. Direct repair of the exocyclic DNA adduct 1,N6-ethenoadenine by the DNA repair AlkB proteins. *J Am Chem Soc* **127**, 14594-14595 (2005).
179. Delaney, J.C., *et al.* AlkB reverses etheno DNA lesions caused by lipid oxidation *in vitro* and *in vivo*. *Nat Struct Mol Biol* **12**, 855-860 (2005).
180. Frick, L.E., Delaney, J.C., Wong, C., Drennan, C.L. & Essigmann, J.M. Alleviation of 1,N6-ethanoadenine genotoxicity by the *Escherichia coli* adaptive response protein AlkB. *Proc Natl Acad Sci U S A* **104**, 755-760 (2007).
181. Ringvoll, J., *et al.* AlkB homologue 2-mediated repair of ethenoadenine lesions in mammalian DNA. *Cancer Res* **68**, 4142-4149 (2008).
182. Saparbaev, M., Kleibl, K. & Laval, J. *Escherichia coli*, *Saccharomyces cerevisiae*, rat and human 3-methyladenine DNA glycosylases repair 1,N6-ethenoadenine when present in DNA. *Nucleic Acids Res* **23**, 3750-3755 (1995).

183. Guliaev, A.B., Hang, B. & Singer, B. Structural insights by molecular dynamics simulations into differential repair efficiency for ethano-A versus etheno-A adducts by the human alkylpurine-DNA *N*-glycosylase. *Nucleic Acids Res* **30**, 3778-3787 (2002).
184. Lee, C.Y., *et al.* Recognition and processing of a new repertoire of DNA substrates by human 3-methyladenine DNA glycosylase (AAG). *Biochemistry* **48**, 1850-1861 (2009).
185. Bleijlevens, B., Shivarattan, T., Sedgwick, B., Rigby, S.E. & Matthews, S.J. Replacement of non-heme Fe(II) with Cu(II) in the alpha-ketoglutarate dependent DNA repair enzyme AlkB: spectroscopic characterization of the active site. *J Inorg Biochem* **101**, 1043-1048 (2007).
186. Chen, H. & Costa, M. Iron- and 2-oxoglutarate-dependent dioxygenases: an emerging group of molecular targets for nickel toxicity and carcinogenicity. *Biometals* **22**, 191-196 (2009).
187. Chen, H., *et al.* Nickel ions inhibit histone demethylase JMJD1A and DNA repair enzyme ABH2 by replacing the ferrous iron in the catalytic centers. *J Biol Chem* **285**, 7374-7383 (2010).
188. Giri, N.C., Sun, H., Chen, H., Costa, M. & Maroney, M.J. X-ray absorption spectroscopy structural investigation of early intermediates in the mechanism of DNA repair by human ABH2. *Biochemistry* **50**, 5067-5076 (2011).
189. Ougland, R., *et al.* AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Mol Cell* **16**, 107-116 (2004).
190. Mishina, Y. & He, C. Probing the structure and function of the *Escherichia coli* DNA alkylation repair AlkB protein through chemical cross-linking. *J Am Chem Soc* **125**, 8730-8731 (2003).
191. Mishina, Y., Lee, C.H. & He, C. Interaction of human and bacterial AlkB proteins with DNA as probed through chemical cross-linking studies. *Nucleic Acids Res* **32**, 1548-1554 (2004).
192. Yu, B., *et al.* Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **439**, 879-884 (2006).
193. Sundheim, O., *et al.* Human ABH3 structure and key residues for oxidative demethylation to reverse DNA/RNA damage. *Embo J* **25**, 3389-3397 (2006).
194. Yang, C.G., *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961-965 (2008).
195. Yi, C. & He, C. AlkB recognition of a bulky DNA base adduct stabilized by chemical crosslinking. *Science China Chemistry* **53**, 86-90 (2009).
196. Holland, P.J. & Hollis, T. Structural and mutational analysis of *Escherichia coli* AlkB provides insight into substrate specificity and DNA damage searching. *PLoS One* **5**, e8680 (2010).
197. Yang, C.G., Garcia, K. & He, C. Damage detection and base flipping in direct DNA alkylation repair. *Chembiochem* **10**, 417-423 (2009).
198. Chen, B., Liu, H., Sun, X. & Yang, C.G. Mechanistic insight into the recognition of single-stranded and double-stranded DNA substrates by ABH2 and ABH3. *Mol Biosyst* **6**, 2143-2149 (2010).
199. Monsen, V.T., *et al.* Divergent ss-hairpins determine double-strand versus single-strand substrate recognition of human AlkB-homologues 2 and 3. *Nucleic Acids Res* **38**, 6447-6455 (2010).
200. Yu, B. & Hunt, J.F. 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 (2009).
201. Bleijlevens, B., *et al.* Dynamic states of the DNA repair enzyme AlkB regulate product release. *EMBO Rep* **9**, 872-877 (2008).
202. Shivarattan, T., Chen, H.A., Simpson, P., Sedgwick, B. & Matthews, S. Resonance assignments of *Escherichia coli* AlkB: a key 2-oxoglutarate and Fe(II) dependent dioxygenase of the adaptive DNA-repair response. *J Biomol NMR* **33**, 138 (2005).
203. Liu, H., Llano, J. & Gauld, J.W. A DFT study of nucleobase dealkylation by the DNA repair enzyme AlkB. *J Phys Chem B* **113**, 4887-4898 (2009).

204. Cisneros, G.A. DFT study of a model system for the dealkylation step catalyzed by AlkB. *Interdiscip Sci* **2**, 70-77 (2010).
205. Yi, C., *et al.* Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase. *Nature* **468**, 330-333 (2010).
206. Pan, Z., *et al.* Impaired placental trophoblast lineage differentiation in Alkbh1(-/-) mice. *Dev Dyn* **237**, 316-327 (2008).
207. Nordstrand, L.M., *et al.* Mice lacking Alkbh1 display sex-ratio distortion and unilateral eye defects. *PLoS One* **5**, e13827 (2010).
208. Bose, J., *et al.* The phosphatidylserine receptor has essential functions during embryogenesis but not in apoptotic cell removal. *J Biol* **3**, 15 (2004).
209. Tsujikawa, K., *et al.* Expression and sub-cellular localization of human ABH family molecules. *J Cell Mol Med* **11**, 1105-1116 (2007).
210. Westbye, M.P., *et al.* Human AlkB homolog 1 is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J Biol Chem* **283**, 25046-25056 (2008).
211. Muller, T.A., Meek, K. & Hausinger, R.P. Human AlkB homologue 1 (ABH1) exhibits DNA lyase activity at abasic sites. *DNA Repair (Amst)* **9**, 58-65 (2010).
212. Agris, P.F., Vendeix, F.A. & Graham, W.D. tRNA's wobble decoding of the genome: 40 years of modification. *J Mol Biol* **366**, 1-13 (2007).
213. Begley, U., *et al.* Trm9-catalyzed tRNA modifications link translation to the DNA damage response. *Mol Cell* **28**, 860-870 (2007).
214. Kalhor, H.R. & Clarke, S. Novel methyltransferase for modified uridine residues at the wobble position of tRNA. *Mol Cell Biol* **23**, 9283-9292 (2003).
215. Songe-Moller, L., *et al.* 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 (2010).
216. Fu, D., *et al.* Human AlkB homolog ABH8 Is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. *Mol Cell Biol* **30**, 2449-2459 (2010).
217. Fu, Y., *et al.* The AlkB domain of mammalian ABH8 catalyzes hydroxylation of 5-methoxycarbonylmethyluridine at the wobble position of tRNA. *Angew Chem Int Ed Engl* **49**, 8885-8888 (2010).
218. Kawakami, M., Takemura, S., Kondo, T., Fukami, T. & Goto, T. Chemical structure of a new modified nucleoside located in the anticodon of *Bombyx mori* glycine tRNA2. *J Biochem* **104**, 108-111 (1988).
219. van den Born, E., *et al.* ALKBH8-mediated formation of a novel diastereomeric pair of wobble nucleosides in mammalian tRNA. *Nat Commun* **2**, 172 (2011).
220. Shimada, K., *et al.* A novel human AlkB homologue, ALKBH8, contributes to human bladder cancer progression. *Cancer Res* **69**, 3157-3164 (2009).
221. Bjornstad, L.G., Zoppellaro, G., Tomter, A.B., Falnes, P.O. & Andersson, K.K. Spectroscopic and Magnetic Studies of wild-type and mutant forms of the Fe(II) and 2-oxoglutarate dependent decarboxylase ALKBH4. *Biochem J* (2010).
222. Thalhammer, A., *et al.* 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 (2011).
223. Peters, T., Ausmeier, K., Dildrop, R. & Ruther, U. 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 (2002).
224. van der Hoeven, F., *et al.* Programmed cell death is affected in the novel mouse mutant Fused toes (Ft). *Development* **120**, 2601-2607 (1994).
225. Banerjee-Basu, S. & Baxevanis, A.D. Molecular evolution of the homeodomain family of transcription factors. *Nucleic Acids Res* **29**, 3258-3269 (2001).
226. Peters, T., Ausmeier, K. & Ruther, U. Cloning of Fatso (Fto), a novel gene deleted by the Fused toes (Ft) mouse mutation. *Mamm Genome* **10**, 983-986 (1999).

227. Lango, H. & Weedon, M.N. What will whole genome searches for susceptibility genes for common complex disease offer to clinical practice? *J Intern Med* **263**, 16-27 (2008).
228. Frayling, T.M., *et al.* A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-894 (2007).
229. Dina, C., *et al.* Variation in FTO contributes to childhood obesity and severe adult obesity. *Nat Genet* **39**, 724-726 (2007).
230. Hunt, S.C., *et al.* Association of the FTO Gene With BMI. *Obesity* (2008).
231. Yajnik, C.S., *et al.* FTO gene variants are strongly associated with type 2 diabetes in South Asian Indians. *Diabetologia* **52**, 247-252 (2009).
232. Ho, A.J., *et al.* A commonly carried allele of the obesity-related *FTO* gene is associated with reduced brain volume in the healthy elderly. *Proc Natl Acad Sci U S A* **107**, 8404-8409 (2010).
233. Speakman, J.R., Rance, K.A. & Johnstone, A.M. Polymorphisms of the FTO gene are associated with variation in energy intake, but not energy expenditure. *Obesity (Silver Spring)* **16**, 1961-1965 (2008).
234. Wardle, J., Llewellyn, C., Sanderson, S. & Plomin, R. The FTO gene and measured food intake in children. *Int J Obes (Lond)* **33**, 42-45 (2009).
235. Cecil, J.E., Tavendale, R., Watt, P., Hetherington, M.M. & Palmer, C.N. An obesity-associated FTO gene variant and increased energy intake in children. *N Engl J Med* **359**, 2558-2566 (2008).
236. Olszewski, P.K., *et al.* Fto colocalizes with a satiety mediator oxytocin in the brain and upregulates oxytocin gene expression. *Biochem Biophys Res Commun* **408**, 422-426 (2011).
237. Andreasen, C.H. & Andersen, G. Gene-environment interactions and obesity – further aspects of Genome Wide Association Studies. *Nutrition* **25**, 998-1003 (2009).
238. Robbens, S., *et al.* The FTO gene, implicated in human obesity, is found only in vertebrates and marine algae. *J Mol Evol* **66**, 80-84 (2008).
239. Jia, G., *et al.* Oxidative demethylation of 3-methylthymine and 3-methyluracil in single-stranded DNA and RNA by mouse and human FTO. *FEBS Lett* **582**, 3313-3319 (2008).
240. Stratigopoulos, G., *et al.* Regulation of Fto/Ftm gene expression in mice and humans. *Am J Physiol Regul Integr Comp Physiol* (2008).
241. Fischer, J., *et al.* Inactivation of the Fto gene protects from obesity. *Nature* **458**, 894-898 (2009).
242. Church, C., *et al.* A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet* **5**, e1000599 (2009).
243. Church, C., *et al.* Overexpression of Fto leads to increased food intake and results in obesity. *Nat Genet* **42**, 1086-1092 (2010).
244. Tung, Y.C., *et al.* Hypothalamic-specific manipulation of Fto, the ortholog of the human obesity gene FTO, affects food intake in rats. *PLoS One* **5**, e8771 (2010).
245. Fredriksson, R., *et al.* The obesity gene, FTO, is of ancient origin, upregulated during food deprivation and expressed in neurons of feeding-related nuclei of the brain. *Endocrinology* (2008).
246. Olszewski, P.K., *et al.* Fto immunoreactivity is widespread in the rodent brain and abundant in feeding-related sites, but the number of Fto-positive cells is not affected by changes in energy balance. *Physiol Behav* **103**, 248-253 (2011).
247. Gao, X., *et al.* The fat mass and obesity associated gene FTO functions in the brain to regulate postnatal growth in mice. *PLoS One* **5**, e14005 (2010).
248. Boissel, S., *et al.* Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. *Am J Hum Genet* **85**, 106-111 (2009).
249. Meyre, D., *et al.* Prevalence of loss-of-function FTO mutations in lean and obese individuals. *Diabetes* **59**, 311-318 (2010).
250. van den Berg, L., *et al.* Investigation of a patient with a partial trisomy 16q including the fat mass and obesity associated gene (FTO): fine mapping and FTO gene expression study. *Am J Med Genet A* **152A**, 630-637 (2010).

251. Stratigopoulos, G., Leduc, C.A., Cremona, M.L., Chung, W.K. & Leibel, R.L. Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated (FTO) and retinitis pigmentosa gtpase regulator interacting protein-1 like (RPGRIPIL) genes, and co-ordinates leptin receptor signaling. *J Biol Chem* (2010).
252. Vierkotten, J., Dildrop, R., Peters, T., Wang, B. & Ruther, U. Ftm is a novel basal body protein of cilia involved in Shh signalling. *Development* **134**, 2569-2577 (2007).
253. Berulava, T. & Horsthemke, B. The obesity-associated SNPs in intron 1 of the FTO gene affect primary transcript levels. *Eur J Hum Genet* **18**, 1054-1056 (2010).
254. Bell, C.G., *et al.* Integrated genetic and epigenetic analysis identifies haplotype-specific methylation in the FTO type 2 diabetes and obesity susceptibility locus. *PLoS One* **5**, e14040 (2010).
255. Wahlen, K., Sjolin, E. & Hoffstedt, J. The common rs9939609 gene variant of the fat mass- and obesity-associated gene *FTO* is related to fat cell lipolysis. *J Lipid Res* **49**, 607-611 (2008).
256. Lappalainen, T., *et al.* Gene expression of FTO in human subcutaneous adipose tissue, peripheral blood mononuclear cells and adipocyte cell line. *J Nutrigenet Nutrigenomics* **3**, 37-45.
257. Bravard, A., *et al.* FTO is increased in muscle during Type 2 Diabetes, and its overexpression in myotubes alters insulin signaling, enhances lipogenesis and ROS production, and induces mitochondrial dysfunction. *Diabetes* **60**, 258-268 (2010).
258. http://www.who.int/mediacentre/factsheets/fs311/en/ World Health Organization. (2011).
259. Lee, I.M., Manson, J.E., Hennekens, C.H. & Paffenbarger, R.S., Jr. Body weight and mortality. A 27-year follow-up of middle-aged men. *J Amer Med Assoc* **270**, 2823-2828 (1993).
260. Chan, J.M., Rimm, E.B., Colditz, G.A., Stampfer, M.J. & Willett, W.C. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. *Diabetes Care* **17**, 961-969 (1994).
261. Stunkard, A.J., Foch, T.T. & Hrubec, Z. A twin study of human obesity. *J Amer Med Assoc* **256**, 51-54 (1986).
262. Maes, H.H., Neale, M.C. & Eaves, L.J. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet* **27**, 325-351 (1997).
263. O'Rahilly, S., Wainscoat, J.S. & Turner, R.C. Type 2 (non-insulin-dependent) diabetes mellitus. New genetics for old nightmares. *Diabetologia* **31**, 407-414 (1988).
264. Han, Z., *et al.* Crystal structure of the FTO protein reveals basis for its substrate specificity. *Nature* **464**, 1205-1209 (2010).
265. Quwailid, M.M., *et al.* A gene-driven ENU-based approach to generating an allelic series in any gene. *Mamm Genome* **15**, 585-591 (2004).
266. Russell, W.L., *et al.* Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc Natl Acad Sci U S A* **76**, 5818-5819 (1979).
267. Justice, M.J., Noveroske, J.K., Weber, J.S., Zheng, B. & Bradley, A. Mouse ENU mutagenesis. *Hum Mol Genet* **8**, 1955-1963 (1999).
268. Cordes, S.P. *N*-ethyl-*N*-nitrosourea mutagenesis: boarding the mouse mutant express. *Microbiol Mol Biol Rev* **69**, 426-439 (2005).
269. McDonough, M.A., Loenarz, C., Chowdhury, R., Clifton, I.J. & Schofield, C.J. Structural studies on human 2-oxoglutarate dependent oxygenases. *Curr Opin Struct Biol* **20**, 659-672 (2010).
270. Graslund, S., *et al.* Protein production and purification. *Nat Methods* **5**, 135-146 (2008).
271. Porath, J. & Flodin, P. Gel filtration: a method for desalting and group separation. *Nature* **183**, 1657-1659 (1959).
272. AgilentTechnologies. QuikChange® Site-Directed Mutagenesis Kit product manual. Vol. 200518-12.
273. Lancaster, D.E., *et al.* Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J* **383**, 429-437 (2004).

274. Flashman, E., *et al.* Kinetic rationale for selectivity toward *N*- and *C*-terminal oxygen-dependent degradation domain substrates mediated by a loop region of hypoxia-inducible factor prolyl hydroxylases. *J Biol Chem* **283**, 3808-3815 (2008).
275. Rost, B. & Liu, J. The PredictProtein server. *Nucleic Acids Res* **31**, 3300-3304 (2003).
276. Ohki, M., Tamura, F., Nishimura, S. & Uchida, H. Nucleotide sequence of the *Escherichia coli* dnaJ gene and purification of the gene product. *J Biol Chem* **261**, 1778-1781 (1986).
277. Beck, J.A., *et al.* Genealogies of mouse inbred strains. *Nat Genet* **24**, 23-25 (2000).
278. Mott, R. A haplotype map for the laboratory mouse. *Nat Genet* **39**, 1054-1056 (2007).
279. AgilentTechnologies. QuikChange® Multi Site Directed Mutagenesis Kit product manual. Vol. 200514-12.
280. Schagger, H., Cramer, W.A. & von Jagow, G. Analysis of molecular masses and oligomeric states of protein complexes by blue native electrophoresis and isolation of membrane protein complexes by two-dimensional native electrophoresis. *Anal Biochem* **217**, 220-230 (1994).
281. Janatova, J., Fuller, J.K. & Hunter, M.J. The heterogeneity of bovine albumin with respect to sulfhydryl and dimer content. *J Biol Chem* **243**, 3612-3622 (1968).
282. Colman, P.M. & Matthews, B.W. Symmetry, molecular weight and crystallographic data for sweet potato -amylase. *J Mol Biol* **60**, 163-168 (1971).
283. AmershamBiosciences. Gel Filtration: Principles and Methods. in *GE Healthcare Handbooks*, Vol. 18-1022-18 (2010).
284. Geoghegan, K.F., *et al.* Spontaneous alpha-*N*-6-phosphogluconoylation of a "His tag" in *Escherichia coli*: the cause of extra mass of 258 or 178 Da in fusion proteins. *Anal Biochem* **267**, 169-184 (1999).
285. Yan, Z., Caldwell, G.W. & McDonell, P.A. Identification of a gluconic acid derivative attached to the *N*-terminus of histidine-tagged proteins expressed in bacteria. *Biochem Biophys Res Commun* **262**, 793-800 (1999).
286. Hardy, A. Structural Studies on the Hypoxic Response, University of Oxford (2009).
287. Wu, Q., *et al.* Carbonic anhydrase inhibitor binding: From solution to the gas phase. *J Am Chem Soc* **119**, 1157-1158 (1997).
288. Smith, R.D., Light-Wahl, K.J., Winger, B.E. & Loo, J.A. Preservation of non-covalent associations in electrospray ionisation mass spectrometry: Multiply charged polypeptide and protein dimers. *Org. Mass Spectrom.* **27**, 811-821 (1992).
289. Busman, M., Knapp, D.R. & Schey, K.L. Observation of large multimers in the electrospray ionization mass spectrometry of peptides. *Rapid Commun Mass Spectrom* **8**, 211-216 (1994).
290. Smith, R.D. & Light-Wahl, K.J. The observation of non-covalent interactions in solution by electrospray ionisation mass spectrometry: Promise, pitfalls and prognosis. *Biol. Mass. Spectrom.* **22**, 493-501 (1993).
291. Sabourin, P.J. & Bieber, L.L. Purification and characterization of an alpha-ketoisocaproate oxygenase of rat liver. *J Biol Chem* **257**, 7460-7467 (1982).
292. Myllyla, R., Majamaa, K., Gunzler, V., Hanauske-Abel, H.M. & Kivirikko, K.I. Ascorbate is consumed stoichiometrically in the uncoupled reactions catalyzed by prolyl 4-hydroxylase and lysyl hydroxylase. *J Biol Chem* **259**, 5403-5405 (1984).
293. Flashman, E., Davies, S.L., Yeoh, K.K. & Schofield, C.J. Investigating the dependence of the hypoxia-inducible factor hydroxylases (factor inhibiting HIF and prolyl hydroxylase domain 2) on ascorbate and other reducing agents. *Biochem J* **427**, 135-142 (2010).
294. Greenfield, N. & Fasman, G.D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108-4116 (1969).
295. Manning, M.C. Underlying assumptions in the estimation of secondary structure content in proteins by circular dichroism spectroscopy - a critical review. *J Pharmaceut Biomed Anal* **7**, 1103-1119 (1989).
296. Durst, R.A. & Staples, B.R. Tris-tris-HCl: a standard buffer for use in the physiologic pH range. *Clin Chem* **18**, 206-208 (1972).

297. Good, N.E., *et al.* Hydrogen ion buffers for biological research. *Biochemistry* **5**, 467-477 (1966).
298. Vedadi, M., *et al.* Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc Natl Acad Sci USA* **103**, 15835-15840 (2006).
299. Matulis, D., Kranz, J.K., Salemme, F.R. & Todd, M.J. Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **44**, 5258-5266 (2005).
300. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545-549 (2010).
301. Dahanukar, A., Walker, J.A. & Wharton, R.P. Smaug, a novel RNA-binding protein that operates a translational switch in Drosophila. *Mol Cell* **4**, 209-218 (1999).
302. Smibert, C.A., Lie, Y.S., Shillinglaw, W., Henzel, W.J. & Macdonald, P.M. Smaug, a novel and conserved protein, contributes to repression of nanos mRNA translation *in vitro*. *RNA* **5**, 1535-1547 (1999).
303. Hewitson, K.S., *et al.* Evidence that two enzyme-derived histidine ligands are sufficient for iron binding and catalysis by factor inhibiting HIF (FIH). *J Biol Chem* **283**, 25971-25978 (2008).
304. Wu, Q., Saunders, R.A., Szkudlarek-Mikho, M., Serna Ide, L. & Chin, K.V. The obesity-associated Fto gene is a transcriptional coactivator. *Biochem Biophys Res Commun* **401**, 390-395 (2010).
305. Russell, M.A. & Morgan, N.G. Conditional expression of the FTO gene product in rat INS-1 cells reveals its rapid turnover and a role in the profile of glucose-induced insulin secretion. *Clin Sci (Lond)* (2010).
306. Bissett, M.L. & Payne, F.E. Development of antigens in human cells infected with Simian virus 40. *J Bacteriol* **91**, 743-749 (1966).
307. Gey, G.O., Coffman, W.D. & Kubick, M.T. Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res* **12**, 264-265 (1952).
308. Pollard, T.D. & Cooper, J.A. Actin, a central player in cell shape and movement. *Science* **326**, 1208-1212 (2009).
309. Shechter, D., Dormann, H.L., Allis, C.D. & Hake, S.B. Extraction, purification and analysis of histones. *Nat Protoc* **2**, 1445-1457 (2007).
310. Kindt, T.J., Osborne, B.A. & Goldsby, R. *Kuby Immunology: International Edition*, (W. H. Freeman; 6th edition, New York, 2006).
311. Muyldermans, S. Single domain camel antibodies: current status. *J Biotechnol* **74**, 277-302 (2001).
312. Janeway, C. *Immunobiology : the immune system in health and disease*, (Churchill Livingstone, London, 2005).
313. Brayer, K.J. & Segal, D.J. Keep your fingers off my DNA: protein-protein interactions mediated by C2H2 zinc finger domains. *Cell Biochem Biophys* **50**, 111-131 (2008).
314. Barlow, J.N., Zhang, Z., John, P., Baldwin, J.E. & Schofield, C.J. Inactivation of 1-aminocyclopropane-1-carboxylate oxidase involves oxidative modifications. *Biochemistry* **36**, 3563-3569 (1997).
315. Welford, R.W.D. Studies on the DNA repair Enzyme AlkB and the Ascorbate Dependence of the 2-Oxoglutarate Oxygenases, University of Oxford (2004).
316. Cockman, M.E., Webb, J.D., Kramer, H.B., Kessler, B.M. & Ratcliffe, P.J. Proteomics-based identification of novel factor inhibiting hypoxia-inducible factor (FIH) substrates indicates widespread asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Mol Cell Proteomics* **8**, 535-546 (2009).
317. Elias, J.E., Haas, W., Faherty, B.K. & Gygi, S.P. Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nat Methods* **2**, 667-675 (2005).
318. Geetha, T. & Wooten, M.W. Structure and functional properties of the ubiquitin binding protein p62. *FEBS Lett* **512**, 19-24 (2002).

319. Nielsen, M.L., *et al.* Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Methods* **5**, 459-460 (2008).
320. Nakano, K., *et al.* Isolation, characterization and structural organization of the gene and pseudogene for the dihydrolipoamide succinyltransferase component of the human 2-oxoglutarate dehydrogenase complex. *Eur J Biochem* **224**, 179-189 (1994).
321. Koike, M. & Koike, K. Structure, assembly and function of mammalian alpha-keto acid dehydrogenase complexes. *Adv Biophys*, 187-227 (1976).
322. Berg, J.M., Tymoczko, J.L. & Stryer, L. *Biochemistry*, (W. H. Freeman, New York, 2007).
323. Squire, J.M. Architecture and function in the muscle sarcomere. *Curr Opin Struct Biol* **7**, 247-257 (1997).
324. Pagliarini, D.J., *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112-123 (2008).
325. Welford, R.W., Schlemminger, I., McNeill, L.A., Hewitson, K.S. & Schofield, C.J. The selectivity and inhibition of AlkB. *J Biol Chem* **278**, 10157-10161 (2003).
326. Koivisto, P., Duncan, T., Lindahl, T. & Sedgwick, B. Minimal methylated substrate and extended substrate range of Escherichia coli AlkB protein, a 1-methyladenine-DNA dioxygenase. *J Biol Chem* **278**, 44348-44354 (2003).
327. Kawamura, A., *et al.* Development of homogeneous luminescence assays for histone demethylase catalysis and binding. *Anal Biochem* **404**, 86-93 (2010).
328. www.proteinatlas.org. (The Human Protein Atlas, 2011).
329. Claros, M.G. & Vincens, P. Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241**, 779-786 (1996).
330. Invitrogen. Lipofectamine 2000^{TM} Protocol manual. (2006).
331. Cole, C., Barber, J.D. & Barton, G.J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* **36**, W197-201 (2008).
332. McNeill, L.A., *et al.* Hypoxia-inducible factor prolyl hydroxylase 2 has a high affinity for ferrous iron and 2-oxoglutarate. *Mol Biosyst* **1**, 321-324 (2005).
333. Shannon, R.D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides *Acta Cryst.* **32**, 751-767 (1976).
334. Munday, R., Munday, C.M. & Winterbourn, C.C. Inhibition of copper-catalyzed cysteine oxidation by nanomolar concentrations of iron salts. *Free Radic Biol Med* **36**, 757-764 (2004).
335. Roy, T.W. & Bhagwat, A.S. Kinetic studies of *Escherichia coli* AlkB using a new fluorescence-based assay for DNA demethylation. *Nucleic Acids Res* **35**, e147 (2007).
336. Mikhailov, S.N., *et al.* Chemical incorporation of 1-methyladenosine into oligonucleotides. *Nucleic Acids Res* **30**, 1124-1131 (2002).
337. Beranek, D.T., Weis, C.C. & Swenson, D.H. A comprehensive quantitative analysis of methylated and ethylated DNA using high pressure liquid chromatography. *Carcinogenesis* **1**, 595-606 (1980).
338. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).
339. Karkhanina, A.A., *et al.* Direct analysis of enzyme-catalyzed DNA demethylation. *Anal Chem* **81**, 5871-5875 (2009).
340. Li, D., *et al.* Repair of DNA alkylation damage by the *Escherichia coli* adaptive response protein AlkB as studied by ESI-TOF mass spectrometry. *J Nucleic Acids* **2010**, 369434 (2010).
341. Beaucage, S.L. & Iyer, R.P. Advances in the Synthesis of Oligonucleotides by the Phosphoramidite Approach. *Tetrahedron* **48**, 2223-2311 (1992).
342. Engel, J.D. Mechanism of the Dimroth rearrangement in adenosine. *Biochem Biophys Res Commun* **64**, 581-586 (1975).
343. Timofeev, E.N., *et al.* Oligodeoxynucleotides containing 2-deoxy-1-methyladenosine and Dimroth rearrangement. *Helvetica Chimica Acta* **90**, 928-937 (2007).
344. Henshaw, T.F., Feig, M. & Hausinger, R.P. Aberrant activity of the DNA repair enzyme AlkB. *J Inorg Biochem* **98**, 856-861 (2004).

345. Lipscomb, S.J., *et al.* The role of arginine residues in substrate binding and catalysis by deacetoxycephalosporin C synthase. *Eur J Biochem* **269**, 2735-2739 (2002).
346. Hopkinson, R.J., Hamed, R.B., Rose, N.R., Claridge, T.D. & Schofield, C.J. Monitoring the activity of 2-oxoglutarate dependent histone demethylases by NMR spectroscopy: direct observation of formaldehyde. *Chembiochem* **11**, 506-510 (2010).
347. Flashman, E., *et al.* Evidence for the slow reaction of hypoxia-inducible factor prolyl hydroxylase 2 with oxygen. *Febs J* **277**, 4089-4099 (2010).
348. Kaule, G. & Gunzler, V. Assay for 2-oxoglutarate decarboxylating enzymes based on the determination of [1-14C]succinate: application to prolyl 4-hydroxylase. *Anal Biochem* **184**, 291-297 (1990).
349. Kanelakis, K.C., *et al.* Characterization of a robust enzymatic assay for inhibitors of 2-oxoglutarate-dependent hydroxylases. *J Biomol Screen* **14**, 627-635 (2009).
350. Crain, P.F. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol* **193**, 782-790 (1990).
351. Le, T., Kim, K.P., Fan, G. & Faull, K.F. A sensitive mass spectrometry method for simultaneous quantification of DNA methylation and hydroxymethylation levels in biological samples. *Anal Biochem* **412**, 203-209 (2011).
352. Munzel, M., *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl* **49**, 5375-5377 (2010).
353. Ando, M., *et al.* Formaldehyde dehydrogenase from *Pseudomonas putida*. Purification and some properties. *J Biochem (Tokyo)* **85**, 1165-1172 (1979).
354. Lohmann, W., Lohmann, C. & Ibrahim, M. Fluorescence spectra of NADH/NAD, kynurenine, tryptophan, and tyrosine. *Naturwissenschaften* **75**, 141-142 (1988).
355. Dawson, R.M.C., Elliott, D.C., Elliott, W.H. & Jones, K.M. *Data for biochemical research*, (Clarendon Press, Oxford, 1986).
356. Atkins, P.W. & De Paula, J. *Atkins' Physical chemistry, 9th Edition*, (Oxford University Press, Oxford, 2009).
357. Oppenheimer, N.J., Henehan, G.T., Huete-Perez, J.A. & Ito, K. *P. putida* formaldehyde dehydrogenase. An alcohol dehydrogenase masquerading as an aldehyde dehydrogenase. *Adv Exp Med Biol* **414**, 417-423 (1997).
358. Rose, N.R., *et al.* Inhibitor scaffolds for 2-oxoglutarate-dependent histone lysine demethylases. *J Med Chem* **51**, 7053-7056 (2008).
359. Sakurai, M., *et al.* A miniaturized screen for inhibitors of Jumonji histone demethylases. *Mol Biosyst* **6**, 357-364 (2010).
360. King, O.N., *et al.* Quantitative high-throughput screening identifies 8-hydroxyquinolines as cell-active histone demethylase inhibitors. *PLoS One* **5**, e15535 (2010).
361. Ehrismann, D., *et al.* Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem J* **401**, 227-234 (2007).
362. Chowdhury, R., *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep* (2011).
363. Mishina, Y., Chen, L.X. & He, C. Preparation and characterization of the native iron(II)-containing DNA repair AlkB protein directly from *Escherichia coli*. *J Am Chem Soc* **126**, 16930-16936 (2004).
364. Zhang, J.H., Chung, T.D. & Oldenburg, K.R. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* **4**, 67-73 (1999).
365. Hewitson, K.S. & Schofield, C.J. The HIF pathway as a therapeutic target. *Drug Discov Today* **9**, 704-711 (2004).
366. Kampranis, S.C. & Tsichlis, P.N. Histone demethylases and cancer. *Adv Cancer Res* **102**, 103-169 (2009).
367. Tung, Y.C. & Yeo, G.S. From GWAS to biology: lessons from FTO. *Ann N Y Acad Sci* **1220**, 162-171 (2010).
368. McDonough, M.A., *et al.* Selective inhibition of factor inhibiting hypoxia-inducible factor. *J Am Chem Soc* **127**, 7680-7681 (2005).

369. Mecinovic, J., Loenarz, C., Chowdhury, R. & Schofield, C.J. 2-Oxoglutarate analogue inhibitors of prolyl hydroxylase domain 2. *Bioorg Med Chem Lett* **19**, 6192-6195 (2009).
370. Thalhammer, A., *et al.* Inhibition of the histone demethylase JMJD2E by 3-substituted pyridine 2,4-dicarboxylates. *Org Biomol Chem* **9**, 127-135 (2011).
371. Chan, D.A. & Giaccia, A.J. PHD2 in tumour angiogenesis. *Br J Cancer* **103**, 1-5 (2010).
372. Frye, S.V. The art of the chemical probe. *Nat Chem Biol* **6**, 159-161 (2010).
373. Rose, N.R., *et al.* Selective inhibitors of the JMJD2 histone demethylases: combined nondenaturing mass spectrometric screening and crystallographic approaches. *J Med Chem* **53**, 1810-1818 (2010).
374. Loenarz, C. & Schofield, C.J. Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem Sci* (2010).
375. Chaney, S.G. & Sancar, A. DNA repair: enzymatic mechanisms and relevance to drug response. *J Natl Cancer Inst* **88**, 1346-1360 (1996).
376. Stevens, M.F., *et al.* Antitumor activity and pharmacokinetics in mice of 8-carbamoyl-3-methyl-imidazo[5,1-d]-1,2,3,5-tetrazin-4(3H)-one (CCRG 81045; M & B 39831), a novel drug with potential as an alternative to dacarbazine. *Cancer Res* **47**, 5846-5852 (1987).
377. Tano, K., Shiota, S., Collier, J., Foote, R.S. & Mitra, S. Isolation and structural characterization of a cDNA clone encoding the human DNA repair protein for O6-alkylguanine. *Proc Natl Acad Sci U S A* **87**, 686-690 (1990).
378. Stupp, R., *et al.* Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol* **10**, 459-466 (2009).
379. Dolan, M.E., Moschel, R.C. & Pegg, A.E. Depletion of mammalian O6-alkylguanine-DNA alkyltransferase activity by O6-benzylguanine provides a means to evaluate the role of this protein in protection against carcinogenic and therapeutic alkylating agents. *Proc Natl Acad Sci U S A* **87**, 5368-5372 (1990).
380. Ueno, T., *et al.* Modulation of chemotherapy resistance in regional therapy: a novel therapeutic approach to advanced extremity melanoma using intra-arterial temozolomide in combination with systemic O6-benzylguanine. *Mol Cancer Ther* **5**, 732-738 (2006).
381. Quinn, J.A., *et al.* Phase II trial of temozolomide plus O6-benzylguanine in adults with recurrent, temozolomide-resistant malignant glioma. *J Clin Oncol* **27**, 1262-1267 (2009).
382. Hewitson, K.S., *et al.* Structural and mechanistic studies on the inhibition of the hypoxia-inducible transcription factor hydroxylases by tricarboxylic acid cycle intermediates. *J Biol Chem* **282**, 3293-3301 (2007).
383. Dang, L., *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739-744 (2009).
384. Reitman, Z.J. & Yan, H. Isocitrate dehydrogenase 1 and 2 mutations in cancer: alterations at a crossroads of cellular metabolism. *J Natl Cancer Inst* **102**, 932-941 (2010).
385. Struys, E.A. D-2-Hydroxyglutaric aciduria: unravelling the biochemical pathway and the genetic defect. *J Inherit Metab Dis* **29**, 21-29 (2006).
386. Chang, K.H., *et al.* Inhibition of histone demethylases by 4-carboxy-2,2'-bipyridyl compounds. *ChemMedChem* **6**, 759-764 (2011).
387. Fedorov, O., *et al.* A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc Natl Acad Sci U S A* **104**, 20523-20528 (2007).
388. Poulsen, S.A. Direct screening of a dynamic combinatorial library using mass spectrometry. *J Am Soc Mass Spectrom* **17**, 1074-1080 (2006).
389. Aplin, R.T., Robinson, C.V., Schofield, C.J. & Westwood, N.J. Does the observation of noncovalent complexes between biomolecules by Electrospray-Ionization Mass-

Spectrometry necessarily reflect specific solution interactions? *Journal of the Chemical Society-Chemical Communications*, 2415-2417 (1994).
390. Erba, E.B. & Zenobi, R. Mass spectrometric studies of dissociation constants of noncovalent complexes. *Annu. Rep. Prog. Chem., Sect. C* (2011).
391. Loenarz, C., *et al.* PHF8, a gene associated with cleft lip/palate and mental retardation, encodes for an N^{ε}-dimethyl lysine demethylase. *Hum Mol Genet* **19**, 217-222 (2010).
392. Baba, T., *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006 0008 (2006).
393. Nichols, R.J., *et al.* Phenotypic landscape of a bacterial cell. *Cell* **144**, 143-156 (2011).
394. Typas, A., *et al.* High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat Methods* **5**, 781-787 (2008).
395. Studier, F.W. & Moffatt, B.A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189**, 113-130 (1986).
396. Gasteiger, E., *et al. The Proteomics Protocols Handbook: Protein identification and analysis tools on the ExPASy server (pp 571-607)*, (Humana Press, Totowa, NJ, 2005).
397. Laemmli, U.K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685 (1970).
398. Niesen, F.H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-2221 (2007).
399. Otwinowski, Z., Minor, W. & Charles W. Carter, Jr. Processing of X-ray diffraction data collected in oscillation mode. in *Methods in Enzymology*, Vol. 276 307-326 (Academic Press, 1997).
400. McCoy, A.J., *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
401. Adams, P.D., *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).
402. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132 (2004).
403. Yamamoto, Y., Katsuki, M., Sekiguchi, M. & Otsuji, N. *Escherichia coli* gene that controls sensitivity to alkylating agents. *J Bacteriol* **135**, 144-152 (1978).

Appendices

A: Structures of methylated bases

Modified nucleic acids are referred to throughout using standard abbreviations. Bases are named using the one letter code, and modifications are denoted using the number of the ring atom to which the modification is attached, with a standard abbreviation describing the group present e.g. 1-methyl adenine = 1meA (Figure A1). In the case of modifications present on atoms that do not directly form part of the base ring structure the identity of the atom to which the modification is joined is denoted by the elemental symbol, with the ring position to which this atom is attached indicated by a superscript figure e.g. 6-methyl guanosine = O^6meG (Figure A2).

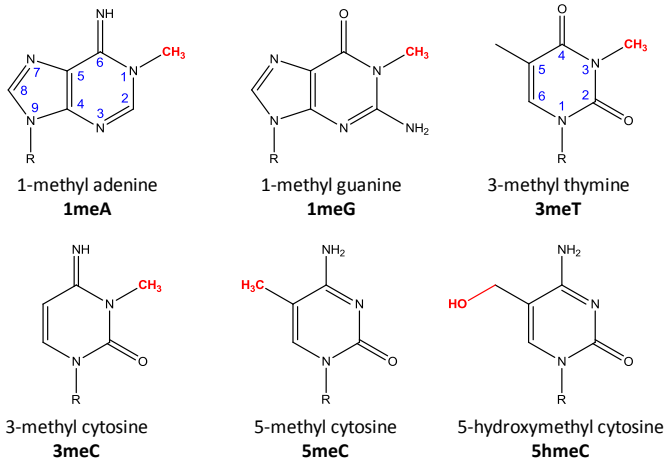


Figure A1: Methylated bases commonly discussed in this study. Numbering of carbon atoms is indicated for representative structures, and modifications to the structures are added in red.

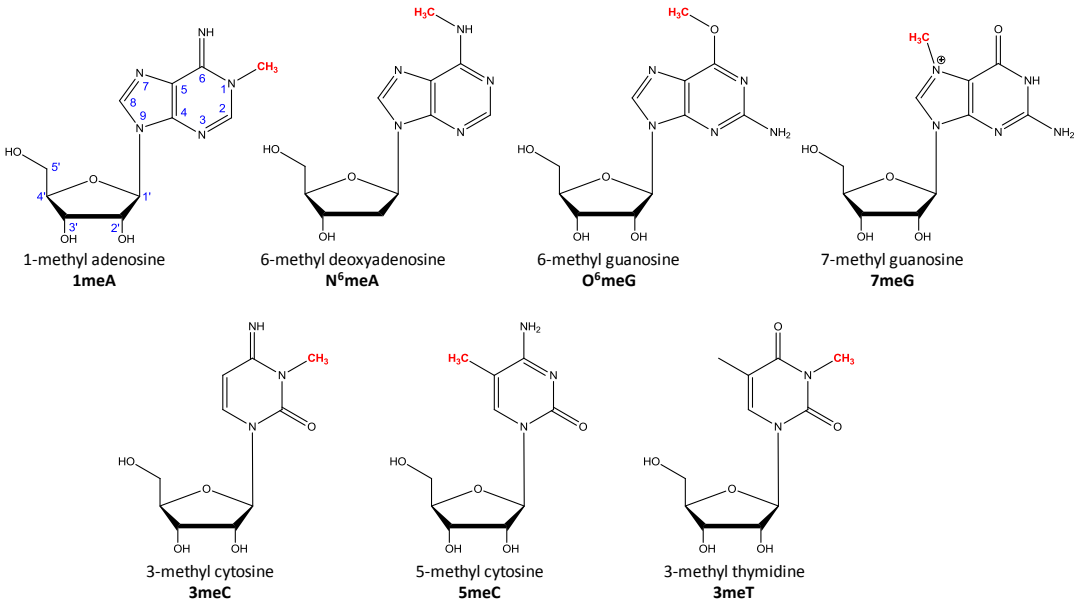


Figure A2: Methylated nucleosides tested in Chapter 4. Numbering of carbon atoms is indicated for representative structures, and modifications to the structures are added in red.

B: Oligonucleotide primers

Construct	Name	Sequence
1- A321 mFTO	*Fwd*	5'-CTCACCGTGTGGCA**TAG**TGCTCAACAGGCACC-3'
	Rev	5'-GCCTGTTGAGCA**CTA**TGCCACACGGTGAGTGG-3'
1- D329 mFTO	*Fwd*	5'-CAGGCACCTTGGAT**TAG**ATCTTAGAACGCTG-3'
	Rev	5'-GCGTTCTAAGAT**CTA**ATCCAAGGTGCCTGTTG-3'
1- K387 mFTO	*Fwd*	5'-GGAAGAAGATGGAG**TGA**GTGACAAATGCGGTGC-3'
	Rev	5'-CCGCATTTGTCAC**TCA**CTCCACCTCTTCTTCCAC-3'
1- E408 mFTO	*Fwd*	5'-GGAAGAAGATGGAG**TGA**GTGCAAATGCGGTGC-3'
	Rev	5'-CCGCATTTGTCAC**TCA**CTCCATCTTCTTCCAC-3'
CmFTOa deletion	*Fwd*	5'-GTGCCGCGCGGCAGCGATTATATCTTAGAACGC-3'
	Rev	5'-GTTCTAAGATATAATCGCTGCCGCGCGGCACC-3'
CmFTO deletion	*Fwd*	5'-GCCATCATCATCATCATCACGATTATATCTTAGAACGC-3'
	Rev	5'-GCGTTCTAAGATATAATCGTGATGATGATGATGATGGC-3'
I367F mFTO	*Fwd*	5'-CAAGGAGAGGAA**TTC**CATAATGAGGTGGAG-3'
	Rev	5'-CCACCTCATTATG**GAA**TTCCTCTCCTTGTTTC-3'
I367A mFTO	*Fwd*	5'-CAAGGAGAGGAA**GCG**CATAATGAGGTGGAG-3'
	Rev	5'-CTCCACCTCATTATG**CGC**TTCCTCTCCTTG-3'
S383N mFTO	*Fwd*	5'-GCAGTTCTGGTTTCAAGGC**AAT**CGATACAAACTTTGC-3'
	Rev	5'-GCAAAGTTTGTATCG**ATT**GCCTTGAAACCAGAACTGC-3'
V409M mFTO	*Fwd*	5'-GGAAGAAGATGGAGAGC**ATG**ACAAATGCGGTGCTCCG-3'
	Rev	5'-CGGAGCACCGCATTTGT**CAT**GCTCTCCATCTTCTTCC-3'
A462V mFTO	*Fwd*	5'-CGTCCGGACTTTACCA**GTA**CAGCAGAAACCAGAC-3'
	Rev	5'-GTCTGGTTTCTGCTG**TAC**TGGTAAAGTCCGGACG-3'
hFTO deletion	*Fwd*	5'-CATCATCATCATCACAAACGTACCCCGACCGCG-3'
	Rev	5'-CGCGGTCGGGGTACGTTTGTGATGATGATGATG-3'
ChFTO	*Fwd*	5'-GCCATCATCATCATCATCACGATTACATCCTGCAACGC -3'
	Rev	5'-GCGTTGCAGGATGTAATCGTGATGATGATGATGATGGC -3'

Table B1: Oligonucleotide primers used in Chapter 2. Red codons identify sites of point mutations.

Construct	Name	Sequence
hFTO	*FwdA*	5'-GG**GGTACC**ATGAAACGTACCCCGACCG-3' KpnI
	FwdB	5'-GG**GGTACC**AATGAAACGTACCCCGACC-3' KpnI
	RevA	5'-CG**GGATCC**TTACGGTTTCGCTTCCAGC-3' BamHI
	RevB	5'-CG**GGATCC**CTCGGTTTCGCTTCCAGCAG-3' BamHI
I370F hFTO	*Fwd*	5'-GAAACAGGGCGAAGAA**TTC**CATAACGAAGTGG-3'
	Rev	5'-CCACTTCGTTATG**GAA**TTCTTCGCCCTGTTTC-3'
H307A hFTO	*Fwd*	5'- CTGAACGCGACCCATCAG**GCT**TGCGTGCTGGCCGGTAGC-3'
	Rev	5'- GCTACCGGCCAGCACGCA**AGC**CTGATGGGTCGCGTTCAG-3'
R316A hFTO	*Fwd*	5'-GCTGGCCGGTAGCCAGCCG**GCT**TTTAGCAGCACCCATCG-3'
	Rev	5'-CGATGGGTGCTGCTAAA**AGC**CGGCTGGCTACCGGCCAGC-3'
FLAG NhFTO deletion	*Fwd*	5'-GGAATGCAGCACCGGCACCCTGTAAGGATCCCGGGTGGC-3'
	Rev	5'-GCCACCCGGGATCCTTACAGGGTGCCGGTGCTGCATTCC-3'
pcDNA NhFTO deletion	*Fwd*	5'-GAATGCAGCACCGGCACCCTGTAAGGATCCACTAGTAACG-3'
	Rev	5'-CGTTACTAGTGGATCCTTACAGGGTGCCGGTGCTGCATTC-3'
N1 EGFP NhFTO deletion	*Fwd*	5'-GGAATGCAGCACCGGCACCCTGAGGGATCCACCGGTCGCC-3'
	Rev	5'-GGCGACCGGTGGATCCCTCAGGGTGCCGGTGCTGCATTCC-3'
C1 EGFP NhFTO deletion	*Fwd*	5'-GGAATGCAGCACCGGCACCCTGTAAGATCCACCGGATC-3'
	Rev	5'-GATCCGGTGGATCTTACAGGGTGCCGGTGCTGCATTCC-3'
NLS-ChFTO deletion	*Fwd*	5'-CTGCGTCTGCTGGAAGAAGATTACATCCTGCAACGC-3'
	Rev	5'-GCGTTGCAGGATGTAATCTTCTTCCAGCAGACGCAG-3'
YFP-CmFTO	*Fwd*	5'-GGACGAGCTGTACAAGGATTATATCTTAGAACGC-3'
	Rev	5'-GCGTTCTAAGATATAATCCTTGTACAGCTCGTCC-3'

Table B2: Oligonucleotide primers used in Chapter 3. Blue bases show restriction sites used. Red codons identify sites of point mutations.

Construct	Name	Sequence
pET24a AlkB ΔN11	*Fwd*	5'-GAAGGAGATATACATATGCAAGAGCCACTGGCGGCTGG-3'
	Rev	5'-CCAGCCGCCAGTGGCTCTTGCATATGTATATCTCCTTC-3'
pET28a AlkB ΔN11	*Fwd*	5'-GCCGCGCGGCAGCCATATGCAAGAGCCACTGGCGGCTGG-3'
	Rev	5'-CCAGCCGCCAGTGGCTCTTGCATATGGCTGCCGCGCGGC-3'

Table B3: Oligonucleotide primers used in Chapter 5.

C: Summary of FTO in vitro *expression constructs*

Construct Name	Description	Optimal Expression Conditions	Creator
mFTOa	I.M.A.G.E. murine FTO cloned into the NheI and EcoRI sites of pET-28a(+)	0.5 mM IPTG, 18 °C, 4 h	T. Gerken
mFTO	Deletion of thrombin cleavage peptide from mFTOa	0.5 mM IPTG, 18 °C, 4 h	T. Gerken
H304A mFTO	Substitution of Fe(II) binding residue	0.5 mM IPTG, 18 °C, 4 h	T. Gerken
R313A mFTO	Substitution of 2OG binding residue	0.5 mM IPTG, 18 °C, 4 h	T. Gerken
hFTOa	Human FTO (synthetic gene) cloned into the NdeI and BamHI sites of pET-28a(+)	0.5 mM IPTG, 15 °C, 4 h	C. Webby
hFTO	Deletion of thrombin cleavage peptide from hFTOa	0.5 mM IPTG, 15 °C, 16 h	E. Bagg
ChFTO	Residues 332-505 of hFTO: *C*-terminal domain	0.2 mM IPTG, 15 °C, 4 h	E. Bagg
CmFTOa	Residues 329-502 of mFTOa (thrombin-cleavable construct): *C*-terminal domain	0.5 mM IPTG, 15 °C, 4 h	E. Bagg
CmFTO	Residues 329-502 of mFTO: *C*-terminal domain	0.5 mM IPTG, 15 °C, 16 h	E. Bagg
mFTO 1-E408	Residues 1-408 of mFTO	Minimal soluble expression	E. Bagg
mFTO 1-K387	Residues 1-387 of mFTO	Minimal soluble expression	E. Bagg
mFTO 1-D329	Residues 1-329 of mFTO: *N*-terminal domain	Minimal soluble expression	E. Bagg
mFTO 1-A321	Residues 1-321 of mFTO	Minimal soluble expression	E. Bagg
I367F mFTO	Substitution observed in mouse model	Minimal soluble expression	E. Bagg
I367F CmFTO	Substitution observed in mouse model produced in *C*-terminal domain construct	Minimal soluble expression	E. Bagg
I367A mFTO	Substitution at site of mouse mutation in mFTO construct	0.2 mM IPTG, 15 °C, 16 h	E. Bagg
I367A CmFTO	Substitution at site of mouse mutation in *C*-terminal domain construct	0.2 mM IPTG, 15 °C, 4 h	E. Bagg
mFTO NCBI	NCBI murine FTO sequence produced from I.M.A.G.E. mFTO by mutagenesis	Low expression: 0.5 mM IPTG, 15 °C, 3 h	E. Bagg
CmFTO NCBI	NCBI sequence produced from I.M.A.G.E. CmFTO by mutagenesis	Low expression: 0.5 mM IPTG, 20 °C, 16 h	E. Bagg
I367F mFTO NCBI	NCBI sequence produced from I.M.A.G.E. I367F mFTO by mutagenesis	Minimal soluble expression	E. Bagg
I367F CmFTO NCBI	NCBI sequence produced from I.M.A.G.E. I367F CmFTO by mutagenesis	Minimal soluble expression	E. Bagg
I367A mFTO NCBI	NCBI sequence produced from I.M.A.G.E. I367A mFTO by mutagenesis	Low expression: 0.5 mM IPTG, 15 °C, 16 h	E. Bagg
I367A CmFTO NCBI	NCBI sequence produced from I.M.A.G.E. I367A CmFTO by mutagenesis	Low expression: 0.5 mM IPTG, 20 °C, 3 h	E. Bagg

Table C1: Summary of FTO constructs and expression conditions described in Chapter 2.

D: 2-Oxoglutarate analogue compound structures

The following compounds are 2OG analogues tested in Chapter 6.

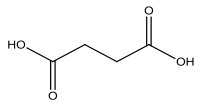
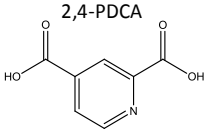
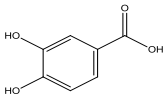
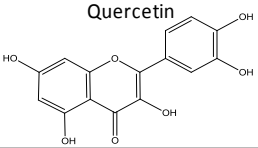
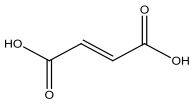
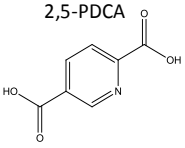
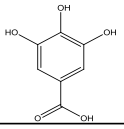
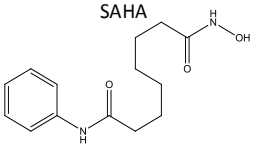
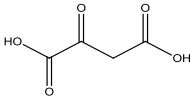
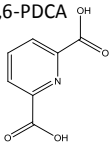
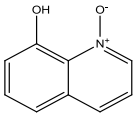
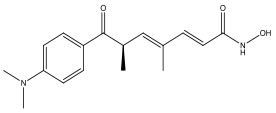
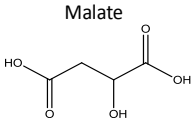
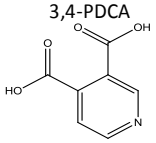
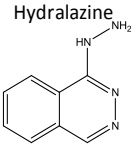
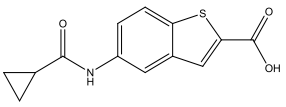
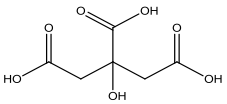
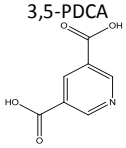
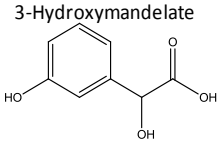
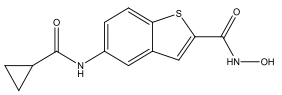
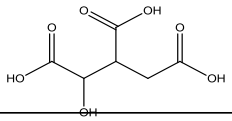
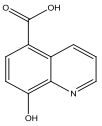
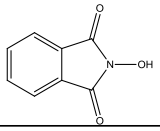
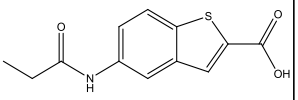
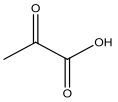
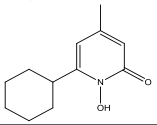
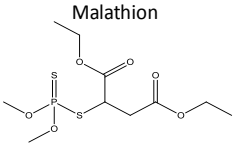
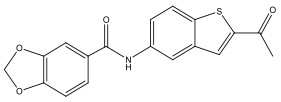
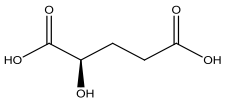
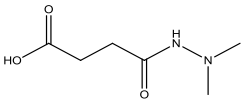
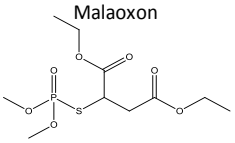
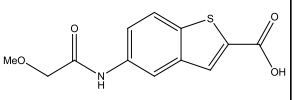
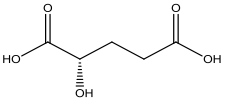
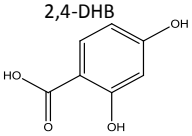
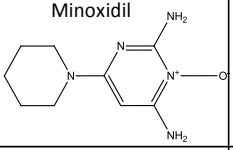
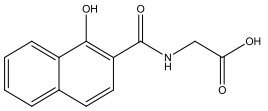
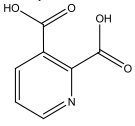
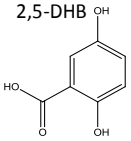
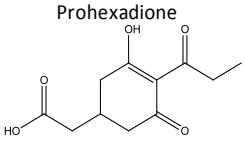
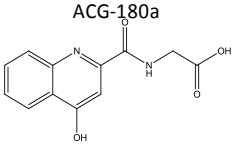
Succinate 	2,4-PDCA 	3,4-DHB 	Quercetin 
Fumarate 	2,5-PDCA 	Gallate 	SAHA 
Oxaloacetate 	2,6-PDCA 	HQNO 	TSA 
Malate 	3,4-PDCA 	Hydralazine 	ACG-109a 
Citrate 	3,5-PDCA 	3-Hydroxymandelate 	ACG-114c 
Isocitrate 	5-Carboxy-8-hydroxyquinoline 	N-Hydroxyphthalimide 	ACG-137a 
Pyruvate 	Ciclopirox Olamine 	Malathion 	ACG-138a 
D-2-HG 	Daminozide 	Malaoxon 	ACG-144a 
L-2-HG 	2,4-DHB 	Minoxidil 	ACG-148a 
2,3-PDCA 	2,5-DHB 	Prohexadione 	ACG-180a 

Table D1: Structures of 2OG analogues.

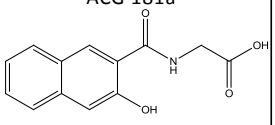
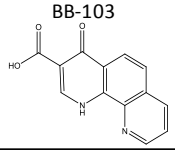
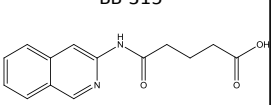
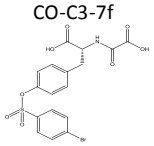
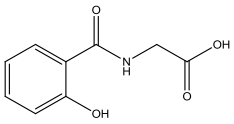
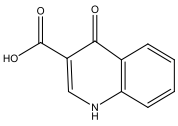
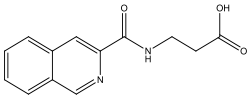
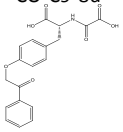
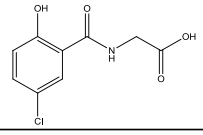
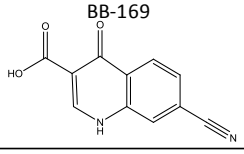
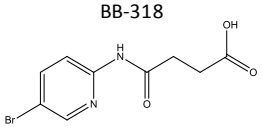
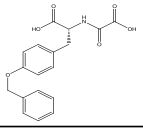
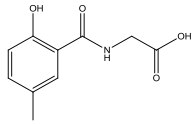
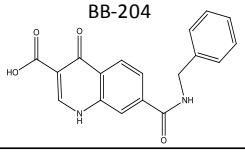
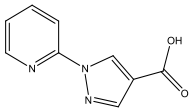
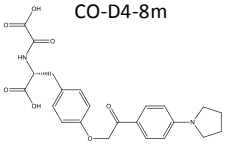
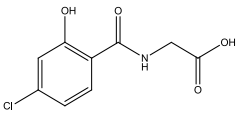
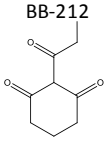
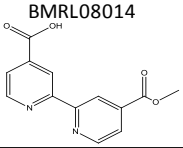
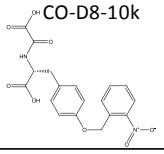
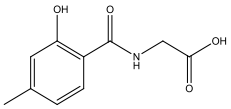
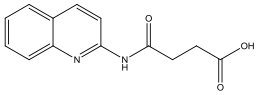
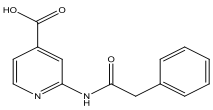
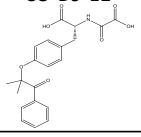
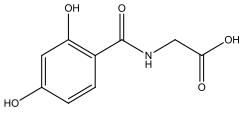
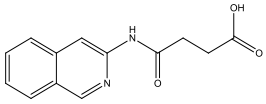
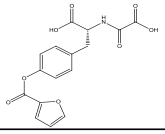
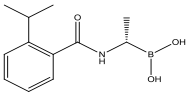
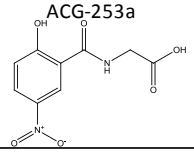
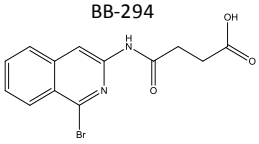
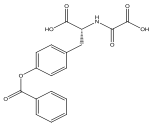
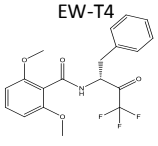
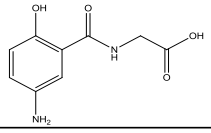
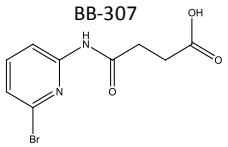
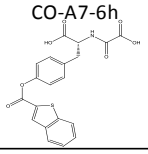
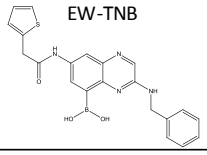
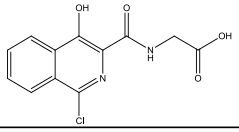
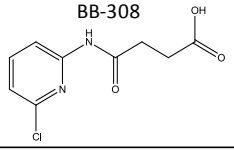
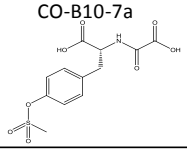
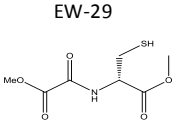
ACG-181a 	BB-103 	BB-315 	CO-C3-7f 
ACG-194b 	BB-110H 	BB-316 	CO-C9-8d 
ACG-223a 	BB-169 	BB-318 	CO-C10-10a 
ACG-226a 	BB-204 	BMRL08012 	CO-D4-8m 
ACG-227a 	BB-212 	BMRL08014 	CO-D8-10k 
ACG-228a 	BB-286 	BMRL08050 	CO-D9-12 
ACG-237a 	BB-287 	CO-A1-6d 	EW-B17 
ACG-253a 	BB-294 	CO-A2-5a 	EW-T4 
ACG-255d 	BB-307 	CO-A7-6h 	EW-TNB 
ACG-263 	BB-308 	CO-B10-7a 	EW-29 

Table D2: Structures of 2OG analogues.

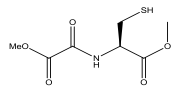
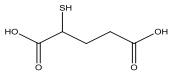
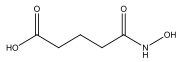
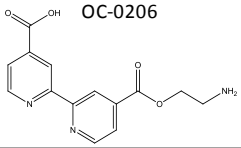
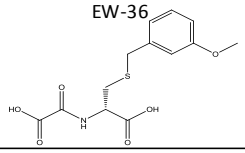
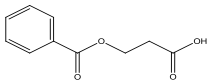
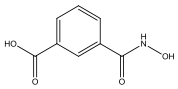
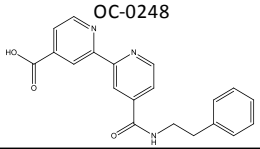
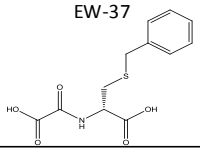
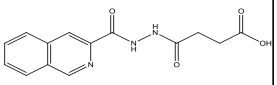
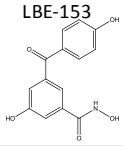
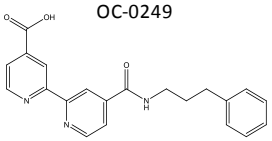
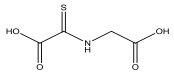
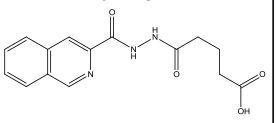
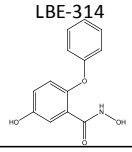
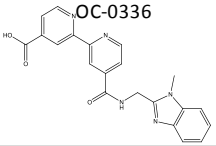
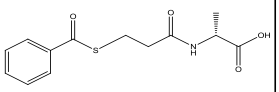
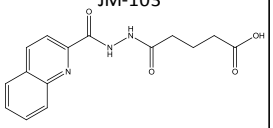
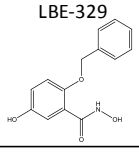
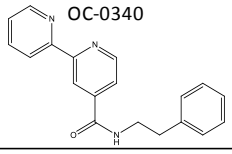
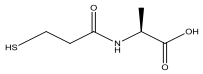
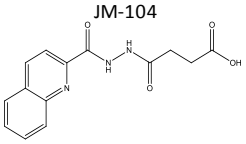
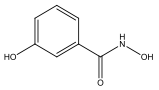
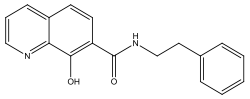
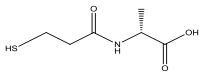
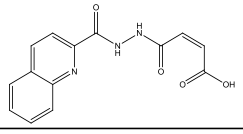
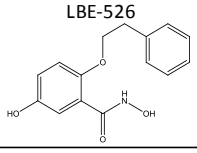
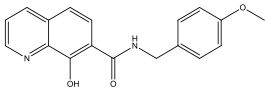
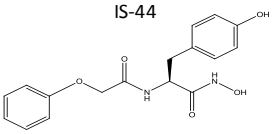
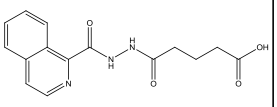
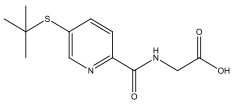
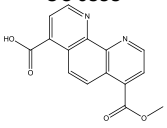
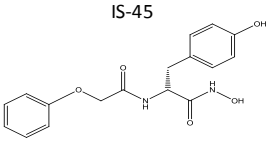
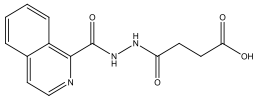
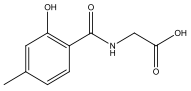
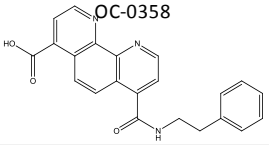
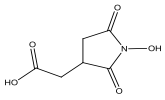
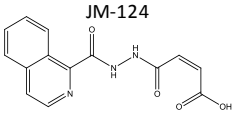
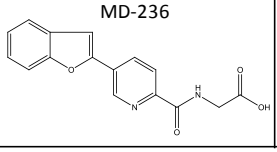
EW-30 	IS-54 	LBE-060 	OC-0206 
EW-36 	IS-84 	LB-096 	OC-0248 
EW-37 	JM-87 	LBE-153 	OC-0249 
IS-3 	JM-102 	LBE-314 	OC-0336 
IS-21 	JM-103 	LBE-329 	OC-0340 
IS-37 	JM-104 	LBE-332-6-3 	OC-0344 
IS-38 	JM-106 	LBE-526 	OC-0345 
IS-44 	JM-109 	MD-104 	OC-0353 
IS-45 	JM-110 	MD-228 	OC-0358 
IS-52 	JM-124 	MD-236 	

Table D3: Structures of 2OG analogues.

E: Publications in Peer-Reviewed Journals

Some of the work described in this thesis has been published in peer-reviewed journals, and the following pages contain copies of papers arising from these studies:

Church C, Lee S, **Bagg EAL**, McTaggart JS, Deacon R, Gerken T, Lee A, Moir L, Mecinović J, Quwailid MM, Schofield CJ, Ashcroft FM, Cox RD. A mouse model for the metabolic effects of the human fat mass and obesity associated FTO gene. *PLoS Genet.* **5** (8) e1000599 (2009).

Krylova SM, Karkhanina AA, Musheev MU, **Bagg EAL**, Schofield CJ, Krylov SN. DNA Aptamers as Analytical Tools for the Quantitative Analysis of DNA-Dealkylating Enzymes. *Anal. Biochem*. **414** (2) 261-5 (2011).

Chowdhury R, Yeoh KK, Tian Y-M, Kawamura A, Hillringhaus L, **Bagg EAL**, Rose NR, Leung I, Woon E, Yang M, McDonough M, King O, Clifton I, Claridge T, Klose R, Ratcliffe P, Schofield CJ. The Oncometabolite 2-Hydroxyglutarate Inhibits Human 2-Oxoglutarate Oxygenases with Varying Potencies. *EMBO Reports*. **12** (5), 463-9 (2011).

Woon ECY, Demetriades M, **Bagg EAL**, Aik W, Krylova SM, Ma JHY, Chan M, Walport LJ, Wegman DW, Dack KN, McDonough MA, Krylov SN, Schofield, CJ. Dynamic Combinatorial Mass Spectrometry Leads to Inhibitors of a 2-Oxoglutarate Dependent Nucleic Acid Demethylase. *J Med Chem.* **55** (5), 2173-84 (2012).

Krylova SM, Koshkin V, **Bagg EAL**, Schofield CJ, Krylov SN. Mechanistic Studies on the Application of DNA Aptamers as Inhibitors of 2-Oxoglutarate-Dependent Oxygenases. *J Med Chem.* **55** (7), 3546-52 (2012)

Thalhammer A, Aik W, **Bagg EAL**, Schofield CJ. The Potential of 2-Oxoglutarate Oxygenases Acting on Nucleic Acids As Therapeutic Targets. *Drug Discovery Today: Therapeutic Strategies.* Accepted. (2012).