STUDIES ON FIH INTERACTIONS

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

Thomas Markus Leissing
Hertford College, Oxford
Michaelmas Term 2017

Thomas Leissing
Hertford College

Abstract

DPhil
Michaelmas Term 2017

The ability to adapt to variations in oxygen levels is essential for higher organisms. In animals, the chronic response to hypoxia (limiting oxygen availability) is substantially regulated by the hypoxia inducible factors (HIFs), which are α,β- heterodimeric transcription factors. The activity and stability of HIF-α subunits are regulated by 2-oxoglutarate (2OG) dependent oxygenases. When sufficient oxygen is present (normoxia), prolyl-4-hydroxylation by the prolyl hydroxylase domain containing proteins (PHDs) signals for HIF-α degradation by promoting binding to the von Hippel-Lindau (VHL) tumour suppressor protein, an E3 ubiquitin ligase. In hypoxia, PHD activity slows and the HIF-α subunits are stabilised leading to increases in HIF-α levels, heterodimerisation with HIF-1β, and consequent increased transcription of HIF target genes (*e.g*. erythropoietin, EPO). Factor Inhibiting HIF (FIH), is an oxygenase that, at least in higher animals, regulates HIF-1/2α activity by catalysing C-3 hydroxylation of a single asparagine in the HIF-α transcriptional activation domain (CAD), thereby blocking the binding of HIF to the CBP/p300 transcriptional coactivators.

Pharmaceutical companies have developed PHD inhibitors to increase endogenous EPO levels *via* HIF-α stabilisation, that are in late stage clinical trials for anaemia treatment in chronic kidney disease. Using FIH as a PHD surrogate, the structural basis for PHD inhibition by clinical candidates was studied. Together with solution studies, the crystallographic results reveal PHD inhibition occurs *via* 2OG competition as well as by varying degrees of substrate competition.

In addition to HIF-α, FIH interacts with multiple ankyrin repeat domain (ARD)-containing proteins, many, but not all, of which are hydroxylated by FIH. The physiological roles of FIH-mediated ankyrin hydroxylation are not fully understood, but ARD proteins are able to compete with HIF for FIH activity and hence may contribute to the regulation of the hypoxic response. To screen for small molecules that could potentially lead to a substrate specific inhibitor, blocking FIH-ARD interactions, thus promoting FIH mediated suppression of HIF transcriptional activity, a crystallographic fragment screen was established and performed for FIH; several small molecules binding to the substrate binding site of FIH were identified.

FIH-ARD interactions were studied for the ARDs of Transient Receptor Potential (TRP) ion channels. Five crystal structures of FIH in complex with TRP channel derived peptides were obtained; analyses of their structures reveal differential modes of FIH substrate binding. Attempts were made to obtain structures of FIH-ARD protein-protein complexes included re-design of the FIH crystallisation construct and disulfide cross-linking approaches.

Evidence was accrued for a novel FIH ARD hydroxylation, *i.e*. the double hydroxylation of two asparagine residues within a single ARD, in the case of Apoptosis Stimulating Proteins of p53 (ASPP) and with related ARD proteins. Potential effects of FIH-mediated hydroxylation on ASPP2 protein function were investigated using interaction studies in cells.

To better understand the mechanisms of 2OG oxygenases, methodologies for X-ray Electron Free Lasers (XFEL) measurement of 2OG oxygenases and related enzymes were developed and the first XFEL room-temperature structures for 2OG oxygenases were achieved.

Overall, the results extend the scope of FIH catalysis, provide new insights into hydroxylase inhibition, and identify XFEL as a useful technique for studying 2OG oxygenase mechanisms.

Für meine Eltern

Acknowledgements

First of all, I want thank my supervisors Prof. Chris Schofield and Prof. Xin Lu for giving me this exciting, challenging and rewarding project and their support, guidance and help throughout my studies.

I would like to thank Dr Richard Hopkinson, Mr Ben Saward, Mr Sam Cahill, Dr Becky Hancock, Dr Adam Hardy, Dr Christopher Lohans, Dr Kundan Sharma, Dr Ludovico Buti, Dr Ian Clifton and Dr Jurgen Brem for proof-reading the thesis, scientific discussions and support throughout the write-up.

Further I am grateful to Dr Rashed Chowdhury, Dr Michael McDonough and Dr Ian Clifton for teaching me the tips and tricks of macromolecular crystallography. Further, I would like to acknowledge Dr Adam Hardy for introducing me to protein purification and enzymology. I would like to thank Mr James Chan, Ms. Jingyi Ma, Dr Patrick Rabe, Mr Jos Kamps, Dr Tzu-Lan Yeh, Dr Martine Abboud and Ms. Anka Lucic for fruitful collaborations. Finally, I want to thank every past and present member of the Lu and Schofield lab who have been exceptional helpful and welcoming over the last four years including Hanna Tarhonskaya, Inga Pfeffer, Ilaria Pettinati, Aiste Skorupskaitė, Stefaine Zentelis, Carlos Ruiz and everyone one else who made the lab such an enjoyable place.

The work would not have been possible without the great help from my collaborators. I would like to thank Prof. Sir Peter Ratcliffe for his helpful input. I thank Prof. Benedikt Kessler, Dr Roman Fischer and Dr Rebecca Konietzny for LC MS/MS data collection and help with data analysis, Dr Holger Kramer for his cooperation with MALDI-MS/MS, Dr Anthony Tumber for rapidFIRE measurements and Mr Joseph Donlan (Prof. Carol Robinson Group) for non-denaturing MS measurements of FIH substrate complexes and Sarah Picaud (Prof. Panagis Filippakopoulos group) for ITC measurements. I would also like to thank Dr Allen Orville, Dr Pierre Aller, Dr Agata Butryn (Diamond Light Source), Dr Aaron Brewster and Ms. Iris Young (University of California, Berkeley) and all members of the XFEL proposals LQ12, PO81 and XFEL2017 for their help in sample preparation, data collection and analysis for the XFEL experiments. I am further grateful to Dr Patrick Collins and Dr Roman Talon and the XChem team for their help and support with the high-throughput crystallographic fragment screen and the in-crystal reaction experiments. Additionally, I would like to thank the scientists of beamlines I02, I03, I04, I04-1 and I24 at Diamond Light Source and MASSIF-1 at the ESRF for assistance. I am grateful to the Chemistry and the Ludwig Institute lab support, especially Dr Stanley Ng, Stephen, Jan and Carly for keeping the lab running, re-supplying consumables and an endless supply of LB medium for bacterial protein production. I am also very grateful to the Systems Biology DTC, Ludwig Institute and Chemistry Department for financial support.

I cannot forget to mention the great members of the Oxford University Canoe and Kayak Club and Falcon Rowing and Canoeing Club for friendship and for making my stay in Oxford such a wonderful experience. Especially I would like to mention the “early morning group”: Andy, Annie, Sam, Matt, Thomas and everyone else coming out in the early mornings and particularly Chris for great white-water trips and fantastic races on flat and white water as well as Lizzie and Jeremey for years of great K2ing with dozens of marathon races and Jeremy also for pulling me over the finish line of the DW. I would also like to thank my “old study mates” Caro and Micha for friendship and great support through my undergrad and graduate studies.

Finally, I want to thank my family for their love and continuously support, giving me the strength and courage to follow my dreams.

Contents

Abstract i

Acknowledgements iii

Abbreviations x

1. Introduction 1

1.1 2OG dependent oxygenases 1

1.2 Human 2OG dependent oxygenases 3

1.3 The hypoxic response and HIF 7

1.4 The Factor Inhibiting HIF- FIH 10

1.5 FIH mediated reactions 12

1.6 FIH substrate binding and recognition 16

1.7 Physiological role of FIH mediated hydroxylation 20

1.8 Aim of the Thesis 21

1.9 References 21

2 Small molecule binding to the HIF hydroxylases 31

2.1 Introduction 31

2.1.1 Structural basis for 2OG inhibitors 32

2.1.2 PHD inhibitors in clinical trials 33

2.1.3 Selectivity and activity of clinical PHD inhibitors and choice of a model systems for structural studies. 35

2.1.4 Structural basis for the inhibition of FIH 37

2.2 Structural basis for the inhibition of the PHDs by clinical drug candidates 39

2.2.1 Crystal structures of FIH in complex with GSK1278863 and Molidustat 39

2.2.2 Crystal structures of FIH in complex with Vadadustat and FG-4592 42

2.2.3 Comparison of the binding modes of clinical candidates to FIH with other 2OG oxygenases 48

2.2.4 Models for PHD2 binding and mode of inhibition of clinically used PHD inhibitors 49

2.3 FIH in complex with non-clinical PHD inhibitors 52

2.3.1 Structures of FIH in complex with N-glycinamide test compounds 52

2.3.2 Structure of FIH in complex with IOX4 55

2.4 FIH as a binding model for other 2OG oxygenases 59

2.5 Conclusions and outlook 62

2.6 References 63

3 High-throughput crystallographic fragment screening 66
3.1 Introduction 66
3.1.1 Fragment based drug discovery 66
3.1.2 Crystallographic fragment screen 66
3.2 Optimisation of fragment screen conditions for FIH 67
3.3 Fragment screen against FIH 69
3.3.1 Fragment screen against FIH: Fragments binding to site 1 72
3.3.2 Fragment screen against FIH: Fragments binding to site 2 75
3.3.3 Fragment screen against FIH: Fragments binding to site 3 78
3.3.4 Fragment screen against FIH: Fragments binding to site 4 81
3.3.5 Fragment screen against FIH: Fragments binding further binding sites of FIH 86
3.3.6 Validation and potential follow-up on fragments hits 87
3.3.7 Data collection statistics 91
3.4 In-crystal fragment reactions – initial experiments 97
3.5 Conclusions and outlook 103
3.6 References 104

4 FIH substrate recognition 106
4.1 Introduction 106
4.2 TRPA1 hydroxylation by FIH 106
4.3 Crystal structures of FIH in complex with TRPA1 and TRPV3 derived peptides 109
4.3.1 Crystal structures of FIH in complex with $TRPA1_{313-339}$ and $TRPV3_{220-246}$ 109
4.3. 2 Crystal structures of FIH in complex with $TRPA1_{321-348}$ and $TRPV3_{229-255}$ 112
4.3.3 Crystal structures of FIH in complex with $TRPA1_{144-171}$ and HIF-2α 116
4.4 Comparison of different ankyrin binding modes to FIH 121
4.5 Conclusions and outlook 123
4.6 References 123

5 Towards an FIH-ankyrin protein co-crystal structure 126
5.1 Introduction 126
5.2 Non cross-linked FIH ankyrin crystallisation trials 127
5.2.1 Initial crystallisation trails of FIH-ARD complexes 127
5.2.2 Variations of FIH constructs 129
5.2.3 Crystallisation trials using limited proteolysis 137
5.3 Enzyme-substrate cross-linking approaches for FIH 139

5.3.1 Introduction to protein cross-linking 139
5.3.2 Design of FIH-substrate cross-linking and first peptide structure 141
5.3.3 Cross-linking of FIH with ARD substrates 145
5.3.4 Impact of cross-linking on substrate binding modes 152
5.4 Conclusions and outlook 154
5.5 References 155

6 Scoping the catalytic landscape of FIH mediated reactions 159
6.1 Introduction: The ASPP protein family 159
6.2 FIH mediated hydroxylation of ASPP proteins 163
6.2.1 FIH interacts with ASPP proteins in U2OS cells 163
6.2.2 ASPP proteins interact with FIH and can compete with HIF-α as a substrate 164
6.2.3 FIH mediated hydroxylation of recombinant iASPP and ASPP2 170
6.2.4 FIH mediated hydroxylation of ASPP proteins in U2OS cells 176
6.2.5 FIH mediated hydroxylation of ASPP peptides 184
6.3 FIH mediated hydroxylation of recombinant ASPP and other VNVN protein constructs 194
6.3.1 Selection of human ARD proteins for FIH hydroxylation assays 194
6.3.2 FIH hydroxylation assays of human VNVN proteins 197
6.4 Conclusions and outlook 206
6.5 References 207

7 Biological roles of ASPP2-FIH interactions 211
7.1 Introduction 211
7.2 Proteomic analysis of U2OS WT, U2OS FIH KO and U2OS ASPP2 KO cells 212
7.3 ASPP2-V5 interactomes in U2OS cells 216
7.4 ASPP2-V5 interactome in AGS cells 228
7.5 Endogenous ASPP2 IP 230
7.6 Conclusions and outlook 236
7.7 References 237

8 XFEL structures of 2OG oxygenases 240
8.1 Introduction 240
8.1.1 Introduction: time-resolved structural biology 240
8.1.2 X-ray electron-free laser in macromolecular serial crystallography 241
8.1.3 Sample requirements for XFEL experiments 241
8.1.4 Sample delivery systems for XFEL 241
8.1.5 XFEL studies on 2OG oxygenases 243
8.2 Sample selection and initial crystallisation optimisation 243
8.2.1 Sample selection 243
8.2.2 Initial crystallisation trials 244
8.3 FIH sample preparation and Data collection 250
8.3.1 FIH sample preparation and data collection at the LCLS 250
8.4 AlkB sample Preparation and Data collection 254
8.4.1 AlkB sample preparation 254
8.4.1 AlkB data collection at the LCLS 255
8.5 XFEL data processing from the LCLS 256
8.6 XFEL crystal structures of FIH and AlkB 261
8.6.1 XFEL crystal structures of FIH in the resting and oxidised states 261
8.6.2 XFEL crystal structures of AlkB in the resting and oxidised states 263
8.6.3 Cryo-crystal structures of FIH and AlkB in the resting state and after oxygen exposure 265
8.6.4 Comparison of room temperature crystal structures with cryo- crystal structures 270
8.7 Preliminary experiments at the European XFEL 272
8.7.1 Sample selection for time-resolved crystallography at the European XFEL 272
8.7.2 VIM-2 and BcII micro-crystal optimisation 272
8.7.3 Batch optimisation of XFEL proteins 274
8.7.4 Pre-screening of micro-crystals from FIH, IPNS, VIM-2 and BcII at DLS I-24 276
8.7.5 Data collection of VIM-2 at the European XFEL 277
8.8 Conclusions and outlook 278
8.9 References 279

9 Summary 282

10 Methods and materials 286
10.1 General 286
10.2 Microbiological techniques 287
10.2.1 Plasmid DNA isolation 287
10.2.2 Site directed mutagenesis 287
10.2.3 Agarose gel electrophoresis 289
10.2.4 Molecular sub-cloning 289
10.2.5 Bacterial growth medium 290
10.2.6 Transformations 290
10.2.7 Glycerol Stocks 290
10.3 Recombinant protein production 292
10.3.1 Large scale recombinant protein production and purification 292
10.3.2 SDS-PAGE 294
10.4 Cell culture 295
10.4.1 General 295
10.4.2 Transfections 295
10.4.3 Antibody cross-linking 295
10.4.4 Small-scale immunoprecipitation 296
10.4.5 Large scale immunoprecipitation 297
10.4.6 Immunoblotting 298
10.4.7 In-gel digestions 300
10.5 Hydroxylation assays 301
10.5.1 Peptide hydroxylation assays 301
10.5.2 Protein hydroxylation assays 301
10.5.3 Methanol/chloroform precipitation 301
10.5.4 In-solution digestions 301
10.5.5 SEP-PAK C18 purification 302
10.6 Mass spectrometry 302
10.6.1 MALDI TOF-MS 302
10.6.2 MALDI TOF-MS/MS 302
10.6.3 Intact protein LC-MS 302
10.6.4 Non-denaturing MassSpec 303
10.6.5 LC-MS/MS analysis 303

10.6.6 Data analysis 303
10.7 Crystallography 304
10.7.1 Crystal preparation and handling 304
10.7.2 Sample preparation 304
10.7.3 Data collection and analysis 305
10.7.4 PANDDA analysis 306
10.8 Peptide synthesis 306
10.9 Affinity measurement 307
10.9.1 Surface plasmon resonance (SPR) 307
10.9.2 Isothermal titration calometry (ITC) 307
10.10 References 308

Appendix I

Abbreviations

1mA	1-methyladenine
2,4 PDCA	2,4-pyridinedicarboxylic acid
2OG	2-oxoglutarate
2TY	2-tryptone yeast
3CA	3 consensus ankyrin
ADE	acoustic droplect ejection
AHR	aryl hydrocarbon receptor
AlkBH	alkylated DNA repair protein alkB homolog
APCL	adenomatous polyposis coli protein 2
ARD	ankyrin repeat domain
ARNT	aryl hydrocarbon receptor nuclear translocator
ASB11	ankyrin repeat and SOCS box protein 11
ASPH	aspartyl/asparaginyl beta-hydroxylase
ASPP1	apoptosis-stimulating of p53 protein 1
ASPP2	apoptosis-stimulating of p53 protein 2
BBOX1	gamma-butyrobetaine dioxygenase
BcII	MBL-2 from *Bacillus cerus*
BDC	background density correction factor
bHLH	basic-helix-loop-helix
BPA	p-benzoyl-L-phenylalanine
BS3	bis-sulfosuccinimidyl
CA	consensus ankyrin
CA9	carbonic anhydrase 9
CAD	C-terminal transactivation domain
CagA	cytotoxin-associated gene A
CAS	clavaminate synthase
CAV1	caveolin-1
CBP	CREB binding protein
CDK	chronic kidney disease
CDN2C	cyclin-dependent kinase 4 inhibitor C
CHCA	α-cyano-4-hydroxycinnamic acid
ChIP	chromatin immunoprecipitation
CP4H	collagen prolyl 4-hydroxylase
CRISPR	clustered regularly interspaced short palindromic repeats
cryo-EM	cryo-electron microscopy
DAOCS	deacetoxycephalosporin C synthase
DCG	2,2-dicholoroglutarate
DCMS	dynamic-combinatorial mass spectrometry
DIC	N,N′-diisopropylcarbodiimide

DLE	D-leucine
DLS	Diamond Light Source
DMEM	dulbecco's modified eagle medium
DMOG	dimethyloxaloylglycine
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DSBH	double-stranded β-helix
DTT	dithiothreitol
DYL	D-allyl-glycine
e.g.	*exempli gratia*
E6AP	E6-associated protein
ECL	enhanced chemilumescent
EDTA	ethylenediaminetetraacetic acid
EE	ethylene glycol
EPO	erythropoietin
ESRF	European Synchrotron Radiation Facility
FBDD	fragment based drug discover
FBS	fetal bovine serum
FIH	factor inhibiting HIF
FRMD6	FERM domain-containing protein 6
FTO	fat mass and obesity-associated protein
GABPB2	GA-binding protein subunit beta-2
GSK	GlaxoSmithKline
GST	glutathione S-transferease
HIF	hypoxia inducible factor
HOBt	hydroxybenzotriazole
HRP	horseradish peroxidase
HTS	high-throughput screening
i.e.	*id est*
iASPP	inhibitor of apoptosis-stimulating protein of p53
IP	immunoprecipitation
IPNS	isopenicillin N synthase
IPTG	isopropyl β-D-1-thiogalactopyranoside
ITC	isothermal titration calorimetry
JmjC	Jumonji-C
JmjD5	JmjC domain-containing protein 5
JmjD6	JmjC domain-containing protein 6
JmjD7	JmjC domain-containing protein 7
KDM	lysine specific demethylase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	lysogen broth
LC	liquid chromatography
LCLS	Linac Coherent Light Source

LCP	lipid cubic phase
LIFT	laser induced fragmentation
LRRK1	leucine-rich repeat serine/threonine-protein kinase 1
MALDI	matrix-assisted laser-desorption ionization
MBL	metallo-β-lactamases
MFX	macromolecular femtosecond crystallography
MINA53	MYC-induced nuclear antigen 53
mir	micro-RNA
MRS	multiple resistance strains
MS	mass spectrometry
MYO16	unconventional myosin-XVI
NMR	nuclear magnetic resonance
NO66	nucleolar protein 66
NOFD	N-oxalyl-D-phenylalanine
NOG	N-oxalylglycine
Notch	neurogenic locus notch homolog protein
NPAS	neuronal PAS domain proteins
ODD	oxygen dependent degradation domains
OGFOD1	2-oxoglutarate and iron-dependent oxygenase domain-containing protein 1
OUTD7	OTU domain-containing protein 7A
OXPHOS	oxidative phosphorylation
PAGE	polyacrylamide gel electrophoresis
PAHX	phytanoyl-CoA hydroxylase
PANDDA	pan-dataset density analysis
PARD3	partitioning defective 3 homolog
PAS	Per ARNT Sim
PBS	phosphate-buffered saline
PCA	principal component analysis
PCM1	pericentriolar material 1 protein
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethyelene gycol
PHD	prolyl hydroxylase domain containing proteins
PP1	serine/threonine-protein phosphatase
PTM	post-translational modification
RASSF7	Ras association domain-containing protein 7
RASSF8	Ras association domain-containing protein 8
rmsd	root-mean-square deviation
RNA	ribonucleic acid
SA	sinapinic acid
SDA	succinimidyl 4,4’-azipentanoate
SDS	sodium dodecyl sulfate

SEC	size exclusion chromatography
SFX	serial femtosecond crystallography
SPR	surface plasmon resonance
TAZ	transcriptional coactivator with PDZ binding motif
TCEP	tris(2-carboxyethyl)phosphine
TEMED	N,N,N′,N′-tetramethylethylenediamine
TET	ten eleven translocation
TEV	tobacco etch virus
TFA	trifluoroacetic acid
THC	delta$_9$-tetrahydrocanabinol
TIPS	triisopropylsilane
TOF	time-of-flight
tRNA	transfer RNA
TRP	transient Receptor Potential
TYW5	tRNA hydroxylase tRNA wybutosine-synthesizing protein 5
UV	ultraviolet
VEGF	vascular endothelial growth factor
VHL	von-Hippel-Lindau
VIM-2	Verona-Integron-Encoded MBL-2
WADA	World Anti-Doping Agency
WB	Western blot
WT	wild-type
XFEL	X-ray free-electron laser
YAP1	yes-associated protein 1

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

1. Introduction

1.1 2OG dependent oxygenases

2-Oxoglutarate and ferrous iron dependent oxygenases (2OG oxygenases) are amongst the most catalytically versatile of enzymes[1]. In animals, 2OG oxygenases have only been reported to catalyse hydroxylation reactions[2-11], including demethylation reactions proceeding *via* initial hydroxylation[12-22]. In microbes and plants, the versatility of reactions catalysed by 2OG oxygenases is much greater. Reactions catalysed by 2OG oxygenases in these organisms include: halogenation[23,24], demethylation[25,26], hydroxylation, sequential oxidation[27], desaturation[28], epimerisation[29], epoxidation[30], endoperoxide formation[31], ring formation[32], and ring expansion reactions[33,34].

Pioneering work on the biochemical characterisation of 2OG oxygenases was performed on the collagen prolyl hydroxylases and related lysyl hydroxylases. This work identified the need for the cofactor Fe(II), the use of 2OG as a cosubstrate and a stimulation of the reaction by ascorbate[35-39]; note that ascorbate dependence of 2OG oxygenases is not fully understood and it is enzyme dependent[40,41]. Despite large efforts, to date, no high-resolution structure of the catalytic domain of human collagen prolyl hydroxylases has been achieved. The first successful structural studies on 2OG oxygenases were in the field of antibiotic biosynthesis.

A milestone in the structural characterisation of 2OG oxygenases was the solution of an X-ray crystal structure of isopenicillin N synthase (IPNS) from *Aspergillus nidulans*[42] (Figure 1A). Although not a 2OG dependent oxygenase, IPNS is structurally and mechanistically related to the 2OG oxygenases. IPNS catalyses the oxidative ring closure of a tripeptide to form the bicyclic penicillin scaffold (Figure 1D)[42]. Surprisingly, it was found that IPNS contains a core double-stranded β-helix (DSBH) fold, a sub-class of the cupin-8 fold[42]. The DSBH fold incorporates a facial triad (also HxD/E…H motif) which binds the non-heme ferrous iron (Figure 1). Following from the structural studies on IPNS, a high resolution X-ray crystal structure of deacetoxycephalosporin C synthase from *Streptomyces clavuligerus* was the first structure of a 2OG oxygenase to be solved[43]. Like IPNS, DAOCS contains a core DSBH fold that incorporates the iron binding motif. The first substrate structure of a 2OG oxygenase was solved for clavaminate synthase (CAS) from *Streptomyces clavuligerus*[44].

The first human 2OG oxygenase structure to be solved was of the oxygen sensor Factor Inhibiting Hypoxia Inducible Factor (FIH)[45,46], that work is further discussed below.

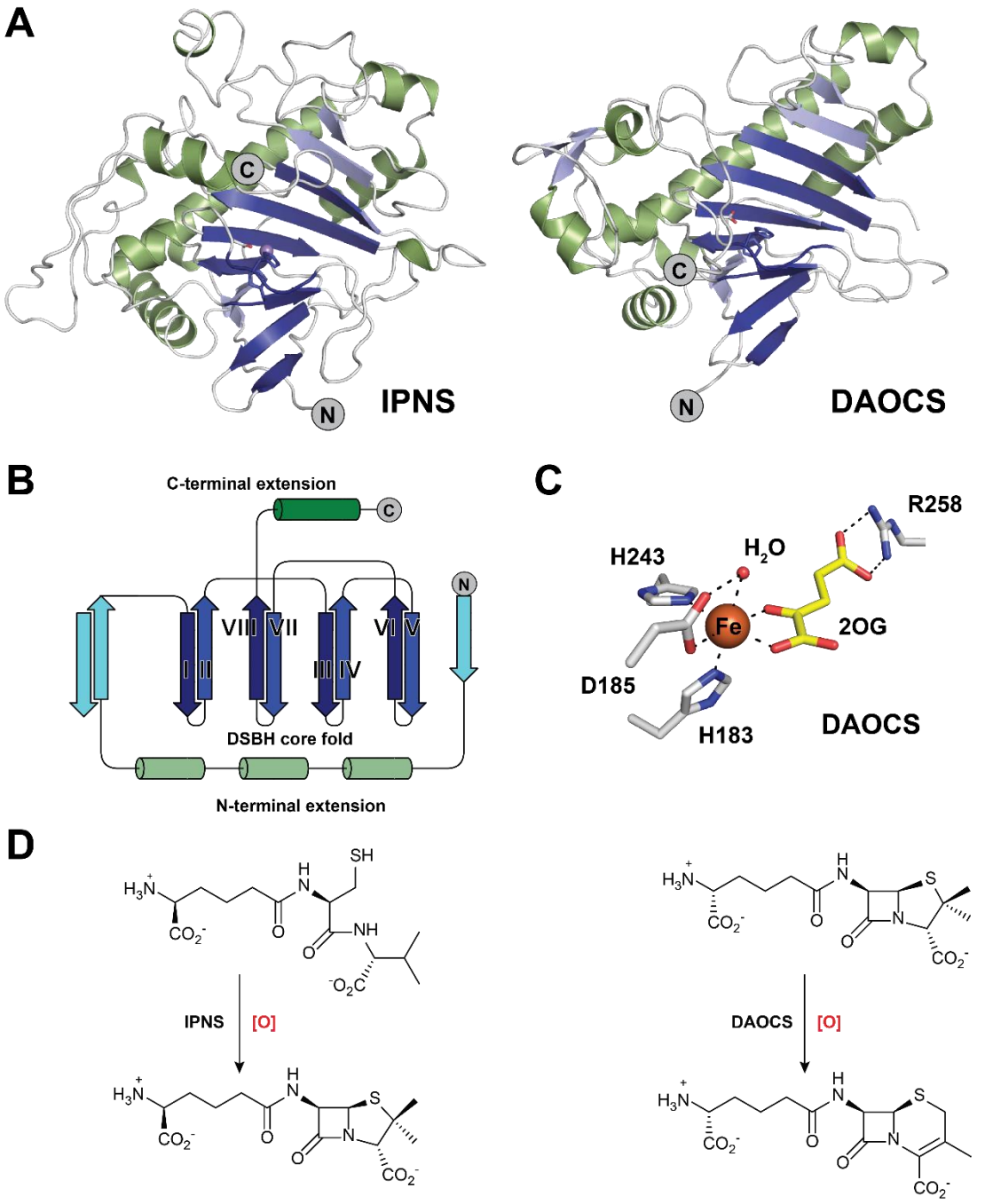


Figure 1 Overview of crystal structures of 2OG dependent oxygenase and related enzymes. **A** Views from crystal structures of IPNS (2.50 Å, PDB code 1IPS[42]) and DAOCS (1.30 Å, PDB code 1DCS[43]) illustrate the overall conserved fold of the two enzymes. The central DSBH fold is coloured in dark-blue. **B** 2 dimensional topology of 2OG oxygenases. The image was created using Topdraw[47]. **C** Views from a crystal structure of DAOCS in complex with iron and 2OG illustrates the conserved facial triad binding the central metal and 2OG (2.1 Å, PDB code 1E5I[48]). **D** The reactions catalysed by IPNS and DAOCS.

1.2 Human 2OG dependent oxygenases

More than 60 human 2OG oxygenases have been characterised[49]. Although 2OG oxygenases have only been reported to catalyse hydroxylation reactions in animals, 2OG oxygenases are involved in several essential cellular processes including collagen biosynthesis[50], fatty acid metabolism[51], DNA repair[52], gene regulation[53-56], protein biosynthesis[57-61] and protein stability[62,63]. Examples of 2OG oxygenase mediated reactions are shown in Figure 2.

Human 2OG oxygenases are proposed to share a largely conserved mechanism (Figure 3). The non-heme iron(II) is coordinated by the characteristic HxD/E…H motif with three coordination sites free for 2OG and oxygen binding. The binding of 2OG followed by substrate binding is believed to weaken a water-metal bond and thereby enable oxygen binding. Following oxidative carboxylation of 2OG to succinate and CO_2, a reactive Fe(IV)=O species is formed. This reacts, under reduction of the Fe(IV) to Fe(II), with the substrate to form an alcohol. In case of the histone demethylases (KDMs) or nucleotide demethylases, a relatively unstable hemiaminal is formed which dissociates to the demethylated product and formaldehyde. Utilising molecular di-oxygen as co-factor for the hydroxylation reactions makes every member of the 2OG oxygenases, at least to some degree, oxygen sensitive, although oxygen dependence varies between different family members[64-68].

Human 2OG oxygenases can be grouped according to sequence and function into: (i) small molecule hydroxylases, including phytanoyl-CoA hydroxylase (PAHX)[10] and gamma-butyrobetaine dioxygenase (BBOX1)[11]; (ii) nucleotide oxygenases, including ten eleven translocation proteins (TETs)[69], Alkylated DNA repair protein alkB homologs (AlkBHs)[52] and fat mass and obesity-associated protein (FTO)[21]; (iii) the prolyl and collagen hydroxylases, including prolyl hydroxylase domain containing proteins 1-3 (PHD 1-3)[5,6], 2-oxoglutarate and iron-dependent oxygenase domain-containing protein 1 (OGFOD1)[60], collagen prolyl 4-hydroxylase 1-3 (CP4H 1-3, see above); and (iv) the Jumonji-C domain-containing (JmjC) oxygenases.

JmjC oxygenases form the largest sub-family of 2OG oxygenases in humans. The JmjC domain originates form the *Jmj* gene in mice (corresponding to human JARID2)[70] which has sequence similarity within the cupin-8 fold[71]. Most of the JmjC oxygenases have been identified as histone demethylases (KDM) sub-classed into 6 subfamilies: KDM 2-7 (Table 1). Some KDMs can also accept methylated arginine residues[22]. All other members of the JmjC oxygenases, with the exception of the tRNA hydroxylase tRNA wybutosine-

synthesizing protein 5 (TYW5)[72], catalyse the hydroxylation of amino acid residues in proteins (referred to as JmjC hydroxylases).

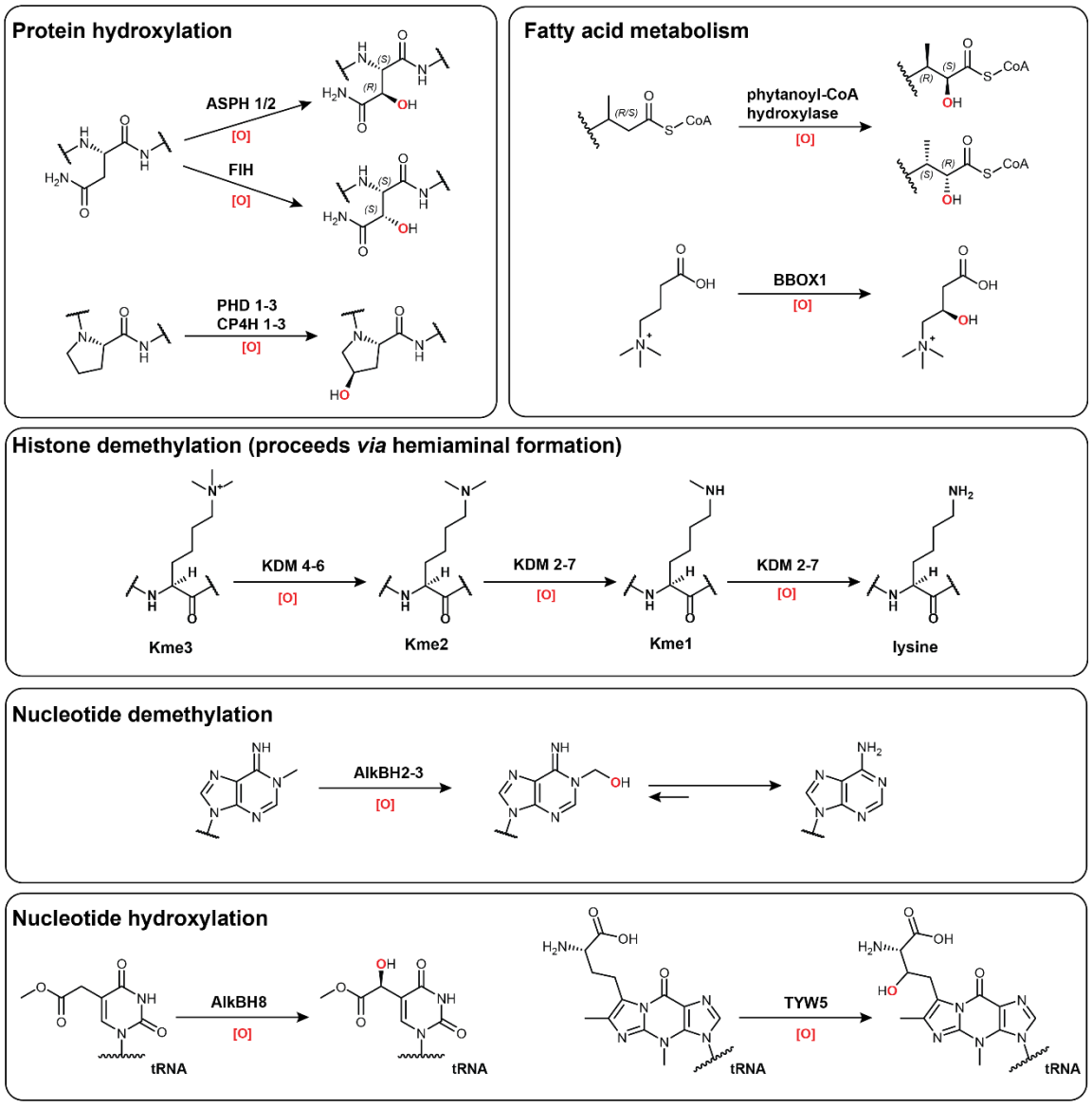


Figure 2 Examples of reactions catalysed by human 2OG oxygenases. Hydroxylation of asparagine residues in proteins is catalysed by the aspartyl/asparaginyl hydroxylase 1/2 (ASPH 1/2) and FIH. Prolyl-4 hydroxylation is catalysed by prolyl hydroxylase domain containing proteins 1-3 (PHD 1-3) and collagen prolyl 4-hydroxylase 1-3 (CP4H 1-3). 2OG oxygenases involved in fatty acid metabolism include phytanoyl-CoA hydroxylase (PAHX) and Gamma-butyrobetaine dioxygenase (BBOX1). Histone demethylation occurs *via* initial hydroxylation followed by spontaneous dissociation of the resulting hemi-aminal. Histone demethylases (KDM) 4-6 accept tri-methyl lysine (Kme3) as substrate; KDM 2-7 accept di- and mono-methyl lysine. Nucleotide demethylation by Alkylated DNA repair protein alkB homolog 2 and 3 (AlkBH2-3) occurs *via* an analogue mechanism as histone demethylation. Nucleotide hydroxylation of transfer RNA (tRNA) is catalysed by Alkylated DNA repair protein alkB homolog 8 (AlkBH8) and tRNA wybutosine-synthesizing protein 5 (TYW5).

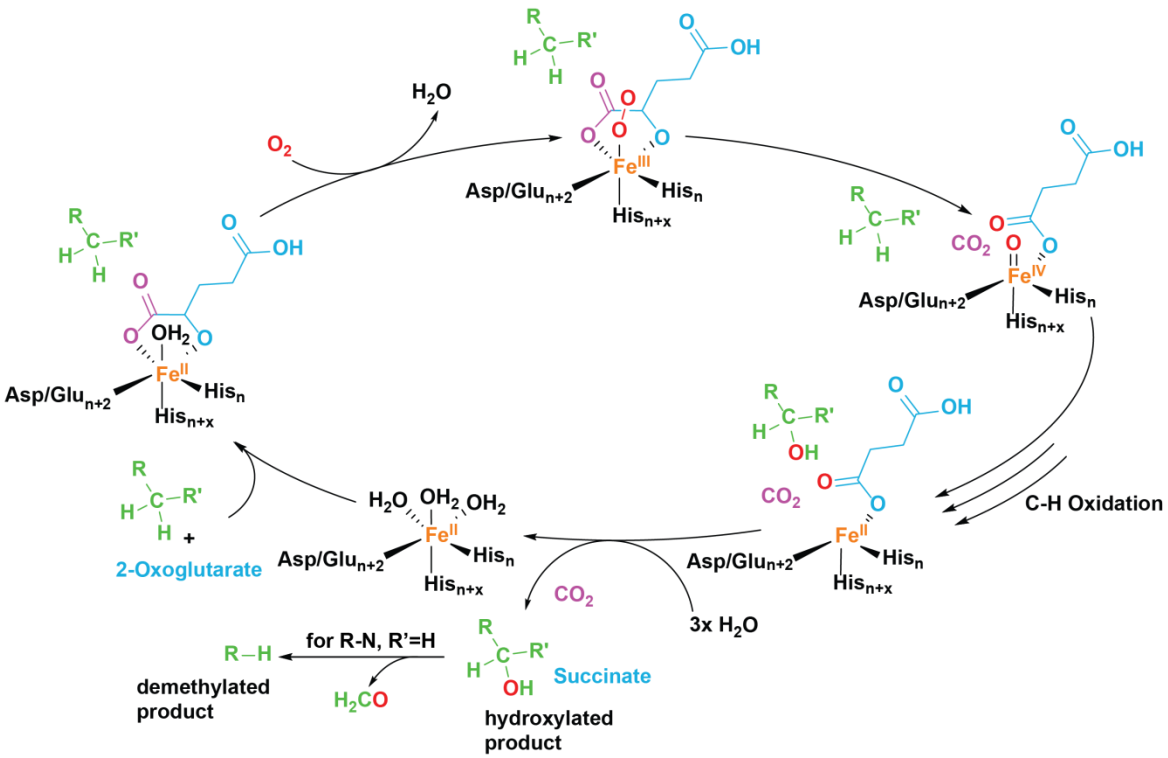


Figure 3 Consensus mechanism for human 2OG oxygenases. 2OG bind to the iron(II), shown in orange, in a bidentate manner with the 2OG C-2 carbonyl *trans*, in the case of JmjC oxygenases, or *cis*, in the case of DAOCS family members, to the acidic reside of the HxD/E…H motif. The substrate is shown in green and oxygen in red and CO_2 in purple. In case of demethylases an unstable hemiaminal is formed which spontaneously leads to the demethylated product. Figure modified from Markolovic and Leissing et al[49] under Creative Commons Attributions License (see Appendix).

In recent years, substantial process has been made in the structural and biochemical characterisation of the JmjC oxygenases. The JmjC oxygenases are the largest subfamily of animal 2OG oxygenases and show the largest substrate versatility[49]. Although there are significant structural differences between JmjC demethylases and hydroxylases, there are some controversial functional assignments for several members, including for JmjD6[7,73-75], JmjD5[76-78], MINA53 and NO66[57,79-82] (Table 1), all of which have been assigned as hydroxylases in Oxford.

The first human JmjC oxygenase to be biochemically and structurally characterised was FIH[2-4,45,46,83]. FIH is to date probably the best studied member of the JmjC hydroxylases. FIH plays an important role as a negative regulator of the transcriptional hypoxic response, as discussed further below.

Table 1 Reported biochemical functions and structures of human JmjC oxygenases. Table modified from Markolovic and Leissing et al.[49]

Family	Subfamily	Name	Catalytic function	Substrates	PDB
Histone demethylases	KDM2	KDM2A	Lys(me2/1) demethylase	H3K36, p65**	4QX8
		KDM2B	Lys(me2/1) demethylase	H3K36, H3K4me3**	---
	KDM3	KDM3A	Lys(me2/1) demethylase	H3K9	---
		KDM3B	Lys(me2/1) demethylase	H3K9	4C8D
		JMJD1C	Lys(me2/1) demethylase	H3K9**, MDC1**	5FZO
	KDM4	KDM4A	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26	2OQ6
		KDM4B	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26	4LXL
		KDM4C	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26; Pc2**	2XML
		KDM4D	Lys(me3/2/1) demethylase	H3K9, H3K36, H1.4K26	5FP4
		KDM4E	Lys(me3/2/1) demethylase	H3K9	---
	KDM5	KDM5A	Lys(me3/2/1) demethylase	H3K4	5E6H
		KDM5B	Lys(me3/2/1) demethylase	H3K4	5FUP
		KDM5C	Lys(me3/2/1) demethylase	H3K4	5FWJ
		KDM5D	Lys(me3/2/1) demethylase	H3K4	---
	KDM6	KDM6A	Lys(me3/2/1) demethylase	H3K27	3AVR
		KDM6B	Lys(me3/2/1) demethylase	H3K27	2XXZ
		KDM6C	Lys(me3/2/1) demethylase	H3K27	5FXV
	KDM7	KDM7A	Lys(me2/1) demethylase	H3K9**, H3K27, H4K20	3KVB
		KDM7B	Lys(me2/1) demethylase	H3K9, H4K20	3KV4
		KDM7C	Lys(me2/1) demethylase	H3K9**, H3K27**, H4K20me3**, ARID5B**	3PUA
Controversial assignments	JmjC -only	JMJD6	Lys (*5S*)-hydroxylase; Arg(me1) demethylase**	Splicing proteins (e.g. U2AF65, LUC7L2), p53,H3R2**; H4R3**	3K2O
		MINA53	His (3S)-hydroxylase; Lys(me3/2) demethylase**	RPL27; H3K9**	4BXF
		NO66	His 3S-hydroxylase; Lys(me3/2) demethylase**	RPL8; H3K4**, H3K36**	4CCO
		JMJD5	Lys(me2) demethylase** Hydroxylase**	H3K36**/ NFATc1**	4GJZ
Hydroxy -lases	JmjC -only	FIH	Asn (3S)-hydroxylase; Asp/His (3S)-hydroxylase	HIF-1α; ARD proteins	1H2K
		JMJD4	Lys C-4 hydroxylase	eRF1	---
		TYW5	Wybutosine β-hydroxylase	$tRNA^{Phe}$ (yW-72)	3AL6
Un-characterised	JmjC -only	HSPBAP1	n/d	n/d	---
		JMJD7	n/d	n/d	4QU2*
		JMJD8	n/d	n/d	---

* murine enzyme

** Controversial functional assignments

1.3 The hypoxic response and HIF

Broad attention was drawn to the field of human 2OG oxygenases when their central role in the oxygen sensing pathway was discovered[5,6,84]. The ability to adapt to variations in ambient oxygen levels is essential for all animals[85,86]. The transcriptional hypoxic response is mediated by the Hypoxia Inducible Factors (HIFs)[87-90], a family of α,β-heterodimeric transcription factors (Figure 4). The HIF transcription factors incorporate a basic-helix-loop-helix (bHLH) DNA binding motif and two Per ARNT Sim (PAS) domains important for dimerisation[91-95]. From the two HIF-subunits only the α-subunit is oxygen sensitive. HIF-1β is the aryl hydrocarbon receptor nuclear translocator (ARNT) and is constitutively expressed[96]. In addition to HIF-α, ARNT was shown to interact with several other bHLH/PAS domain containing transcription factors, including the neuronal PAS domain proteins 1-3 (NPAS 1-3) and the aryl hydrocarbon receptor (AHR)[97-99].

In total, three HIF-α isoforms are known in humans, of which HIF-1/2α are transcriptional activators; HIF-3α, and splicing isoforms thereof, are transcriptional repressors[100]. HIF-1α and HIF-2α are differentially expressed and have complementary rather than redundant functions[101-103].

Under normoxic conditions (*i.e*. normal oxygen levels), HIF-1/2α are hydroxylated at three distinct sites; at two proline residues within the oxygen dependent degradation domains (ODDs) of HIF by the PHDs, and at an asparagine residue in the C-terminal transactivation domain by FIH[2-4,83]. Prolyl hydroxylation by the PHDs leads to the binding of the von-Hippel-Lindau (VHL) tumour suppressor, an E3 ubiquitin ligase. Ubiquitylation of HIF-α targets the protein for proteasomal degradation by the 26S proteasome[5,6,104-106]. FIH mediated asparagine hydroxylation within the C-terminal transactivation domain does not affect HIF-α protein stability but impairs the ability of HIF-α to recruit transcriptional co-activator p300/CBP and thereby creates a second layer of regulation of HIF target gene expression. Notably, only a subset of HIF target genes are affected by FIH activity[107-110].

Under hypoxic conditions (*i.e*. low oxygen levels), hydroxylation cannot occur efficiently and HIF-α is stabilised[5,6,111]. Stabilised HIF-α can translocate to the nucleus, dimerise with HIF-1β and induce the transcription of genes important for the hypoxic response (see above). Genes upregulated under acute hypoxia include genes involved in hormonal regulation, energy metabolism, angiogenic signalling, vasomotor regulation, cell migration, matrix and barrier function, transport, transcriptional regulation and growth and apoptosis[86].

All 2OG dependent oxygenases utilise molecular oxygen. For the HIF-hydroxylases, extensive biochemical analyses have been performed to investigate their oxygen sensitivity[64-68]. PHD2, which is believed to be the primary oxygen sensor in humans, shows a very high K_M for oxygen (>400 μM)[64] which makes it sensitive to variations of oxygen levels within the physiological range. FIH on the other hand was shown to have an oxygen K_M of ~100 μM which makes the enzyme significantly less sensitive to variations in physiological oxygen levels[112]. This raises the question of whether FIH is a real oxygen sensor in the HIF pathway or rather a secondary oxygen sensor to "fine-tune" the hypoxic response[109]. It is notable that the HIF hydroxylases are members of two different sub-families of 2OG oxygenases; PHDs are members of the DAOCS subfamily while FIH is a member of the JmjC subfamily.

Subsequent studies on JmjC KDMs have shown high oxygen K_M values for some KDMs which may make them, at least to some degree, oxygen sensitive under physiological conditions[113] and which has the potential to add another HIF independent layer to the transcriptional hypoxic response.

Modulating the hypoxic response is of pharmaceutical interest. The HIF-mediated up-regulation of erythropoietin (EPO) and thereby increased number in erythrocytes was the first phenotype of the hypoxic response observed in humans and led ultimately to the discovery of the HIF-transcription factors[114]. The discovery of PHDs opened the possibility of mimicking hypoxia and upregulating endogenous EPO for treatment of anaemia[115]. Several drug candidates are, to date, in clinical trials for anaemia treatment in patients with chronic kidney disease. Mechanisms of PHD inhibition is further discussed in chapter 2.

A second aspect of the hypoxic response is the change in metabolism. HIF was found to inhibit oxidative phosphorylation (OXPHOS) while up-regulating genes involved in glycolysis[116]. This is important for cancer treatment as HIF is frequently found to be stabilised under normoxic conditions in cancer leading to a normoxic switch in metabolism, also known as the Warburg effect[117]. The change in metabolism is believed to lead to radio-resistance of certain cancer types, limiting the success of radiotherapy[118,119].

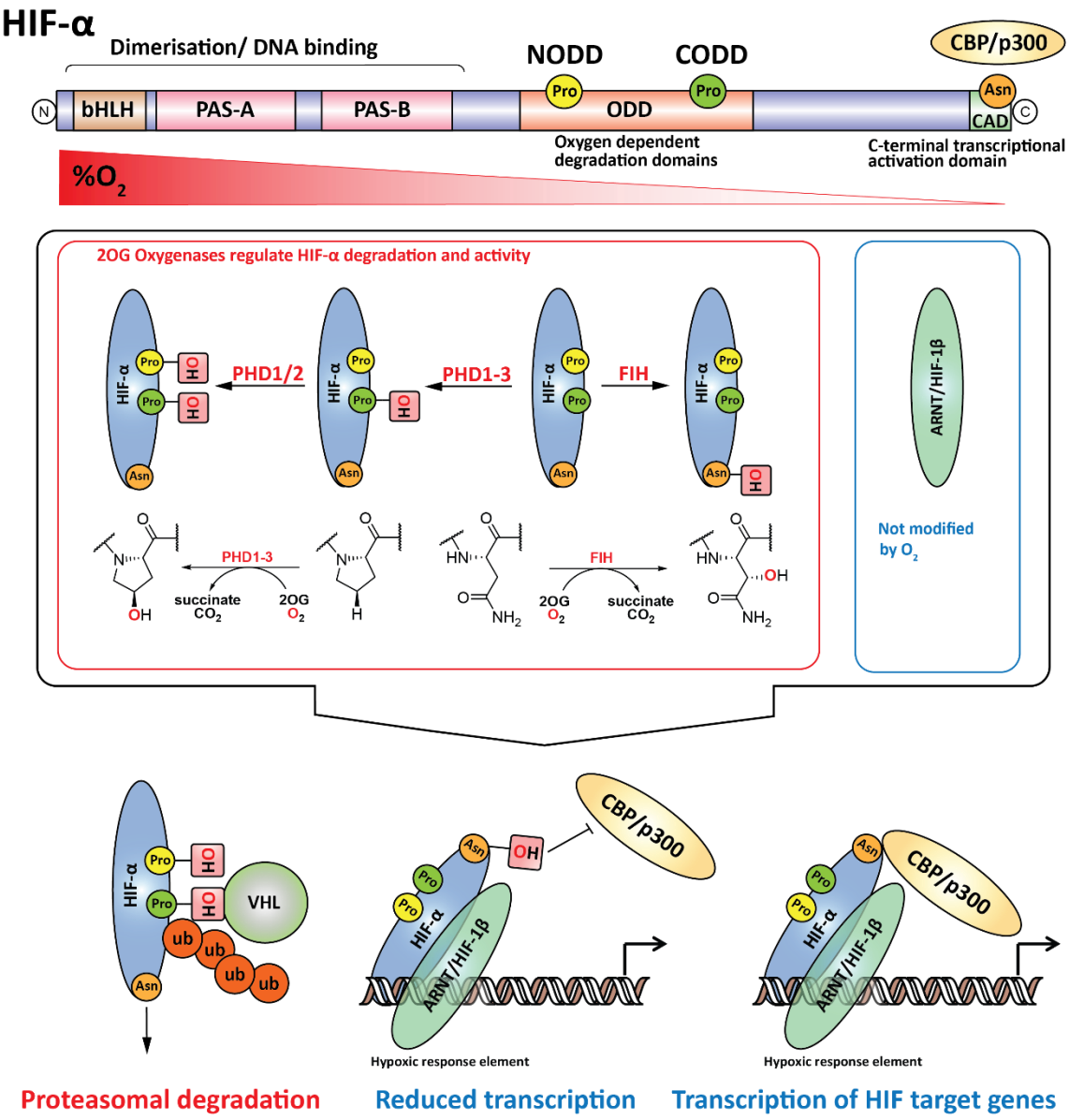


Figure 4 Regulation of the transcriptional hypoxic response by 2OG oxygenases. Human HIF-1/2α, but not HIF-1β, incorporate three hydroxylation sites within their oxygen dependent degradation domains and C-terminal transcriptional activation domains. PHD1-3 accept the C-terminal degradation domain as substrates whereas only PHD1/2 accept the N-terminal degradation domain as substrate. Hydroxylation on either of the ODDs enhances binding of the Von-Hippel-Lindau (VHL) tumour suppressor, part of an E3 ubiquitin ligase complex, leading to ubiquitylation and subsequent proteasomal degradation of the HIF-α subunit. FIH mediated asparagine hydroxylation of the C-terminal transcriptional activation domain of HIF-1/2α blocks the interaction with CBP/p300 transcriptional co-activators suppressing the expression of a subset of HIF target genes. Figure modified from Islam et al, *in press*.

1.4 The Factor Inhibiting HIF- FIH

Following the discovery of its role in the HIF pathway, several structural and biochemical studies have been performed on FIH, the archetypical JmjC hydroxylase. FIH is encoded by the *hif1an* gene with a primary sequence of 349 amino acids[83], comparable to other JmjC oxygenases like TYW5 (315 amino acids), JmjD7 (316 amino acids) and JmjD6 (406 amino acids). The JmjC hydroxylases are generally shorter than JmjC demethylases, which often contain subsidiary non-catalytic domains important for substrate recognition[49].

FIH has a distorted DSBH fold comprising 8 anti-parallel β-sheets (I-VIII) containing also the HxD/E…H motif for iron(II) and 2OG binding (Figure 5A), further discussed below. The proximal histidine is located on the β-II strand and the aspartate at the β-II/β-III loop. The distal histidine is located on the β-VII strand, spatial adjacent to the β-II strand (Figure 5B). As it is typical for human JmjC oxygenases, the basic residue binding the C-5 carboxylate of 2OG is a lysine, lysine-214 for FIH, located at the top of the β-IV strand (Figure 5A).

The DSBH fold of FIH is further extended by N-terminal parallel and anti-parallel β-sheets at the “bottom” and “top” of the DSBH, which is frequently observed for JmjC oxygenases[49]. An often observed feature for JmjC hydroxylases is the presence of an oligomerisation domain, as seen in FIH, JmjD6, TYW5, MINA53 or NO66 (Figure 5B,C). Oligomerisation domains have not been observed for histone demethylases[49]. JmjD6 shows an overall complicated oligomerisation pattern with different oligomeric states observed in solution[120-123] whereas FIH, TYW5, NO66 and MINA53 dimerize (Figure 5C)[45,46,81,124,125]. FIH incorporates a C-terminal dimerisation domain consisting of two helices forming a bundle that mainly interacts *via* hydrophobic interactions (Figure 5B).

FIH forms, at least in the crystal state, a dimer with rotational symmetry through the dimerisation domain (Figure 5B). The active sites of both proteins of the dimer are facing in the same direction and both proteins can bind substrates[46]. The dimerisation domain is involved in substrate binding and FIH mutants with impaired dimerisation are inactive[124]. Cross-talk between two substrates bound to one FIH dimer has not been observed but appears possible due to spatial proximity and is further discussed below.

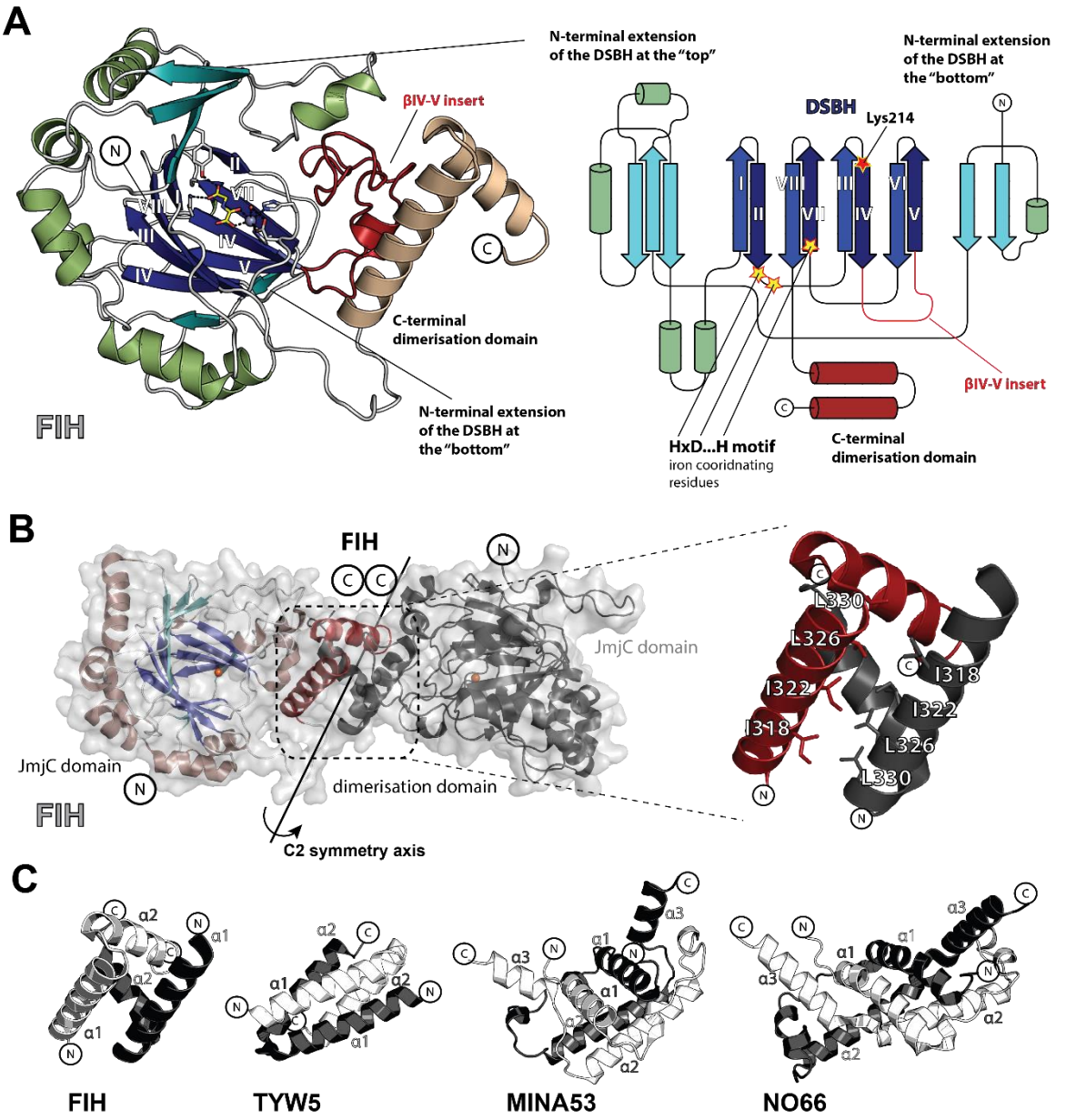


Figure 5 Structure of FIH A Left: View from a crystal structure of FIH (2.15 Å, PDB code 1H2K[46]) illustrating its overall fold and domain structure. The DSBH core fold is coloured in dark-blue, the βIV-V insert is coloured in red and the dimerisation domain is coloured in brown. Right: 2D representation of FIH; image generated using Topdraw[47]. **B** View from a crystal structure of FIH illustrates the dimeric structure of FIH. The C-terminal helices form a bundle *via* predominantly hydrophobic interactions. At least in the crystal state, the FIH dimer manifests a perfect rotational symmetry around a C_2-symmetry axis. **C** Comparison of the dimerisation domains of FIH with TYW5 (2.80 Å, PDB code 3AL6[125]), MINA53 (2.05 Å, PDB code 4BXF[81]) and NO66 (2.30 Å, PDB code 4CCO[81]). Images from **B** and **C** are modified from Markolovic and Leissing et al[49].

1.5 FIH mediated reactions

FIH was original identified as a HIF binding protein[83]. Following the identification of FIH as a 2OG oxygenase, FIH was assigned as an asparagine hydroxylase, catalysing hydroxylation of the β-carbon of asparagine to form (*S,S*) hydroxyl asparagine (Figure 6)[2-4]. Notably, the diastereomer of the product formed by ASPH, the only other asparagine hydroxylase known to date in humans[126].

In addition to the HIF transcription factors, FIH was subsequently shown to catalyse hydroxylation of a large number of ankyrin repeat domain (ARD) containing proteins[8,108,127-134]. To date, >30 ARD proteins have been reported to be hydroxylated by FIH to date (Table 2), including non-human ankyrins, such as mouse Notch1 (mNotch1), a consensus ankyrin (CA) protein and viral ARD proteins. Some ankyrin substrates have only been identified using 2OG turnover assays (Table 2), in which the release of radioactive $^{14}CO_2$ from labelled 2OG is measured. The oxidative decarboxylation of 2OG however, does not correlate directly with substrate hydroxylation (Figure 3). Indeed, examples of induced 2OG turnover upon ARD addition to the FIH protein were observed in catalytically inactive FIH mutants[135].

Investigations of FIH mediated hydroxylation of tankyrase-2 revealed that FIH can also accept histidine and aspartate residues *in vivo* within an ankyrin context[8]. Subsequent studies using a consensus ankyrin sequence showed that FIH can also accept serine and even leucine residues *in vitro*[136]. Follow-up reactions on hydroxylated serine and histidine residues were also observed *in vitro* and are summarised in Figure 6. Whether these reactions are FIH catalysed or non-enzymatic is unknown.

Structures of FIH in complex with D-amino acids derived from natural amino acids have also been reported (see below), which suggests an even broader substrate specificity of FIH.

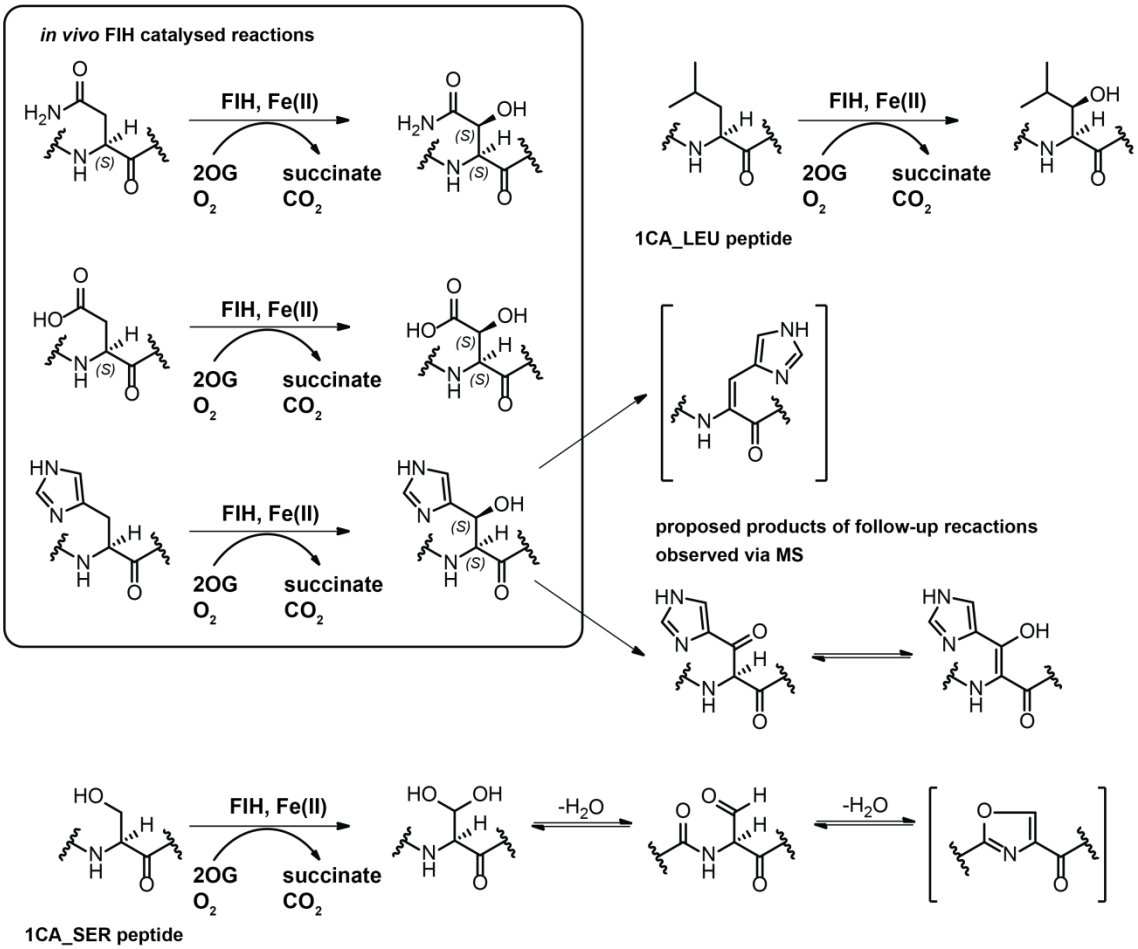


Figure 6 Overview of FIH mediated reactions. FIH mediated hydroxylation reactions observed *in vivo* are highlighted. FIH mediated reactions observed within a consensus ankyrin sequence include the oxidation of leucine and serine, proposed follow-up reaction of serine and histidine observed in vitro are shown.

Sequence alignment of known FIH substrates (Figure 7) shows a preference for FIH to accept a recognition sequence close to the consensus ankyrin. Important residues for FIH binding are valine in -1 position of the hydroxylation site and a -8 leucine (Figure 7).

Table 2 Reported FIH reaction sites in HIF and ankyrin substrates.

protein	hydroxylation sites	detection method	references
HIF-1α	N803	NMR, Edman degradation, MS, amino acid analysis, crystallography, 2OG turnover	Lando et al (2002), Elkins et al (2003)
HIF-2α	N851	recombinant protein MS-MS	Lando et al (2002)
ASPP2	N986	recombinant protein MS-MS/ overexpression IP-MS/MS	Janke et al (2013)
iκBα	N210, N244	overexpression, IP-MS/MS, recombinant protein MS/MS	Cockman et al (2006)
p105	N678	overexpression, IP-MS/MS, 2OG turnover	Cockman et al (2006)
p19-Ink4d*	N101	2OG turnover	Cockmann et al (2006)
ILK-1*	N94	2OG trunover	Cockmann et al (2006)
Notch1	N1956, N2023	endogenous IP-MS/MS hypoxia control, crystallography	Coleman et al (2007) Wilkins et al (2009)
Notch2	N1902, N1969	recombinant protein MS-MS/MS, IP	Coleman et al (2007), Wilkins et al (2009)
Notch3	N1867, N1934	recombinant protein MS-MS/MS, IP	Coleman et al (2007), Wilkins et al (2009)
Tankyrase1	N864, H396, H711	Peptide-MS/MS	Cockman et al (2006),Tarhonskaya et al (2015), Yang et al (2011)
Tankyrase2	N203, N271, N427, N586 N671, N739, H238, H553	peptide-MS/MS, crystallography	Yang et al (2011)
TRPV3	N242	2OG turnover; IP, overexpressed IP-MS/MS	Karttunen et al (2015)
TRPV4	H265	peptide-MS/MS	Yang et al (2011)
GABP-β2	H131	peptide-MS/MS	Yang et al (2011)
hMASK	D1352	peptide LC-MS/MS	Yang et al (2011)b
GTAR	D1380	peptide LC-MS/MS	Yang et al (2011)b
ASB4	N246	overexpressed IP-MS/MS	Ferguson et al (2007)
ASB8	N80	endogenous IP- MS/MS IP-MS (FIH bait)	Rodriguez et al (2016)
Rabankyrin-5	N316, N649, N752, N797	overexpressed IP-MS/MS	Cockman et al (2009)
MYPT1	N67, N100, N226	overexpressed IP-MS/MS	Webb at al (2009)
Gankyrin*	N100	2OG trunover	Cockman et al (2006) Wilkins et al (2012)
RNasL	N233	overexpressed IP-MS/MS	Cockman et al (2009)
RIPK4	N579, N646	IVT product in cell lyste, MS/MS	Cockman et al (2008)
GABP-β1*	N98	2OG trunover	Cockman et al (2006)
FEM1β*	N526	2OG trunover	Cockman et al (2006)
FGIF*	N168	2OG trunover	Cockman et al (2006)
Myotrophin*	N62	2OG trunover	Cockman et al (2006)
AnkyrinR	N105	MS/MS-recombinant	Yang et al 2011
OV008**	N40	co-IP/ 2OG turnover, in vitro MS/MS	Chen and Fabrizio et al (2016)
OV126**	N285	co-IP/ 2OG turnover, in vitro MS/MS	Chen and Fabrizio et al (2016)
OV129**	N44	co-P/ 2OG turnover, in vitro MS/MS	Chen and Fabrizio et al (2016)

* only 2OG turn-over data, no direct evidence for hydroxylation

** viral proteins

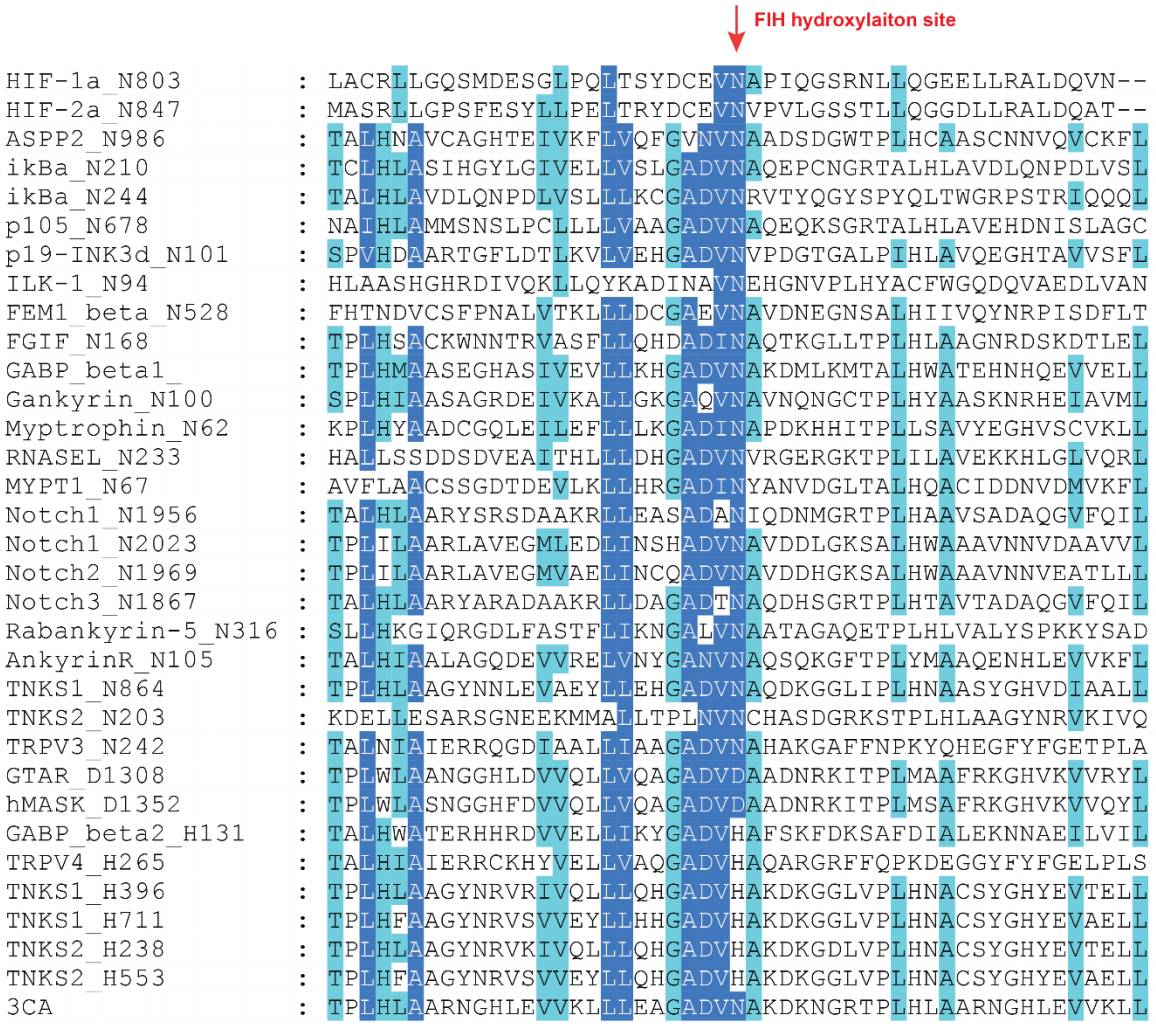


Figure 7 Sequence alignment of reported FIH substrates. Sequences were aligned around the FIH hydroxylation site. Image generated using GeneDoc. The FIH hydroxylation site is marked by a red arrow.

In addition to the reported substrates, FIH was also found to interact with several proteins for which no hydroxylation has been reported. Notably, not all of the reported FIH interactors are ARD proteins. Reported FIH interactors include the ARD proteins 4 consensus ankyrin (4CA[62]), Notch4[137], LRRK1[138]; the ubiquitin ligases VHL[83] and E6AP[139] and the de-ubiquitinylase OTUD7[140]. With the exception of 4CA, all of the above were identified from co-immunoprecipitation assays with FIH suggesting interactions in a cellular context. Note, indirect interactions, like binding to the same protein in cells, are possible.

1.6 FIH substrate binding and recognition

In contrast to the considerable number of identified FIH substrates, to date, only structures of FIH in complex with peptides derived from HIF-1α[141], human and mouse Notch1[129] and Tankyrase-2[8] have been reported.

In addition to naturally occurring substrates, structures of FIH in complex with peptides derived from the consensus ankyrin sequence have been reported[132,136]. Note, only the peptides derived from HIF-1α have elongated C-termini that allow for further interaction with the protein downstream of the hydroxylation site (Table 3).

Table 3 List of peptides crystallised with FIH

Substrate	Sequence	PDB code	Resolution[Å]
HIF-1a	SMDESGLPQLTSYDCEV **N** APIQGSRNLLQGEELLRALDQVN	1H2K	2.15
hNotch1	RSDAAKRLLEASADA **N** IQDN	3P3P	2.60
mNotch1	LEGMLEDLINSHADV **N** AVDD	3P3N	2.40
Tkns-2	RVSVVEYLLQHGADV **H** AKDKG	2Y0I	2.28
1CA_ASP*	HLEVVKLLLEAGADV (**Asp**) AKDK	2XUM	2.20
1CA_SER	HLEVVKLLLEAGADV (**Ser**) AKDK	4B7K	2.39
1CA_LEU	HLEVVKLLLEAGADV (**Leu**) AKDK	4B7E	2.50
1CA_DLE	HLEVVKLLLEAGADV (**DLE**) AKDK	4JAA	2.39
1CA_DYL	HLEVVKLLLEAGADV (**DYL**) AKDK	4NR1	2.68
1CA	HLEVVKLLLEAGADV (**Asn**) AKDK		

hNotch1: human Notch1 derived peptide; mNotch1: mouse derived Notch1 peptide; * 1CA_ASP was crystallised with FIH Q239H mutant, DLE: D-leucine, DYL: D-allyl-glycine.

The active sites of both FIH proteins of the FIH dimer face towards the same site, at least in the crystalline state (Figure 8A). Early structural studies on FIH revealed the binding mode of HIF-1α derived peptides to the FIH dimer. The substrates binds in a substrate binding groove, rather than pocket, with two distinct binding sites, site one and site two[142].

Site one is N-terminal to the hydroxylation site and spans from the dimerisation domain into the active site (Figure 8A). Note that the HIF-1α substrate is interacting with both enzymes of the FIH dimer. Subsequent studies confirmed the importance of the FIH dimerisation for substrate binding[45,124]. Interaction of the FIH with its substrate in site one can be sufficient for substrate recognition and turnover[46,127,129].

In addition to site one binding, further interactions of FIH with HIF-1α are observed C-terminal of the hydroxylation site, referred to as site two binding (Figure 8A). Site two binding is possibly not necessary for recognition of all substrates by FIH, but for several substrates, *i.e*. HIF-1α and Notch1, it has been shown that site two binding is important for efficient substrate turnover[112,137].

The Fe(II) is coordination in the active site of FIH by the facial triad that leaves 3 coordination sites of the iron(II) free, which are occupied by the C1 and C2 oxygen of 2OG, and one for oxygen binding. The latter is frequently found to be occupied by water in the crystal state (Figure 8B,C).

The C-5 carbonyl of 2OG interacts with a basic residue in the back of the active site (lysine-214, Figure 8B). The interactions of the C-5 carbonyl of 2OG with the lysine is conserved amongst JmjC family members and characteristic for this sub-family of 2OG oxygenases. In the case of *e.g.* DAOCS family members, an arginine binds the C-5 carbonyl in place of the lysine. In addition to the metal coordination and interaction with lysine, further hydrogen bonds stabilise the binding of 2OG to the active site (Figure 8B).

For stereo-selective oxidation of the β-carbon, the substrate has to be correctly orientated in the active site. The hydroxylated asparagine residue interacts with multiple residues in the active site, which apparently ensure the exact positioning of the oxidised C-H bond. Interaction of the mNotch1 substrate with the active site residues is shown in Figure 8C. Asparagine-1945 of mNotch1 forms hydrogen bonds with glutamine-239 and arginine-238 of FIH in addition to hydrophobic interactions with histidine-199 and tyrosine-102 of FIH, fixing its position. In addition to interactions with the iron(II), asparatate-201 of FIH forms hydrogen bonds with the backbone of asparagine-1945 of mNotch1. Important, but not essential are interactions of valine-1944 of mNotch1 with tryptophan-296 of FIH.

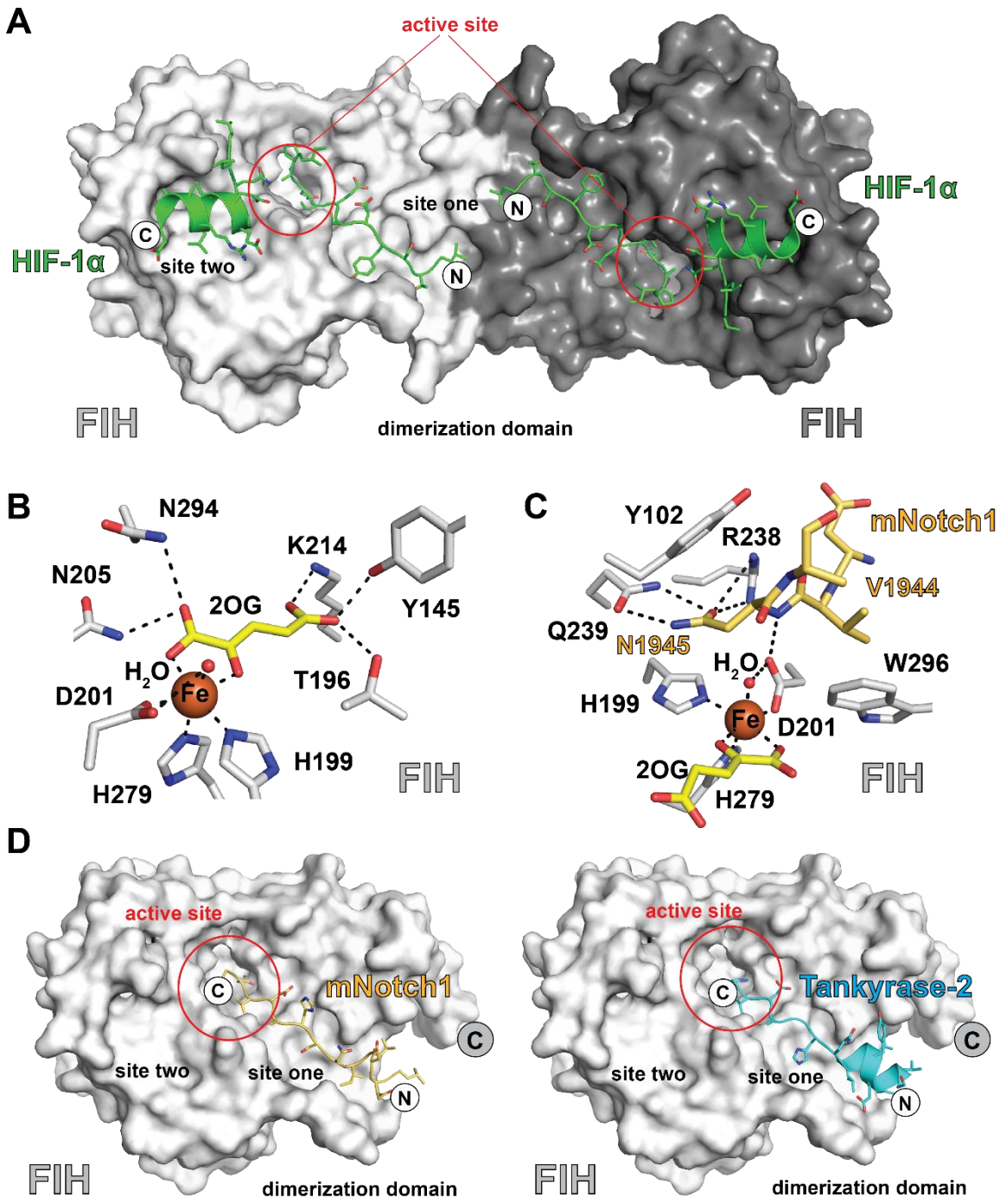


Figure 8 Substrate and co-factor binding to FIH. A Views from crystal structures of FIH show the dimeric structure of FIH and binding of HIF-1α peptide interacting with both dimer partners (2.15 Å, PDB code 1H2K[46]). FIH is coloured in grey, the FIH dimer is coloured in dark grey and the HIF-1α substrates are coloured in green. Binding sites on FIH are labelled as site one (N-terminal to the hydroxylation site) and site 2 (C-terminal to the hydroxylation site). **B** Views from the active site of FIH show the 2OG binding mode (2.40 Å, PDB code 3P3N[129]). **C** Views from the active site of FIH illustrate the substrate binding mode as exemplified by the binding of mNotch1 (2.40 Å, PDB code 3P3N[127]). The substrate is bound and held in position by several hydrogen bonds and hydrophobic interactions. **D** Overall binding modes of the ankyrin substrates mNotch1 (2.40 Å, PDB code 3p3N[127]) and Tankyrase-2 (2.28 Å, PDB code 2Y0I[8]) show an overall conserved site one binding mode.

FIH accepts two structurally unrelated groups of substrates[3,111,127,143]. The HIF-α transcription factors are intrinsically disordered in their C-terminal transactivation domain; the available evidence is that they only show the formation of secondary structure upon binding to other proteins[144]. Ankyrin repeat domain containing proteins on the other hand are fully-folded proteins with a well-defined tertiary structure. Interestingly, both classes of substrates characterised for FIH show a conserved binding mode (Figure 8D). No structures of FIH in complex with ankyrin substrates covering potential site two binding have been reported to date, despite accumulating data on the importance of this site (see below).

In order to bind to FIH, the ankyrin repeats have to unfold, at least partially (Figure 9). It is noted that the mNotch1 substrate forms an α-helix at the N-terminus binding to FIH reassembling the native ankyrin fold (Figure 9).

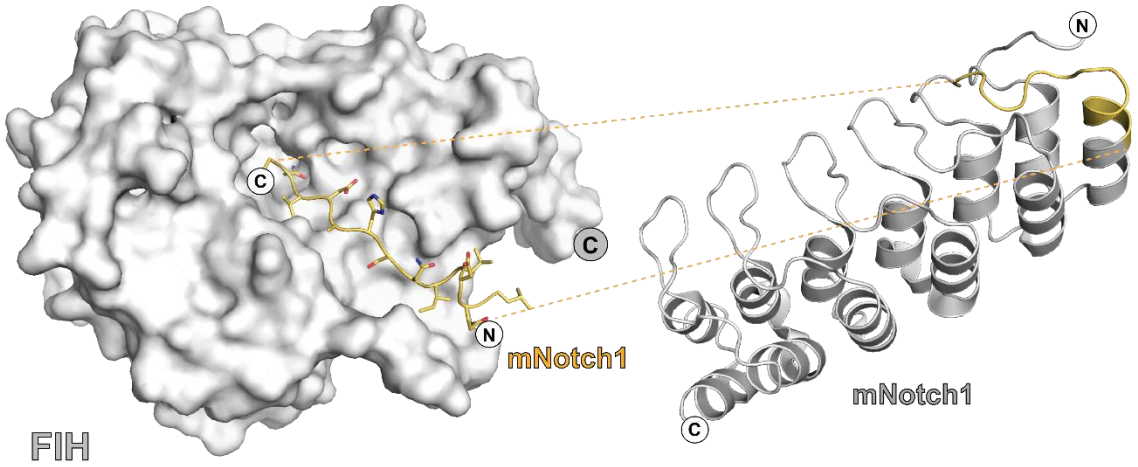


Figure 9 Unfolding of the ankyrin repeats is needed for FIH binding. Views from a crystal structure of FIH in complex with mNotch1 (2.40 Å, PDB code 3P3N[129]) and the ARDs of mNotch1 alone (2.35 Å, PDB code 2QC9[129]) illustrate the requirement of the ARD unfolding to bind to FIH.

1.7 Physiological role of FIH mediated hydroxylation

The function of FIH as a transcriptional repressor of HIF-1/2α in the hypoxic response is relatively well studied. Although the exact molecular mechanisms remain unclear, it is generally accepted that FIH mediated hydroxylation of the C-terminal transactivation domains of HIF-1/2α impairs their binding with p300/CBP like transcriptional co-activators (see above). Effects on HIF target gene expression are however highly dependent on the cellular context and, to date, are not fully understood[109,110].

The large number of ARD substrates of FIH led to the hypothesis that the ankyrins function as a pool forming a “hypoxic memory” by competing with HIF-α more strongly in their non-hydroxylated form than in their hydroxylated form[131,143,145]. During prolonged hypoxia, a pool of non-hydroxylated ankyrins could buffer FIH activity during short term re-oxygenation[145]. The potentially long life-times of ankyrin proteins make them unsuitable for switch-like functions as observed in the HIF pathway.

FIH mediated hydroxylation of ankyrin repeats has been shown to stabilise the ankyrin protein fold[62,63], but the potential relevance of this observation into the cellular functions of the proteins remains unclear. Further, the large number of FIH substrates challenges the investigation of single substrates. Studies using alanine mutation of the ankyrins as an “non-hydroxylated ankyrin mimic” to show potential effects of FIH mediated effects do not account for structural disturbance of the protein fold by the mutation[146,147], which may contribute to the observed effects. Substrate specific FIH inhibitors to dissect different cellular functions of FIH are not available.

In vivo studies on FIH KO mouse models imply that FIH is a non-essential gene. The mice showed reduced weight gain in high-fat diets and increased insulin resistance, but the loss of weight was otherwise unobtrusive[148].

In humans, FIH loss or mutation has not yet been reported to be linked to a pathological phenotype. In the cosmic database[149], sporadic mutations in cancer are listed including miss-sense mutations in a flexible-loop involved in substrate binding (R305I, I306V, K311N and P309H). Several studies investigated a potential link between FIH and cancer progression. The downregulation of FIH by microRNA-31 was reported to regulate keratinocyte differentiation and immortalisation[150-154]. Further studies reported several microRNAs (mir), mir-98, mir-184 and mir-135b, downregulate FIH and contribute to cell differentiation and cancer progression[153,155,156]. Whether the observed phenotypes are directly related to FIH remains however unclear. Finally, cytoplasmic location of FIH in invasive breast cancer was linked to enhanced HIF activity and poor clinical prognosis[157-160].

1.8 Aim of the Thesis

The aim of the work described in this thesis is to explore the interactions of FIH with small molecules and its substrates to broaden the understanding of FIH substrate recognition and human oxygenase catalysed reactions.

With increasing pharmaceutical interest in 2OG oxygenases, reliable crystallisation models for structural information are needed. The work thus also aimed to establish FIH as a surrogate for otherwise inaccessible 2OG oxygenase-inhibitor structures including for PHD inhibitors in clinical trials.

Potential differences in the binding of FIH substrates were investigated with the aim of helping development of substrate specific inhibitors of FIH.

Further work focused on characterising the ASPP protein family as FIH substrates, and included cellular studies assessing the potential physiological effects of FIH mediated ankyrin hydroxylation.

1.9 References

1 Flashman, E. & Schofield, C. J. The most versatile of all reactive intermediates? *Nat Chem Biol* **3**, 86-87, doi:10.1038/nchembio0207-86 (2007).

2 Hewitson, K. S. *et al.* Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J Biol Chem* **277**, 26351-26355, doi:10.1074/jbc.C200273200 (2002).

3 Lando, D. *et al.* FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev* **16**, 1466-1471, doi:10.1101/gad.991402 (2002).

4 McNeill, L. A. *et al.* Hypoxia-inducible factor asparaginyl hydroxylase (FIH-1) catalyses hydroxylation at the beta-carbon of asparagine-803. *Biochem J* **367**, 571-575, doi:10.1042/BJ20021162 (2002).

5 Ivan, M. *et al.* HIFalpha targeted for VHL-mediated destruction by proline hydroxylation: implications for O2 sensing. *Science* **292**, 464-468, doi:10.1126/science.1059817 (2001).

6 Jaakkola, P. *et al.* Targeting of HIF-alpha to the von Hippel-Lindau ubiquitylation complex by O2-regulated prolyl hydroxylation. *Science* **292**, 468-472, doi:10.1126/science.1059796 (2001).

7 Webby, C. J. *et al.* Jmjd6 catalyses lysyl-hydroxylation of U2AF65, a protein associated with RNA splicing. *Science* **325**, 90-93, doi:10.1126/science.1175865 (2009).

8 Yang, M. *et al.* Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidinyl residues within ankyrin repeat domains. *FEBS J* **278**, 1086-1097, doi:10.1111/j.1742-4658.2011.08022.x (2011).

9 Leśniak, R. K., Markolovic, S., Tars, K. & Schofield, C. J. Human carnitine biosynthesis proceeds via (2S,3S)-3-hydroxy-Nε-trimethyllysine. *Chem Commun (Camb)* **53**, 440-442, doi:10.1039/c6cc08381a (2016).

10 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).

11 Vaz, F. M., van Gool, S., Ofman, R., Ijlst, L. & Wanders, R. J. Carnitine biosynthesis: identification of the cDNA encoding human gamma-butyrobetaine hydroxylase. *Biochem Biophys Res Commun* **250**, 506-510, doi:10.1006/bbrc.1998.9343 (1998).

12 Ng, S. S. *et al.* Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87-91, doi:10.1038/nature05971 (2007).

13 Chen, Z. *et al.* Structural insights into histone demethylation by JMJD2 family members. *Cell* **125**, 691-702, doi:10.1016/j.cell.2006.04.024 (2006).

14 Tsukada, Y. *et al.* Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439**, 811-816, doi:10.1038/nature04433 (2006).

15 Yamane, K. *et al.* JHDM2A, a JmjC-containing H3K9 demethylase, facilitates transcription activation by androgen receptor. *Cell* **125**, 483-495, doi:10.1016/j.cell.2006.03.027 (2006).

16 Whetstine, J. R. *et al.* Reversal of histone lysine trimethylation by the JMJD2 family of histone demethylases. *Cell* **125**, 467-481, doi:10.1016/j.cell.2006.03.028 (2006).

17 Duncan, T. *et al.* Reversal of DNA alkylation damage by two human dioxygenases. *Proc Natl Acad Sci U S A* **99**, 16660-16665, doi:10.1073/pnas.262589799 (2002).

18 Aas, P. A. *et al.* Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* **421**, 859-863, doi:10.1038/nature01363 (2003).

19 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, doi:10.1074/jbc.M803776200 (2008).

20 Zheng, G. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol Cell* **49**, 18-29, doi:10.1016/j.molcel.2012.10.015 (2013).

21 Gerken, T. *et al.* The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* **318**, 1469-1472, doi:10.1126/science.1151710 (2007).

22 Walport, L. J. *et al.* Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases. *Nat Commun* **7**, 11974, doi:10.1038/ncomms11974 (2016).

23 Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O2-dependent halogenase. *Proc Natl Acad Sci U S A* **102**, 10111-10116, doi:10.1073/pnas.0504412102 (2005).

24 Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat Chem Biol* **10**, 921-923, doi:10.1038/nchembio.1625 (2014).

25 Falnes, P., Johansen, R. F. & Seeberg, E. AlkB-mediated oxidative demethylation reverses DNA damage in Escherichia coli. *Nature* **419**, 178-182, doi:10.1038/nature01048 (2002).

26 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, doi:10.1038/nature00908 (2002).

27 Holme, E., Lindstedt, G., Lindstedt, S. & Tofft, M. 18-O studies of the 2-ketoglutarate-dependent sequential oxygenation of thymine to 5-carboxyuracil. *J Biol Chem* **246**, 3314-3319 (1971).

28 Townsend, C. A. Structural studies of natural product biosynthetic proteins. *Chem Biol* **4**, 721-730 (1997).

29 Chang, W. C. *et al.* Mechanism of the C5 stereoinversion reaction in the biosynthesis of carbapenem antibiotics. *Science* **343**, 1140-1144, doi:10.1126/science.1248000 (2014).

30 Li, J., van Belkum, M. J. & Vederas, J. C. Functional characterization of recombinant hyoscyamine 6β-hydroxylase from Atropa belladonna. *Bioorg Med Chem* **20**, 4356-4363, doi:10.1016/j.bmc.2012.05.042 (2012).

31 Steffan, N., Grundmann, A., Afiyatullov, S., Ruan, H. & Li, S. M. FtmOx1, a non-heme Fe(II) and alpha-ketoglutarate-dependent dioxygenase, catalyses the endoperoxide formation of verruculogen in Aspergillus fumigatus. *Org Biomol Chem* **7**, 4082-4087, doi:10.1039/b908392h (2009).

32 Busby, R. W. & Townsend, C. A. A single monomeric iron center in clavaminate synthase catalyzes three nonsuccessive oxidative transformations. *Bioorg Med Chem* **4**, 1059-1064 (1996).

33 Jakubczyk, D. *et al.* Discovery and reconstitution of the cycloclavine biosynthetic pathway--enzymatic formation of a cyclopropyl group. *Angew Chem Int Ed Engl* **54**, 5117-5121, doi:10.1002/anie.201410002 (2015).

34 Hamed, R. B. *et al.* The enzymes of β-lactam biosynthesis. *Nat Prod Rep* **30**, 21-107, doi:10.1039/c2np20065a (2013).

35 Hutton, J. J., Trappel, A. L. & Udenfriend, S. Requirements for alpha-ketoglutarate, ferrous ion and ascorbate by collagen proline hydroxylase. *Biochem Biophys Res Commun* **24**, 179-184 (1966).

36 Rhoads, R. E. & Udenfriend, S. Decarboxylation of alpha-ketoglutarate coupled to collagen proline hydroxylase. *Proc Natl Acad Sci U S A* **60**, 1473-1478 (1968).

37 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).

38 Puistola, U., Turpeenniemi-Hujanen, T. M., Myllylä, R. & Kivirikko, K. I. Studies on the lysyl hydroxylase reaction. I. Initial velocity kinetics and related aspects. *Biochim Biophys Acta* **611**, 40-50 (1980).

39 Hutton, J. J. & Udenfriend, S. Soluble collagen proline hydroxylase and its substrates in several animal tissues. *Proc Natl Acad Sci U S A* **56**, 198-202 (1966).

40 Kivirikko, K. I. & Myllyharju, J. Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol* **16**, 357-368 (1998).

41 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, doi:10.1042/BJ20091609 (2010).
42 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, doi:10.1038/375700a0 (1995).
43 Valegård, K. *et al.* Structure of a cephalosporin synthase. *Nature* **394**, 805-809, doi:10.1038/29575 (1998).
44 Zhang, Z. *et al.* Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat Struct Biol* **7**, 127-133, doi:10.1038/72398 (2000).
45 Dann, C. E., Bruick, R. K. & Deisenhofer, J. Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci U S A* **99**, 15351-15356, doi:10.1073/pnas.202614999 (2002).
46 Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806, doi:10.1074/jbc.C200644200 (2003).
47 Bond, C. S. TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* **19**, 311-312 (2003).
48 Lee, H. J. *et al.* Kinetic and crystallographic studies on deacetoxycephalosporin C synthase (DAOCS). *J Mol Biol* **308**, 937-948, doi:10.1006/jmbi.2001.4649 (2001).
49 Markolovic, S. *et al.* Structure-function relationships of human JmjC oxygenases-demethylases versus hydroxylases. *Curr Opin Struct Biol* **41**, 62-72, doi:10.1016/j.sbi.2016.05.013 (2016).
50 Myllyharju, J. & Kivirikko, K. I. Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* **20**, 33-43, doi:10.1016/j.tig.2003.11.004 (2004).
51 Vaz, F. M. & Wanders, R. J. Carnitine biosynthesis in mammals. *Biochem J* **361**, 417-429 (2002).
52 Zdżalik, D. *et al.* Protozoan ALKBH8 oxygenases display both DNA repair and tRNA modification activities. *PLoS One* **9**, e98729, doi:10.1371/journal.pone.0098729 (2014).
53 Schofield, C. J. & Ratcliffe, P. J. Signalling hypoxia by HIF hydroxylases. *Biochem Biophys Res Commun* **338**, 617-626, doi:10.1016/j.bbrc.2005.08.111 (2005).
54 Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat Rev Genet* **7**, 715-727, doi:10.1038/nrg1945 (2006).
55 Scott-Browne, J. P., Cw, L. & Rao, A. TET proteins in natural and induced differentiation. *Curr Opin Genet Dev* **46**, 202-208, doi:10.1016/j.gde.2017.07.011 (2017).
56 Parashar, N. C., Parashar, G., Nayyar, H. & Sandhir, R. N6-adenine DNA methylation demystified in eukaryotic genome: From biology to pathology. *Biochimie*, doi:10.1016/j.biochi.2017.10.014 (2017).
57 Ge, W. *et al.* Oxygenase-catalyzed ribosome hydroxylation occurs in prokaryotes and humans. *Nat Chem Biol* **8**, 960-962, doi:10.1038/nchembio.1093 (2012).
58 Katz, M. J. *et al.* Sudestada1, a Drosophila ribosomal prolyl-hydroxylase required for mRNA translation, cell homeostasis, and organ growth. *Proc Natl Acad Sci U S A* **111**, 4025-4030, doi:10.1073/pnas.1314485111 (2014).
59 Loenarz, C. *et al.* Hydroxylation of the eukaryotic ribosomal decoding center affects translational accuracy. *Proc Natl Acad Sci U S A* **111**, 4019-4024, doi:10.1073/pnas.1311750111 (2014).
60 Singleton, R. S. *et al.* OGFOD1 catalyzes prolyl hydroxylation of RPS23 and is involved in translation control and stress granule formation. *Proc Natl Acad Sci U S A* **111**, 4031-4036, doi:10.1073/pnas.1314482111 (2014).

61 Horita, S. *et al.* Structure of the ribosomal oxygenase OGFOD1 provides insights into the regio- and stereoselectivity of prolyl hydroxylases. *Structure* **23**, 639-652, doi:10.1016/j.str.2015.01.014 (2015).
62 Hardy, A. P., Prokes, I., Kelly, L., Campbell, I. D. & Schofield, C. J. Asparaginyl beta-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J Mol Biol* **392**, 994-1006, doi:10.1016/j.jmb.2009.07.070 (2009).
63 Kelly, L., McDonough, M. A., Coleman, M. L., Ratcliffe, P. J. & Schofield, C. J. Asparagine beta-hydroxylation stabilizes the ankyrin repeat domain fold. *Mol Biosyst* **5**, 52-58, doi:10.1039/b815271c (2009).
64 Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-inducible Factor (FIH) as an Oxygen Sensor. *J Biol Chem* **290**, 19726-19742, doi:10.1074/jbc.M115.653014 (2015).
65 Tarhonskaya, H. *et al.* Investigating the contribution of the active site environment to the slow reaction of hypoxia-inducible factor prolyl hydroxylase domain 2 with oxygen. *Biochem J* **463**, 363-372, doi:10.1042/BJ20140779 (2014).
66 Hancock, R. L., Masson, N., Dunne, K., Flashman, E. & Kawamura, A. The Activity of JmjC Histone Lysine Demethylase KDM4A is Highly Sensitive to Oxygen Concentrations. *ACS Chem Biol* **12**, 1011-1019, doi:10.1021/acschembio.6b00958 (2017).
67 Hirsilä, M., Koivunen, P., Günzler, V., Kivirikko, K. I. & Myllyharju, J. Characterization of the human prolyl 4-hydroxylases that modify the hypoxia-inducible factor. *J Biol Chem* **278**, 30772-30780, doi:10.1074/jbc.M304982200 (2003).
68 Ehrismann, D. *et al.* Studies on the activity of the hypoxia-inducible-factor hydroxylases using an oxygen consumption assay. *Biochem J* **401**, 227-234, doi:10.1042/BJ20061151 (2007).
69 He, Y. F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303-1307, doi:10.1126/science.1210944 (2011).
70 Takeuchi, T. *et al.* Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev* **9**, 1211-1222 (1995).
71 Clissold, P. M. & Ponting, C. P. JmjC: cupin metalloenzyme-like domains in jumonji, hairless and phospholipase A2beta. *Trends Biochem Sci* **26**, 7-9 (2001).
72 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, doi:10.1093/nar/gkq265 (2010).
73 Chang, B., Chen, Y., Zhao, Y. & Bruick, R. K. JMJD6 is a histone arginine demethylase. *Science* **318**, 444-447, doi:10.1126/science.1145801 (2007).
74 Mantri, M. *et al.* The 2-oxoglutarate-dependent oxygenase JMJD6 catalyses oxidation of lysine residues to give 5S-hydroxylysine residues. *Chembiochem* **12**, 531-534, doi:10.1002/cbic.201000641 (2011).
75 Unoki, M. *et al.* Lysyl 5-hydroxylation, a novel histone modification, by Jumonji domain containing 6 (JMJD6). *J Biol Chem* **288**, 6053-6062, doi:10.1074/jbc.M112.433284 (2013).
76 Hsia, D. A. *et al.* KDM8, a H3K36me2 histone demethylase that acts in the cyclin A1 coding region to regulate cancer cell proliferation. *Proc Natl Acad Sci U S A* **107**, 9671-9676, doi:10.1073/pnas.1000401107 (2010).
77 Youn, M. Y. *et al.* JMJD5, a Jumonji C (JmjC) domain-containing protein, negatively regulates osteoclastogenesis by facilitating NFATc1 protein degradation. *J Biol Chem* **287**, 12994-13004, doi:10.1074/jbc.M111.323105 (2012).
78 Del Rizzo, P. A., Krishnan, S. & Trievel, R. C. Crystal structure and functional analysis of JMJD5 indicate an alternate specificity and function. *Mol Cell Biol* **32**, 4044-4052, doi:10.1128/MCB.00513-12 (2012).

79 Sinha, K. M., Yasuda, H., Coombes, M. M., Dent, S. Y. & de Crombrugghe, B. Regulation of the osteoblast-specific transcription factor Osterix by NO66, a Jumonji family histone demethylase. *EMBO J* **29**, 68-79, doi:10.1038/emboj.2009.332 (2010).
80 Lu, Y. *et al.* Lung cancer-associated JmjC domain protein mdig suppresses formation of tri-methyl lysine 9 of histone H3. *Cell Cycle* **8**, 2101-2109, doi:10.4161/cc.8.13.8927 (2009).
81 Chowdhury, R. *et al.* Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature*, doi:10.1038/nature13263 (2014).
82 Wang, C. *et al.* Structure of the JmjC domain-containing protein NO66 complexed with ribosomal protein Rpl8. *Acta Crystallogr D Biol Crystallogr* **71**, 1955-1964, doi:10.1107/S1399004715012948 (2015).
83 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, doi:10.1101/gad.924501 (2001).
84 Epstein, A. C. *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).
85 Kaelin, W. G. & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol Cell* **30**, 393-402, doi:10.1016/j.molcel.2008.04.009 (2008).
86 Schofield, C. J. & Ratcliffe, P. J. Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Biol* **5**, 343-354, doi:10.1038/nrm1366 (2004).
87 Semenza, G. L., Roth, P. H., Fang, H. M. & Wang, G. L. Transcriptional regulation of genes encoding glycolytic enzymes by hypoxia-inducible factor 1. *J Biol Chem* **269**, 23757-23763 (1994).
88 Wang, G. L. & Semenza, G. L. Characterization of hypoxia-inducible factor 1 and regulation of DNA binding activity by hypoxia. *J Biol Chem* **268**, 21513-21518 (1993).
89 Wang, G. L. & Semenza, G. L. Desferrioxamine induces erythropoietin gene expression and hypoxia-inducible factor 1 DNA-binding activity: implications for models of hypoxia signal transduction. *Blood* **82**, 3610-3615 (1993).
90 Firth, J. D., Ebert, B. L. & Ratcliffe, P. J. Hypoxic regulation of lactate dehydrogenase A. Interaction between hypoxia-inducible factor 1 and cAMP response elements. *J Biol Chem* **270**, 21021-21027 (1995).
91 Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proc Natl Acad Sci U S A* **92**, 5510-5514 (1995).
92 Jiang, B. H., Rue, E., Wang, G. L., Roe, R. & Semenza, G. L. Dimerization, DNA binding, and transactivation properties of hypoxia-inducible factor 1. *J Biol Chem* **271**, 17771-17778 (1996).
93 Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature*, doi:10.1038/nature14883 (2015).
94 Card, P. B., Erbel, P. J. & Gardner, K. H. Structural basis of ARNT PAS-B dimerization: use of a common beta-sheet interface for hetero- and homodimerization. *J Mol Biol* **353**, 664-677, doi:10.1016/j.jmb.2005.08.043 (2005).
95 Erbel, P. J., Card, P. B., Karakuzu, O., Bruick, R. K. & Gardner, K. H. Structural basis for PAS domain heterodimerization in the basic helix--loop--helix-PAS transcription factor hypoxia-inducible factor. *Proc Natl Acad Sci U S A* **100**, 15504-15509, doi:10.1073/pnas.2533374100 (2003).
96 Wood, S. M., Gleadle, J. M., Pugh, C. W., Hankinson, O. & Ratcliffe, P. J. The role of the aryl hydrocarbon receptor nuclear translocator (ARNT) in hypoxic induction of gene expression. Studies in ARNT-deficient cells. *J Biol Chem* **271**, 15117-15123 (1996).
97 Dioum, E. M. *et al.* NPAS2: a gas-responsive transcription factor. *Science* **298**, 2385-2387, doi:10.1126/science.1078456 (2002).

98 Murray, I. A., Patterson, A. D. & Perdew, G. H. Aryl hydrocarbon receptor ligands in cancer: friend and foe. *Nat Rev Cancer* **14**, 801-814, doi:10.1038/nrc3846 (2014).
99 Wu, D., Potluri, N., Lu, J., Kim, Y. & Rastinejad, F. Structural integration in hypoxia-inducible factors. *Nature* **524**, 303-308, doi:10.1038/nature14883 (2015).
100 Makino, Y., Kanopka, A., Wilson, W. J., Tanaka, H. & Poellinger, L. Inhibitory PAS domain protein (IPAS) is a hypoxia-inducible splicing variant of the hypoxia-inducible factor-3alpha locus. *J Biol Chem* **277**, 32405-32408, doi:10.1074/jbc.C200328200 (2002).
101 Gordan, J. D. & Simon, M. C. Hypoxia-inducible factors: central regulators of the tumor phenotype. *Curr Opin Genet Dev* **17**, 71-77, doi:10.1016/j.gde.2006.12.006 (2007).
102 Lau, K. W., Tian, Y. M., Raval, R. R., Ratcliffe, P. J. & Pugh, C. W. Target gene selectivity of hypoxia-inducible factor-alpha in renal cancer cells is conveyed by post-DNA-binding mechanisms. *Br J Cancer* **96**, 1284-1292, doi:10.1038/sj.bjc.6603675 (2007).
103 Qing, G. & Simon, M. C. Hypoxia inducible factor-2alpha: a critical mediator of aggressive tumor phenotypes. *Curr Opin Genet Dev* **19**, 60-66, doi:10.1016/j.gde.2008.12.001 (2009).
104 Maxwell, P. H. *et al.* The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271-275, doi:10.1038/20459 (1999).
105 Wykoff, C. C., Pugh, C. W., Maxwell, P. H., Harris, A. L. & Ratcliffe, P. J. Identification of novel hypoxia dependent and independent target genes of the von Hippel-Lindau (VHL) tumour suppressor by mRNA differential expression profiling. *Oncogene* **19**, 6297-6305, doi:10.1038/sj.onc.1204012 (2000).
106 Salceda, S. & Caro, J. Hypoxia-inducible factor 1alpha (HIF-1alpha) protein is rapidly degraded by the ubiquitin-proteasome system under normoxic conditions. Its stabilization by hypoxia depends on redox-induced changes. *J Biol Chem* **272**, 22642-22647 (1997).
107 Pugh, C. W., O'Rourke, J. F., Nagao, M., Gleadle, J. M. & Ratcliffe, P. J. Activation of hypoxia-inducible factor-1; definition of regulatory domains within the alpha subunit. *J Biol Chem* **272**, 11205-11214 (1997).
108 Singleton, R. S. *et al.* Quantitative mass spectrometry reveals dynamics of factor-inhibiting hypoxia-inducible factor-catalyzed hydroxylation. *J Biol Chem* **286**, 33784-33794, doi:10.1074/jbc.M111.262808 (2011).
109 Stolze, I. P. *et al.* Genetic analysis of the role of the asparaginyl hydroxylase factor inhibiting hypoxia-inducible factor (FIH) in regulating hypoxia-inducible factor (HIF) transcriptional target genes [corrected]. *J Biol Chem* **279**, 42719-42725, doi:10.1074/jbc.M406713200 (2004).
110 Chan, M. C. *et al.* Tuning the transcriptional response to hypoxia by inhibiting HIF prolyl- and asparaginyl-hydroxylases. *J Biol Chem*, doi:10.1074/jbc.M116.749291 (2016).
111 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, doi:10.1126/science.1068592 (2002).
112 Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-Inducible Factor (FIH) as an Oxygen Sensor. *J Biol Chem*, doi:10.1074/jbc.M115.653014 (2015).
113 Hancock, R. L., Masson, N., Dunne, K., Flashman, E. & Kawamura, A. The activity of JmjC histone lysine demethylase KDM4A is highly sensitive to oxygen concentrations. *ACS Chem Biol*, doi:10.1021/acschembio.6b00958 (2017).
114 Semenza, G. L. Regulation of erythropoietin production. New insights into molecular mechanisms of oxygen homeostasis. *Hematol Oncol Clin North Am* **8**, 863-884 (1994).
115 Yan, L., Colandrea, V. J. & Hale, J. J. Prolyl hydroxylase domain-containing protein inhibitors as stabilizers of hypoxia-inducible factor: small molecule-based therapeutics for anemia. *Expert Opin Ther Pat* **20**, 1219-1245, doi:10.1517/13543776.2010.510836 (2010).

116 Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029-1033, doi:10.1126/science.1160809 (2009).
117 Bayley, J. P. & Devilee, P. The Warburg effect in 2012. *Curr Opin Oncol* **24**, 62-67, doi:10.1097/CCO.0b013e32834deb9e (2012).
118 Harada, H. Hypoxia-inducible factor 1-mediated characteristic features of cancer cells for tumor radioresistance. *J Radiat Res* **57 Suppl 1**, i99-i105, doi:10.1093/jrr/rrw012 (2016).
119 Koukourakis, M. I. *et al.* Normal tissue radioprotection by amifostine via Warburg-type effects. *Sci Rep* **6**, 30986, doi:10.1038/srep30986 (2016).
120 Han, G. *et al.* The hydroxylation activity of Jmjd6 is required for its homo-oligomerization. *J Cell Biochem* **113**, 1663-1670, doi:10.1002/jcb.24035 (2012).
121 Wolf, A. *et al.* The polyserine domain of the lysyl-5 hydroxylase Jmjd6 mediates subnuclear localization. *Biochem J* **453**, 357-370, doi:10.1042/BJ20130529 (2013).
122 Hahn, P. *et al.* Analysis of Jmjd6 cellular localization and testing for its involvement in histone demethylation. *PLoS One* **5**, e13769, doi:10.1371/journal.pone.0013769 (2010).
123 Tibrewal, N., Liu, T., Li, H. & Birge, R. B. Characterization of the biochemical and biophysical properties of the phosphatidylserine receptor (PS-R) gene product. *Mol Cell Biochem* **304**, 119-125, doi:10.1007/s11010-007-9492-8 (2007).
124 Lancaster, D. E. *et al.* Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J* **383**, 429-437, doi:10.1042/BJ20040735 (2004).
125 Kato, M. *et al.* Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res* **39**, 1576-1585, doi:10.1093/nar/gkq919 (2011).
126 Lancaster, D. E., McDonough, M. A. & Schofield, C. J. Factor inhibiting hypoxia-inducible factor (FIH) and other asparaginyl hydroxylases. *Biochem Soc Trans* **32**, 943-945, doi:10.1042/BST0320943 (2004).
127 Cockman, M. E. *et al.* Posttranslational hydroxylation of ankyrin repeats in IkappaB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc Natl Acad Sci U S A* **103**, 14767-14772, doi:10.1073/pnas.0606877103 (2006).
128 Ferguson, J. E. *et al.* ASB4 is a hydroxylation substrate of FIH and promotes vascular differentiation via an oxygen-dependent mechanism. *Mol Cell Biol* **27**, 6407-6419, doi:10.1128/MCB.00511-07 (2007).
129 Coleman, M. L. *et al.* Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038, doi:10.1074/jbc.M704102200 (2007).
130 Webb, J. D., Murányi, A., Pugh, C. W., Ratcliffe, P. J. & Coleman, M. L. MYPT1, the targeting subunit of smooth-muscle myosin phosphatase, is a substrate for the asparaginyl hydroxylase factor inhibiting hypoxia-inducible factor (FIH). *Biochem J* **420**, 327-333, doi:10.1042/BJ20081905 (2009).
131 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, doi:10.1074/mcp.M800340-MCP200 (2009).
132 Yang, M. *et al.* Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. *J Biol Chem* **286**, 7648-7660, doi:10.1074/jbc.M110.193540 (2011).
133 Janke, K. *et al.* Factor inhibiting HIF-1 (FIH-1) modulates protein interactions of apoptosis-stimulating p53 binding protein 2 (ASPP2). *J Cell Sci* **126**, 2629-2640, doi:10.1242/jcs.117564 (2013).

134 Karttunen, S. *et al.* Oxygen-dependent hydroxylation by Factor Inhibiting HIF (FIH) regulates the TRPV3 ion channel. *J Cell Sci*, doi:10.1242/jcs.158451 (2014).
135 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, doi:10.1074/jbc.M804999200 (2008).
136 Yang, M. *et al.* Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl* **52**, 1700-1704, doi:10.1002/anie.201208046 (2013).
137 Wilkins, S. E. *et al.* Differences in hydroxylation and binding of Notch and HIF-1alpha demonstrate substrate selectivity for factor inhibiting HIF-1 (FIH-1). *Int J Biochem Cell Biol* **41**, 1563-1571, doi:10.1016/j.biocel.2009.01.005 (2009).
138 Peng, H., Kaplan, N., Yang, W., Getsios, S. & Lavker, R. M. FIH-1 disrupts an LRRK1/EGFR complex to positively regulate keratinocyte migration. *Am J Pathol* **184**, 3262-3271, doi:10.1016/j.ajpath.2014.08.014 (2014).
139 Martínez-Noël, G. *et al.* Identification and proteomic analysis of distinct UBE3A/E6AP protein complexes. *Mol Cell Biol* **32**, 3095-3106, doi:10.1128/MCB.00201-12 (2012).
140 Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389-403, doi:10.1016/j.cell.2009.04.042 (2009).
141 Elkins, J. M. *et al.* Oligomeric structure of proclavaminic acid amidino hydrolase: evolution of a hydrolytic enzyme in clavulanic acid biosynthesis. *Biochem J* **366**, 423-434, doi:10.1042/BJ20020125 (2002).
142 Markolovic, S., Wilkins, S. E. & Schofield, C. J. Protein Hydroxylation Catalyzed by 2-Oxoglutarate-dependent Oxygenases. *J Biol Chem* **290**, 20712-20722, doi:10.1074/jbc.R115.662627 (2015).
143 Cockman, M. E., Webb, J. D. & Ratcliffe, P. J. FIH-dependent asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Ann N Y Acad Sci* **1177**, 9-18, doi:10.1111/j.1749-6632.2009.05042.x (2009).
144 Freedman, S. J. *et al.* Structural basis for recruitment of CBP/p300 by hypoxia-inducible factor-1 alpha. *Proc Natl Acad Sci U S A* **99**, 5367-5372, doi:10.1073/pnas.082117899 (2002).
145 Schmierer, B., Novák, B. & Schofield, C. J. Hypoxia-dependent sequestration of an oxygen sensor by a widespread structural motif can shape the hypoxic response--a predictive kinetic model. *BMC Syst Biol* **4**, 139, doi:10.1186/1752-0509-4-139 (2010).
146 Zheng, X. *et al.* Interaction with factor inhibiting HIF-1 defines an additional mode of cross-coupling between the Notch and hypoxia signaling pathways. *Proc Natl Acad Sci U S A* **105**, 3368-3373, doi:10.1073/pnas.0711591105 (2008).
147 Scholz, C. C. *et al.* FIH Regulates Cellular Metabolism through Hydroxylation of the Deubiquitinase OTUB1. *PLoS Biol* **14**, e1002347, doi:10.1371/journal.pbio.1002347 (2016).
148 Zhang, N. *et al.* The asparaginyl hydroxylase factor inhibiting HIF-1alpha is an essential regulator of metabolism. *Cell Metab* **11**, 364-378, doi:10.1016/j.cmet.2010.03.001 (2010).
149 Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* **43**, D805-811, doi:10.1093/nar/gku1075 (2015).
150 Peng, H. *et al.* MicroRNA-31 targets FIH-1 to positively regulate corneal epithelial glycogen metabolism. *FASEB J* **26**, 3140-3147, doi:10.1096/fj.11-198515 (2012).
151 Peng, H. *et al.* microRNA-31/factor-inhibiting hypoxia-inducible factor 1 nexus regulates keratinocyte differentiation. *Proc Natl Acad Sci U S A* **109**, 14030-14034, doi:10.1073/pnas.1111292109 (2012).
152 Chen, T. *et al.* MicroRNA-31 contributes to colorectal cancer development by targeting factor inhibiting HIF-1α (FIH-1). *Cancer Biol Ther* **15**, 516-523, doi:10.4161/cbt.28017 (2014).

153 Liu, Z. *et al.* Enhanced functional properties of human limbal stem cells by inhibition of the miR-31/FIH-1/P21 axis. *Acta Ophthalmol* **95**, e495-e502, doi:10.1111/aos.13503 (2017).
154 Zhang, J. *et al.* Downregulation of microRNA-31 inhibits proliferation and induces apoptosis by targeting HIF1AN in human keloid. *Oncotarget* **8**, 74623-74634, doi:10.18632/oncotarget.20284 (2017).
155 Hu, D. *et al.* microRNA-98 mediated microvascular hyperpermeability during burn shock phase via inhibiting FIH-1. *Eur J Med Res* **20**, 51, doi:10.1186/s40001-015-0141-5 (2015).
156 Aakula, A. *et al.* MicroRNA-135b regulates ERα, AR and HIF1AN and affects breast and prostate cancer cell growth. *Mol Oncol* **9**, 1287-1300, doi:10.1016/j.molonc.2015.03.001 (2015).
157 Soilleux, E. J. *et al.* Use of novel monoclonal antibodies to determine the expression and distribution of the hypoxia regulatory factors PHD-1, PHD-2, PHD-3 and FIH in normal and neoplastic human tissues. *Histopathology* **47**, 602-610, doi:10.1111/j.1365-2559.2005.02280.x (2005).
158 Tan, E. Y. *et al.* Cytoplasmic location of factor-inhibiting hypoxia-inducible factor is associated with an enhanced hypoxic response and a shorter survival in invasive breast cancer. *Breast Cancer Res* **9**, R89, doi:10.1186/bcr1838 (2007).
159 Giatromanolaki, A. *et al.* Expression of prolyl-hydroxylases PHD-1, 2 and 3 and of the asparagine hydroxylase FIH in non-small cell lung cancer relates to an activated HIF pathway. *Cancer Lett* **262**, 87-93, doi:10.1016/j.canlet.2007.11.041 (2008).
160 Tan, E. Y. *et al.* The key hypoxia regulated gene CAIX is upregulated in basal-like breast tumours and is associated with resistance to chemotherapy. *Br J Cancer* **100**, 405-411, doi:10.1038/sj.bjc.6604844 (2009).

2 Small molecule binding to the HIF hydroxylases

This chapter describes work leading to the first crystal structures of the PHD inhibitors in clinical trials, Vadadustat, GSK1278863 and Molidustat, in complex with a HIF hydroxylase. Crystal structures of other small molecules bound to the active site of FIH are presented and their active site interactions and metal coordination are described. For the structures of FIH in complex with Vadadustat and FG-4592, Pan-Dataset Density Analysis was used, which enabled the low occupancy conformational states of the ligands to be determined. Crystal structures of FIH in complex with PHD inhibitor IOX4 and three further test compounds are presented and analysed. The value of FIH as a model 2OG oxygenase for structural studies is demonstrated by structural analysis of FIH complexed with the instable ligand 2,2-dicholoro glutarate.

2.1 Introduction

Members of the human 2OG oxygenase family are crucial for animal cell homeostasis and survival and are frequently dysregulated in disease (see chapter 1). Consequently, there is a significant “ongoing” pharmaceutical and academic interest in inhibiting specific 2OG oxygenases, including KDMs, BBOX1, and the PHDs, further discussed below.

BBOX1 is proposed to be targeted clinically by Meldonium for a cardiac protection treatment[1]. Meldonium has received public attention as a consequence of its abuse by prominent athletes for performance enhancement[2]. However, the clinical value of BBOX1 inhibition remains controversial[3].

Efforts to selectively inhibit KDM5 and KDM6 for cancer treatment are ongoing. The high sequence and structural similarity of KDM5 and KDM6, as well as a lack of structural information has hindered the development of potent and isoform specific inhibitors[4-6].

The PHDs are being targeted for anaemia treatment with several drugs currently in clinical trials[7,8]. Common complications in chronic kidney diseases (CDK) are low renal erythropoietin (EPO) production and/or EPO resistance[9,10]. These lead to low erythrocyte levels, commonly referred to as anaemia. Current routine treatment for patients with CKD linked anaemia involves intravenous administration of recombinant EPO, often accompanied by oral or intravenous iron supplementation[11]. Although this has proven to be clinically efficient, it comes with relatively high treatment costs, potential health risks, and considerable patient discomfort[12]. The levels of EPO injected are up to ~1,000 fold higher than endogenous levels which can cause an imbalance of erythropoiesis; also, the exogenous protein can cause an immune reaction[13-15]. In addition to the clinical need of cheaper and safer treatment, there is a major commercial interest in new EPO enhancing drugs with a market volume of ~$10 billion per annum[12].

2.1.1 Structural basis for 2OG inhibitors

Many 2OG oxygenase inhibitors are 2OG competitors or substrate competitors; some show dual modes of action: substrate and 2OG competitors have been reported (Figure 1).

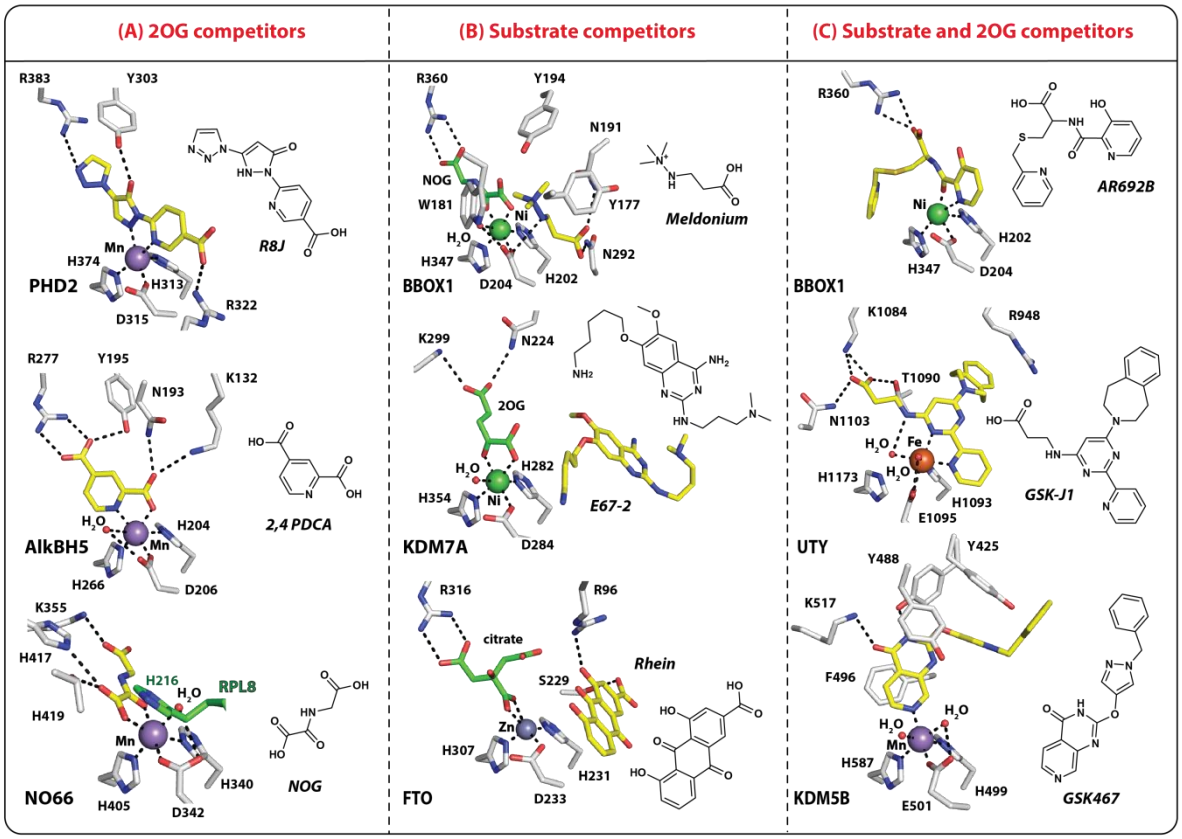


Figure 1 Views from crystal structures of inhibitors bound to the active site of selected 2OG oxygenases illustrate different modes of binding and inhibition. Figure modified from Islam et al, *in press* **(A)** Views from a crystal structure of PHD2 in complex with the IOX4 derivative R8J (3.30 Å, PDB code 5A3U[16]), ALKBH5 in complex with the broad spectrum 2OG oxygenase inhibitor 2,4 PDCA (2.50 Å, PDB code 4NRQ[17]) and NO66 in complex with the 2OG analogue and broad spectrum 2OG oxygenase inhibitor NOG (2.30 Å, PDB code 4CCO[18]). **(B)** Views from crystal structures of BBOX1 in complex with Meldonium (1.82 Å, PDB code 3MS5), KDM7A in complex with E67-2 (2.69 Å, PDB code 3U78[19]) and FTO in complex with Rhein (2.60 Å, PDB code 4IE7[20]). **(C)** BBOX1 in complex with AR692B (2.82 Å, PDB code 4C8R[21]), UTY in complex with GSK-J1 (2.00 Å, PDB code 3ZPO[22]) and KDM5B in complex with GSK467 (2.30 Å, PDB code 5FUN[23]).

2.1.2 PHD inhibitors in clinical trials

Upregulation of endogenous EPO levels *via* HIF-2α stabilisation with small molecule inhibitors is an attractive therapeutic strategy. Following the discovery of PHDs as oxygen sensors in cells and regulators of the HIF transcription factors[8,12], various pharmaceutical companies have developed PHD inhibitors that are currently in clinical trials. Three drug candidates, FG-4592 from FibroGen in cooperation with AstraZeneca[24,25], Vadadustat from Akebia[26], and GSK1278863 from GlaxoSmithKline (GSK)[27,28], are currently in clinical phase III; Molidustat from Bayer[29] is in clinical phase II. Further candidates are in clinical phase I or pre-clinical studies, as summarised in Figure 2[8].

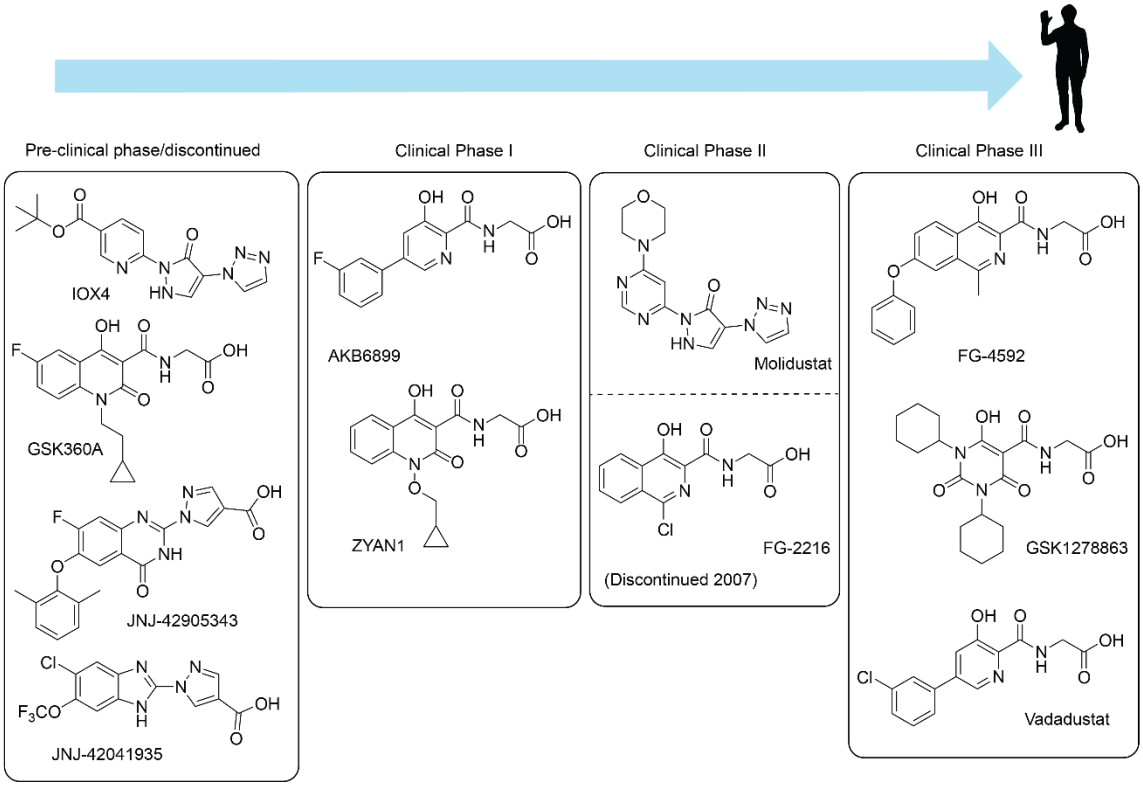


Figure 2: Progression PHD inhibitors in clinical trials. Overview of PHD inhibitors currently in clinical and pre-clinical trials.

The current generation of PHD inhibitors are being proposed as anaemia therapeutics, but PHD inhibitors are also under investigation for other clinical indications such as tissue protection after ischemic insult[30-32].

To date, FG-4592 has progressed the furthest in clinical trials and is currently in clinical phase III[12]. Trials of the related compound FG-2216 were discontinued after a death case, which was later shown to be unrelated to the trials[8]. Notably, all of the compounds, which have advanced to clinical phase III, incorporate an N-glycinamide group attached to a heteroaromatic ring system; Molidustat is the only PHD inhibitor in late clinical trials not incorporating an N-glycinamide motif. Although already administered to a large number of patients, very little is known about the mechanism of action of these inhibitors.

FG-4592 recently aroused public interest when it was abused by athletes as illegal performance enhancer. All agents upregulating the HIF pathway are prohibited substances by the World Anti-Doping Agency (WADA) which also explicitly states FG-4592 and Molidustat and are routinely checked for in doping controls[33-37].

2.1.3 Selectivity and activity of clinical PHD inhibitors and choice of a model systems for structural studies.

The clinical used PHD2 inhibitors are being administered to a large number of patients during clinical trials. However, very little is being published about the mechanism of action for any of these compounds. Solving the structures of the clinical inhibitors in complex with oxygenases is therefore of key public interest.

Investigations against recombinant forms of selected 2OG oxygenases have shown that the inhibitors are very potent and highly selective for PHD2[7] (IC_{50} values in the low nanomolar range, Table 1).

Table 1: Potency of clinical candidates FG-4592, GSK1278863, Molidustat and Vadadustat and the non-clinical PHD inhibitor IOX4 against a selective panel of 2OG oxygenases. Table Modified from Yeh et al[7].

	IC_{50} [μM]									
Inhibitor	PHD2	FIH	Ofd1	Tpa1p	OGFOD1	JARID1A (KDM5A)	JARID1B (KDM5B)	JARID1C (KDM5C)	JARID1D (KDM5D)	JMJD3 (KDM6B)
FG-4592	0.027	>100	8.5	12.8	<1	>100	>100	>100	>100	>100
GSK1278863	0.067	21	23.9	5.1	2.1	>100	>100	>100	>100	>100
Molidustat	0.007	66	10.3	3.8	5.1	>100	>100	>100	>100	35
IOX4	0.003	31	41.3	2.3	<1	53	87	97	91	<1
Vadadustat	0.029	29	ND*	ND*	1.4	>100	30	>100	>100	37

While the inhibitors' IC_{50} values are higher for FIH than for PHD2, FIH is considerably more robust at crystallising relative to PHD2 and is, like the PHDs, part of the HIF pathway. FIH and PHD2 share an overall conserved fold and are similar in their overall active site composition (Figure 3). This makes FIH an obvious choice for a PHD2 surrogate for crystallisation trials.

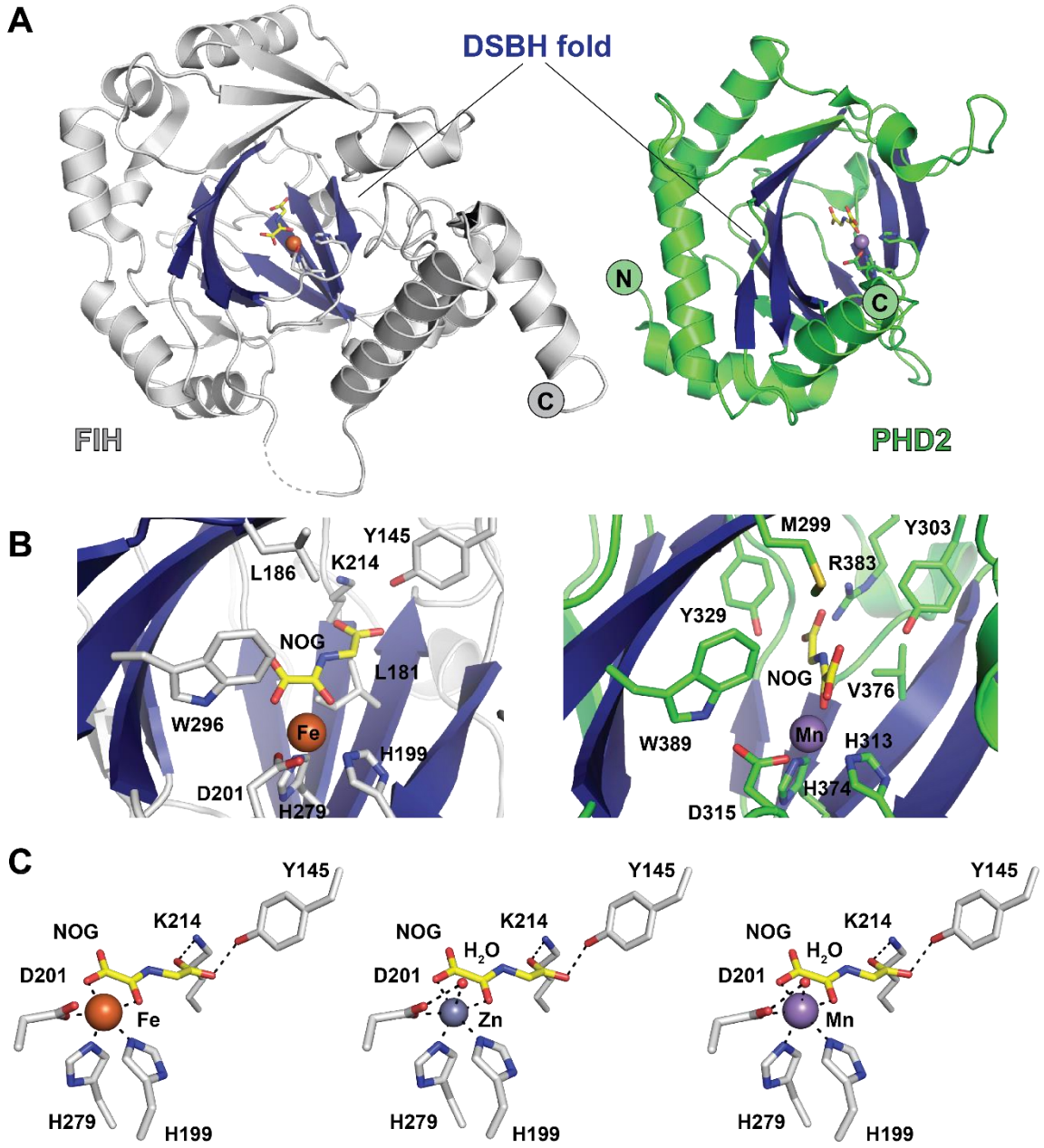


Figure 3 Comparison of crystal structures of FIH and PHD2. Views from crystal structures of FIH (2.15 Å, PDB code 1H2K[38]) and PHD2 (1.81 Å, PDB code 5L9R[39]). The central DSBH fold is coloured in dark blue. **A** Views of the overall fold of FIH and PHD2. **B** Active site views of FIH and PHD2 illustrate similarities and differences in their active sites. **C** Comparison of the active site metal coordination modes in FIH; iron (2.15 Å, PDB code 1H2K[38]), zinc (1.95 Å, see also chapter 5) and manganese (2.01 Å, see also chapter 5).

2.1.4 Structural basis for the inhibition of FIH

Since its discovery[40], considerable progress has been achieved in the structural characterisation of FIH. Several crystal structures of small molecules binding to the active site of FIH have been reported including several small natural products, TCA cycle intermediates, 2-hydroxyglutarate, and NOG (Figure 4).

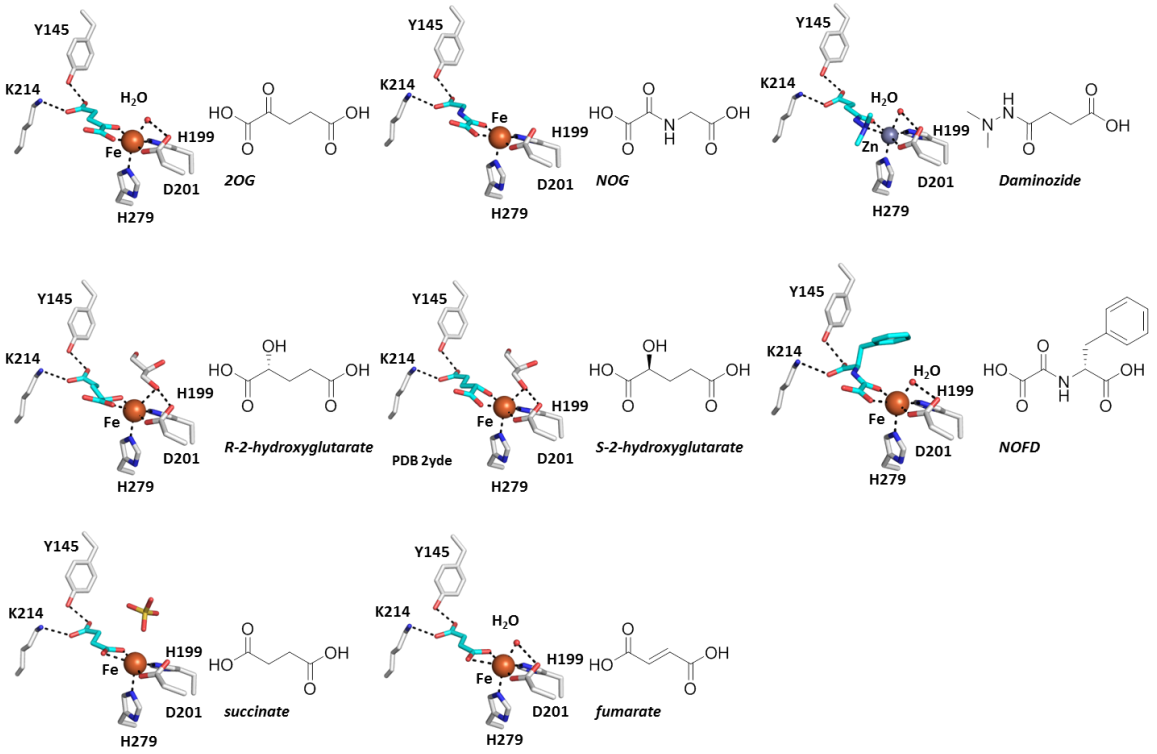


Figure 4 Views from crystal structures of FIH in complex with naturally occurring dicarboxylic acids. FIH in complex with 2OG (2.28 Å, PDB code 2Y0I[41]); FIH in complex with NOG (2.15 Å, PDB code 1H2K[38]), FIH in complex with succinate (2.40 Å, PDB code 2CGN[42]); FIH in complex with R-2-hydroxyglutarate (2.28 Å, PDB code 2YD0[43]); FIH in complex with S-2-hydroxyglutarate (2.28 Å, PDB: 2YDE[43]), FIH complex with NOFD (2.70 Å, PDB code 1YCI[44]) and FIH in complex with fumarate (2.30 Å, PDB code 2CGO[42]).

All of the naturally occurring dicarboxylic acids bind to FIH *via* electrostatic interactions with lysine-214 and tyrosine-145 which are located at the back of the active site; they also coordinate the active site metal in a bidentate manner (Figure 4). 2OG, NOG and 2-hydroxyglutarate, which all have a 5 carbon backbone, coordinate the metal via their C-1 oxygen *trans* to the proximal histidine (histidine-199) of the HxD/E…H motif and with their C-2 keto/hydroxyl group *trans* to the acidic metal binding residue (aspartate-201, Figure 4). This binding mode is typical for JmjC domain containing proteins, at least in the crystal state[7]. Succinate and fumarate, which both have a 4 carbon backbone, coordinate the central metal via their two C-1 oxygens *trans* to histidine-199 and aspartate-201. The coordination site of the metal *trans* to the distal histidine (histidine-279) is not occupied. None of the other dicarboxylic acids can be turned over by FIH, except 2OG. Therefore,

these binders can act as 2OG competitors and inhibitors of FIH activity. The NOG analogue N-oxalyl-D-phenylalanine (NOFD) is to date the only reported specific FIH inhibitor[44].

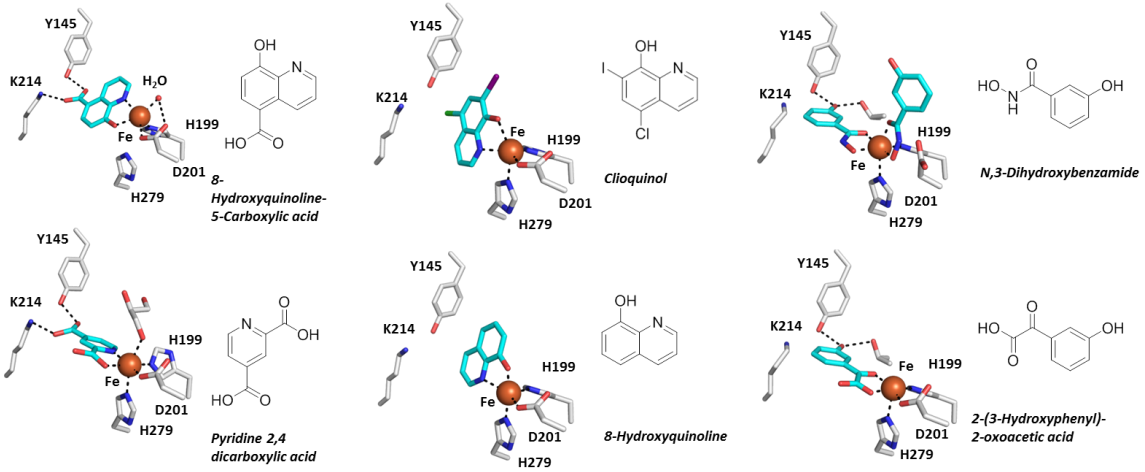


Figure 5 Views from crystal structures of FIH in complex small molecule inhibitors. FIH in complex with hydroxyquinoline-5-carboxylic acid (2.45 Å, PDB code 4BIO[45]); FIH in complex with cliquinol (2.60 Å, PDB code 3KCX[46]); FIH in complex with 2,4-PDCA (2.12 Å, PDB code 2W0X[47]); FIH in complex with 8-hydroxyquinoline (2.59 Å, PDB code 3KCY[46]); FIH in complex with daminozide (2.40 Å, PDB code 4AI8[48]); FIH in complex with 2-(3-hydroxyphenyl)-2-oxoacetic acid (2.50 Å, PDB code 2WA3[47]); FIH in complex with N,3-dihydroxybenzamide (2.50 Å, PDB code 2WA4[47])**.**

In addition to naturally occurring dicarboxylic acids, several structures of FIH in complex with small molecule inhibitors have been reported (Figure 5). Although structurally not related, nearly all of the reported structures of small molecule 2OG inhibitors in complex with FIH reveal analogous binding to 2OG (Figure 5). The active site metal is coordinated *via* one oxygen of the C1-carboxylate *trans* to histidine-199 and *via* an oxygen or nitrogen *trans* to aspartate-201. Exceptions include FIH binding to N,3-dihydroxybenzamide which includes, at least in the crystal state, two inhibitor molecules binding in the active site; the first molecule binds analogously to other overserved inhibitors whereas the second molecule displaces the active site water and aspartate-201. Clioquinol and 8-hydroxyquinoline share the same metal coordination motif as 8-hydroxyquinoline-5-carboxylic acid, but the structures reported show a different mode of metal coordination. The hydroxyl group of Clioquinol coordinates the central metal *trans* to histidine-279 whereas the hydroxyl group of 8-hydroxyquinoline-5-carboxylic acid coordinates the active site metal *trans* to histidine-199 (Figure 5). On close inspection of the structures, the exact binding modes should be considered speculative as the structures are of comparably low resolution, have high B-factors, and ambiguous electron density maps.

2.2 Structural basis for the inhibition of the PHDs by clinical drug candidates

2.2.1 Crystal structures of FIH in complex with GSK1278863 and Molidustat

Full length WT FIH was incubated in the presence of zinc(II) acetate (2 eq), substituting for iron(II) to avoid metal oxidation, and small molecule inhibitor (from Yeh et al.[7] ,DMSO stocks, 0.1 M, final concentration: 3-7 eq; >4% DMSO) for 30-60 min on ice prior to crystallisation trials (0.1 M HEPES pH = 7.5, 1.1 M - 2.2 M ammonium sulphate, 2.5 - 6.0% PEG400), using the sitting drop vapour diffusion technique. Crystals appeared within 24 h and grew to full size within 48 h. Crystals of the FIH.Molidustat complex were stable for several weeks at room temperature whereas crystals of the FIH.GSK1278863 complex started to degrade after 72 h (by visual inspection). Data from single crystals were collected using Diamond MX beamlines; detailed data collection and refinement statistics are shown in Table 2. For detailed crystallography experimental procedures see chapter 10.

Solution of crystal structures of FIH in complex with Molidustat and GSK1278863 (using molecular replacement, search model PDB code 1H2K[38]) revealed clear density corresponding to the ligands, enabling the unambiguous modelling of Molidustat and GSK1278863 into the FIH active site (Figure 6). Previous reported crystal structures of structurally related compounds in complex with PHD2 suggested that the triazole ring of Molidustat might bind to the 2OG C-5 binding pocket[16]. Similar binding was also expected for the N-glycinamide group of GSK1278863[49] (Figure 1). Somewhat surprisingly, it was found that Molidustat coordinates to the active site metal *via* the nitrogens of its pyrazolone (*trans* to aspartate-201) and pyrimidine rings (*trans* to histidine-279). Interestingly, GSK1278863 was found to bind the active site metal *via* its amide oxygen (*trans* to aspartate-201) and one oxygen of its pyrimidine-trione rings (*trans* to histidine-279), in a general manner similar to Molidustat. This metal coordination is surprising as in previous structural studies, nearly all inhibitors binding to the active site coordinate the central metal in a bidentate manner *trans* to aspartate-201 and histidine-199 and not histidine-279 (see section 2.1.4). The metal-coordination site *trans* to histidine-279 is proposed to be the oxygen binding site of FIH (see chapter 1). The observed metal coordination is closer to that observed for the inhibitors binding to the active site of DAOCS family members, *e.g*. PHD2, see below. Higher steric demands of the clinical inhibitors compared to previously investigated compounds may contribute to this “switch” in metal coordination. Notably, there is a difference in the “metal binding planes” (as defined by the pyrazolone and pyrimidine rings of Molidustat and the pyrimindine-trione rings of GSK1278863) of ~20 °. The steric demands of the cyclo-hexyl groups of

GSK1278863 might force the compound to bind at a steeper angle compared to Molidustat to minimize steric clashes (Figure 6).

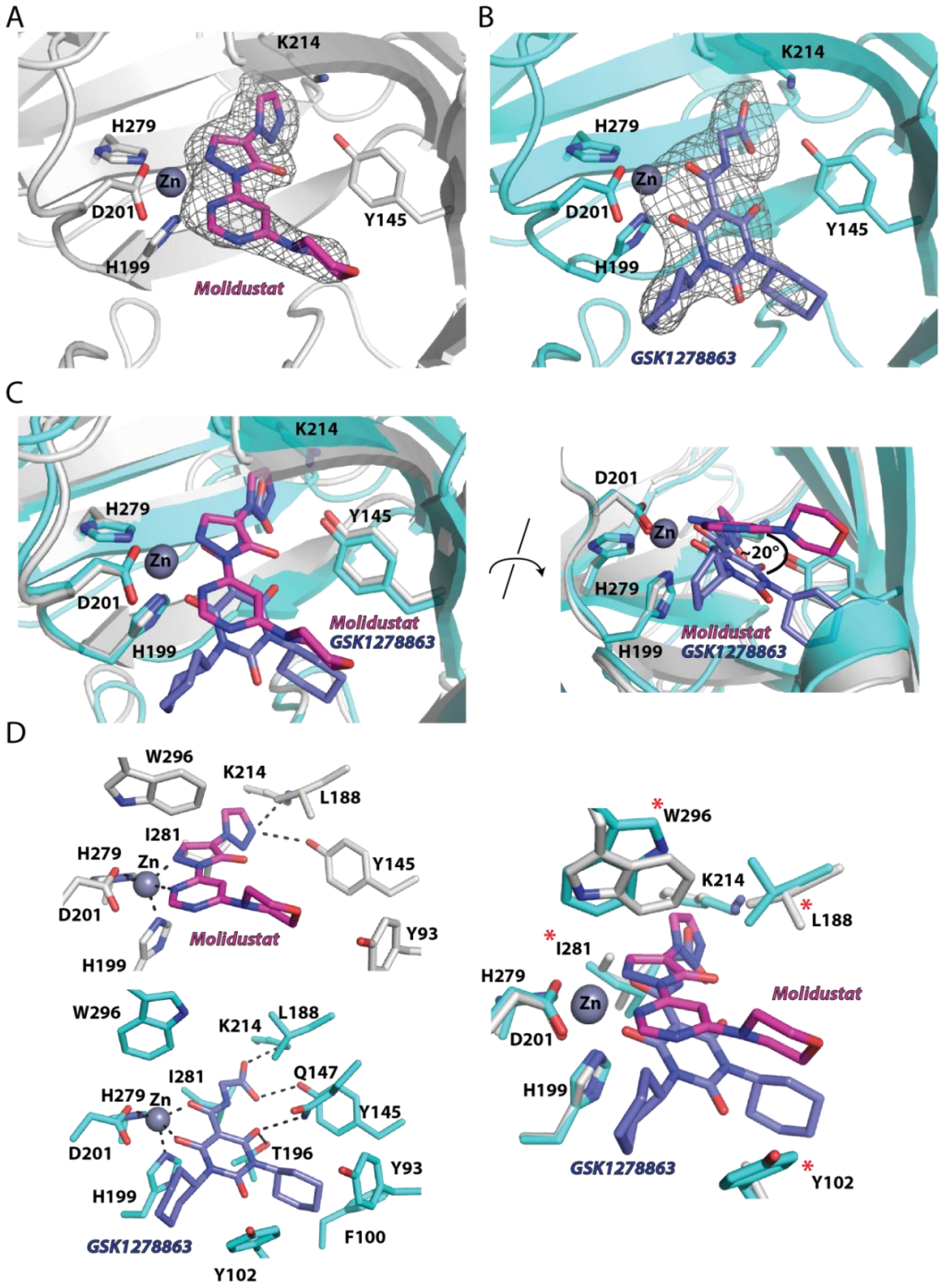


Figure 6 Views from crystal structures of FIH in complex with Molidustat and GSK1278863. Figure reprinted from Yeh et al[7]. **A, B:** Views from crystal structures of Molidustat and GSK1278863 in complex with FIH showing the mFo-Fc OMIT maps contoured to 3σ around the ligand. **C** Overlay of crystal structures of FIH in complex with Molidustat and GSK1278863 illustrates similarities and differences in inhibitor binding. **D** Views from crystal structures of Molidustat and GSK1278863 in complex with FIH illustrate active site interactions and local conformational changes in the active site induced by ligand binding.

Table 2 Data collection and refinement statistics of FIH complexes with "clinical" PHD inhibitors Molidustat and GSK1278863.

Datasets	**FIH**.Zn(II).Molidustat	**FIH**.Zn(II).GSK1278863
PDB code	**5OP8**	**5OP6**
Data Collection		
Beamline (Wavelength, Å)	DLS I04 (0.9795)	DLS I04 (0.9795)
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	MOSFLM/SCALA	MOSFLM/SCALA
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	86.2, 86.2, 148.2	86.2, 86.2, 147.1
α,β,γ(°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. reflections	25593 (3657)*	21126 (3014)*
Resolution (Å)	30.47-2.30 (2.41-2.30)*	30.48-2.45 (2.58-2.45)*
R_{merge}	0.148 (1.229)*	0.138 (1.253)*
I/σI	8.9 (1.9)*	11.5 (2.0)*
CC1/2	0.995 (0.695)*	0.998 (0.570)*
Completeness (%)	99.5 (100)*	100 (100)*
Multiplicity	11.3 (11.5)*	11.5 (11.1)*
Wilson B value (Å^2)	48.87	52.45
Refinement	REFMAC /PHENIX	REFMAC /PHENIX
R_{work}/R_{free}†	0.1992/0.2142	0.2037/ 0.2274
No. atoms		
-Enzyme (A/B)	2664	2619
-Ligand	23	28
-Water	62	46
Avg. B-factors		
Enzyme	71.7	66.5
Ligand	109.1	67.4
Water	60.9	54.9
R.m.s deviations		
-Bond lengths (Å)	0.012	0.005
-Bond angles (°)	0.664	0.569

* Highest resolution shell in parentheses.
†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

2.2.2 Crystal structures of FIH in complex with Vadadustat and FG-4592

2.2.2.1 Challenges in determining the crystal structures of FIH in complex with Vadadustat and FG-4592

Vadadustat and FG-4592 (from Yeh et al.[7]) were co-crystallised with FIH as described for the FIH-Molidustat complex (see chapter 2.2.1). As the resulting electron densities for the bound ligand could not be unambiguously modelled, soaking experiments were performed to rule out potential degradation of the compound under the crystallisation conditions. Stock solutions of Vadadustat and FG-4592 (100 mM) were added to the cryo-protectant solution (final concentration 4 mM inhibitor and 4% DMSO, 20% glycerol in well solution). The crystals were soaked (1-10 min) before data were collected using either MX beamlines at Diamond Light Source or an in-house X-ray source to reduce potential radiation damage. However, these strategies did not improve information on the ligand density (data not shown).

2.2.2.2 Determining multiple confirmations of Vadadustat in complex with FIH using PANDDA event maps

Upon solving the structure of FIH in complex with Vadadustat and FG-4592, the bound ligand could not be modelled unambiguously. In particular, the undefined density for the chlorophenyl group of Vadadustat suggested some degree of flexibility, *i.e*. multiple conformations. To investigate this further, and to potentially determine low occupancy conformations of the chlorophenyl group, the recently published Pan-Dataset Density Analysis (PANDDA) was employed (Pearce et al 2017[50]). Firstly, FIH was crystallised in the presence of zinc but in the absence of any inhibitor. Several of these “apo”-crystal were sent to the fully automated beamline MASSIF1 at the European Synchrotron Radiation Facility (ESRF) for data collection. In total 8 crystals diffracted to 2.6 Å or better and were used to determine an averaged ground state. Using the ground state maps and the maps obtained from the data collections of the FIH.Vadadustat and FIH.FG-4592 crystals, PANDDA event maps were generated. Using the PANDDA event maps, two major conformations of the chlorophenyl group were identified and modelled (Figure 7). The structures suggest that the chlorophenyl group is likely stabilised *via* a weak hydrogen bond to arginine-328 and interactions with tryptophan-298. It is possible that there are further low occupancy conformations not visible in the event maps.

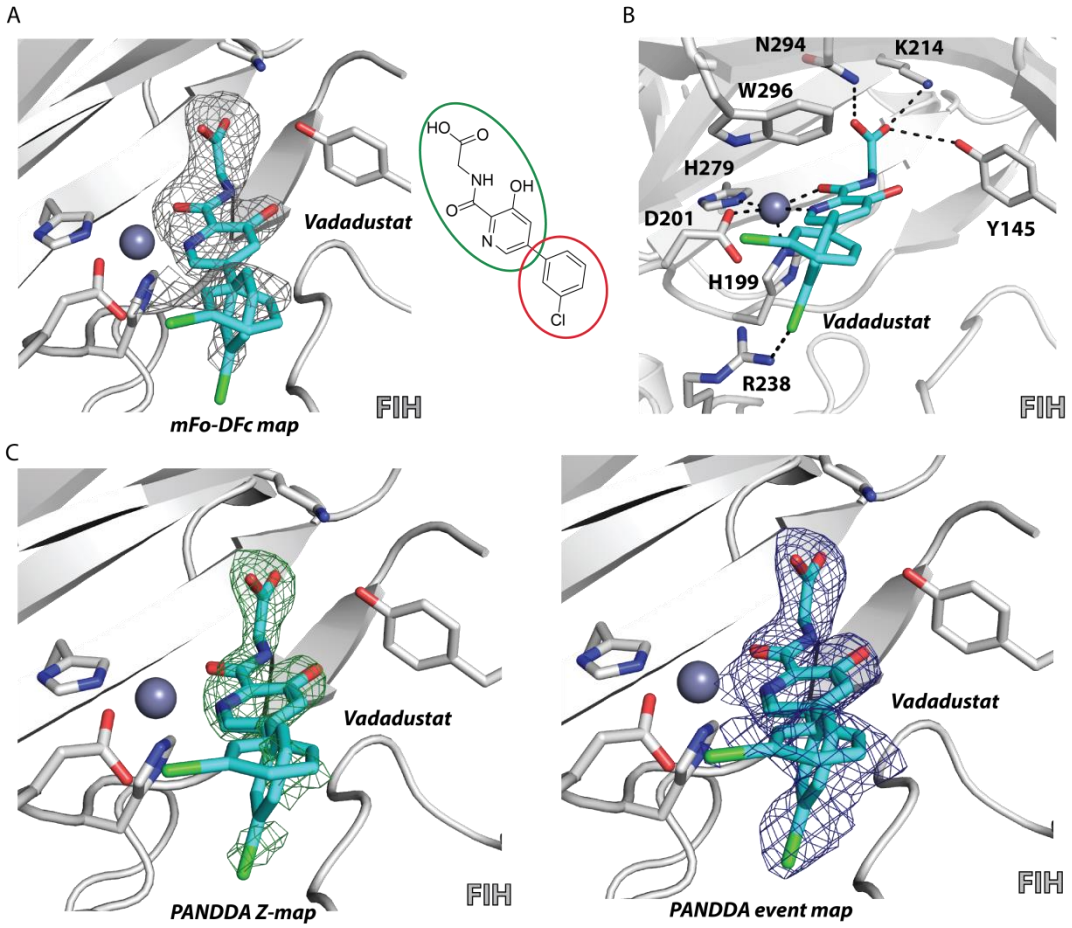


Figure 7 Views from a crystal structure of FIH in complex with Vadadustat. Figure reprinted from Yeh et al[7]**. A** Views from a crystal structure of FIH in complex with Vadadustat showing the mFo-Fc OMIT map contoured to 3σ around the ligand. **B** Views from a crystal structure of FIH in complex with FIH illustrates the interactions of Vadadustat with active site residues of FIH. **C** Left: the PANDDA Z-map contoured to 6σ; Right: the PANDDA event map at 2.6 Å resolution. The chloride of Vadadustat shows less density in the PANDDA map than expected, likely resulting from background solvent subtraction. Solvent exposed parts of ligands are less dominant in PANDDA event maps[50].

2.2.2.3 Comparison of crystal structures of FIH and PHDs in complex with Vadadustat

Following the structural studies on FIH-Vadadustat interactions, a crystal structure of $PHD2_{181-426}$ in complex with manganese and Vadadustat was solved[7]. A comparison of the binding mode of Vadadustat to FIH and PHD is shown in Figure 8.

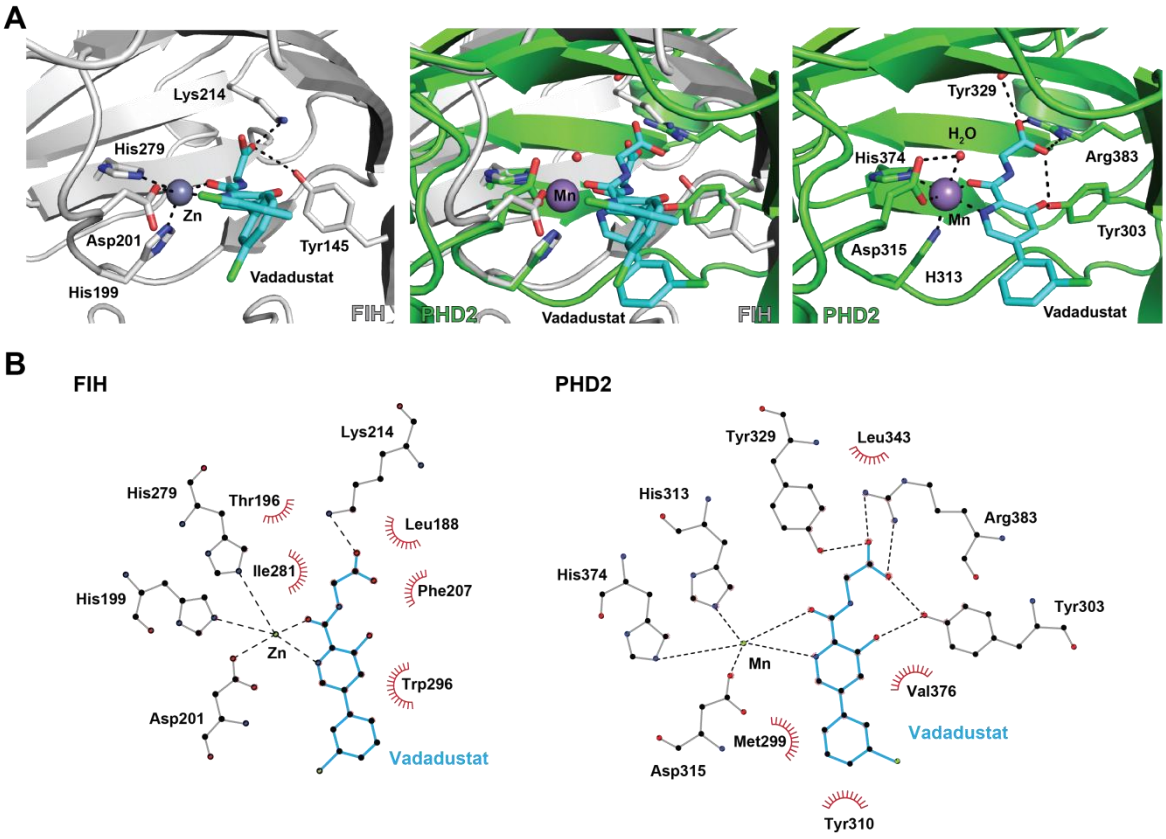


Figure 8 Comparison of crystal structures of FIH and PHD2 in complex with Vadadustat. **A** Views from crystal structures of FIH in complex with Zn and Vadadustat and $PHD2_{181-426}$ in complex with Mn and Vadadustat (1.99 Å, PDB code 5OX6[7]). **B** 2D representation of active site interactions of Vadadustat in complex with FIH and PHD2. Images generated using LIGPLOT[51].

Comparing crystal structures of Vadadustat in complex with FIH and PHD2 in complex with Vadadustat reveals overall conservation of the binding mode (Figure 8). With both FIH and PHD2, the inhibitor chelates the active site metal in a bidentate manner with its pyrimidine ring nitrogen and the N-glycinamide oxygen *trans* to the aspartate and distal histidine of the HxD/E…H motif respectively. The carboxylate of the N-glycinamide group interacts with the 2OG C-5 carboxylate binding pocket in the back of the FIH/PHD2 active site. The phenoxy group points into the active site, a larger degree of rotational freedom is apparent in FIH. Nevertheless, the binding of Vadadustat to FIH largely mirrors that observed in PHD2.

2.2.2.4 Attempts to determine a crystal structure of the FIH.FG-4592 complex

Following from the structural characterization of GSK1278863, Molidustat and Vadadustat in complex with FIH, attempts were made to solve a crystal structure of FIH in complex with FG-4592. No structural information was previously available on FG-4592 binding to 2OG oxygenases. Attempts to co-crystallise FIH with FG-4592 have as yet been unsuccessful. Solving the crystal structure from co-crystallisation trials of FIH with FG-4592 did not reveal clear electron density in the active site of FIH corresponding to the bound ligand. The solubility of the compound proved to be a challenge. It was then attempted to soak the compound into “apo”-FIH crystals. Solving the structure revealed electron density in the active site which originates most likely from the ligand. The electron density maps for the FIH.FG-4592 complex show clear density for the N-glycinamide and the pyridinyl part of the quinoline rings but insufficient density for the second of the quinoline rings and very little density for the phenoxy group (Figure 9).

The phenoxy group of FG-4592 is expected to be flexible as observed for the chloro-phenyl group of Vadadustat (see Figure 7). The quinoline rings however do not show substantial flexibility. Firstly, stability of the ligand in the presence of 2OG oxygenases were confirmed by NMR (Dr. Martine Abboud, University of Oxford). Reactions of quinolone rings with 2OG oxygenases have been observed with FG-2216, where a chlorine (not present in FG-4592) undergoes a nucleophilic aromatic substation by sulphur of a surface cysteine[52]. FG-4592 has a methyl group at the analogous position (see Figure 2) which is chemically inert to this reaction. Further, FIH does not display any cysteines at the surface around the active site and the electron density observed in several crystal structures of the FIH.FG-4592 complex did not indicate any reactions of FG-4592 with FIH.

To identify possible low occupancy states, PANDDA analysis was performed as previously described (see above). The PANDDA event map is shown in Figure 9B. Unlike in the FIH.Vadadustat complex, the PANDDA event maps did not improve the ligand density and the ligand could not be modelled unambiguously. As the crystal structure of FG-2216 in complex with PHD2 is known and the binding mode of Vadadustat to FIH could be determined, further attempts to solve a crystal structure of FG-4592 bound to FIH were not pursued.

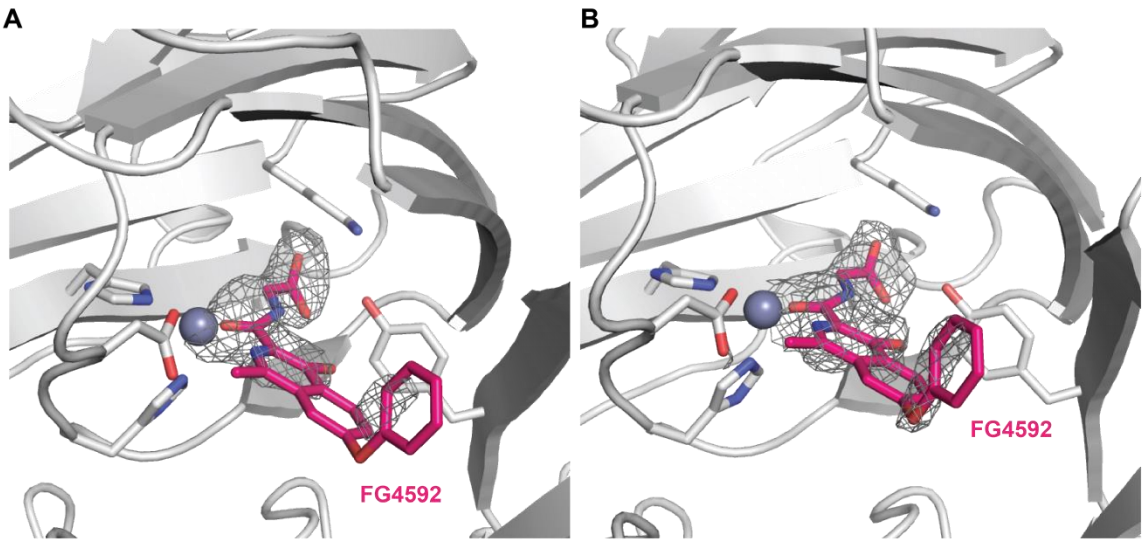


Figure 9 Views from a crystal structures of FIH from crystallization and soaking trials. A mFo-Fc OMIT maps contoured to 3σ around the ligand. **B** PANDDA event map with 0.27 BDZ contoured to 3 σ around the presumed binding mode of FG-4592.

Table 3 Data collection and refinement statistics of FIH complexes with "clinical" PHD inhibitors Vadadustat and FG-4592.

Datasets	**FIH**.Zn(II).Vadadustat	**FIH**.Zn(II).FG-4592
PDB code	**5OPC**	---
Data Collection		
Beamline (Wavelength, Å)	DLS I03 (0.9763)	In-house (RIGAKU FR-E Superbright) (1.5419)
Detector	PILATUS 6M-F	Saturn 944+
Data Processing	MOSFLM/SCALA	HKL2000
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	86.8, 86.8, 149.5	86.6, 86.6, 148.0
α,β,γ(°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. reflections	26216 (3748)*	29198 (1420)*
Resolution (Å)	86.80-2.30 (2.42-2.30)*	50.0-2.20 (2.24-2.20)*
R_{merge}	0.152 (1.834)*	0.085(955)*
I/σI	10.9 (1.8)*	19.4 (2.1)*
CC1/2	0.998 (0.689)*	(0.822)*
Completeness (%)	100 (100)*	99.3 (100)*
Multiplicity	17.5 (17.5)*	8.2 (8.0)*
Wilson B value (Å^2)	55.78	51.05
Refinement		
R_{work}/R_{free}†	0.1925/0.2140	------
No. atoms		------
-Enzyme (A/B)	2697	------
-Ligand	21	------
-Water	72	------
Avg. B-factors		------
Enzyme	74.0	------
Ligand	93.1	------
Water	63.1	------
R.m.s deviations		------
-Bond lengths (Å)	0.003	------
-Bond angles (°)	0.522	------

* Highest resolution shell in parentheses.
†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

2.2.3 Comparison of the binding modes of clinical candidates to FIH with other 2OG oxygenases

Structures of Molidustat, GSK1278863 and Vadadustat in complex with FIH reveal a similar binding mode for all three compounds. The clinical PHD inhibitors bind to the active site and coordinate the central metal in a bidentate manner *trans* to aspartate-201 and histidine-279, the distal histidine of the HxD/E…H motif. The same metal coordination is also observed for crystal structures of 2OG and CCT6 in complex with PHD (Figure 10) where the central metal is coordinated *trans* to aspartate-315 and histidine-374.

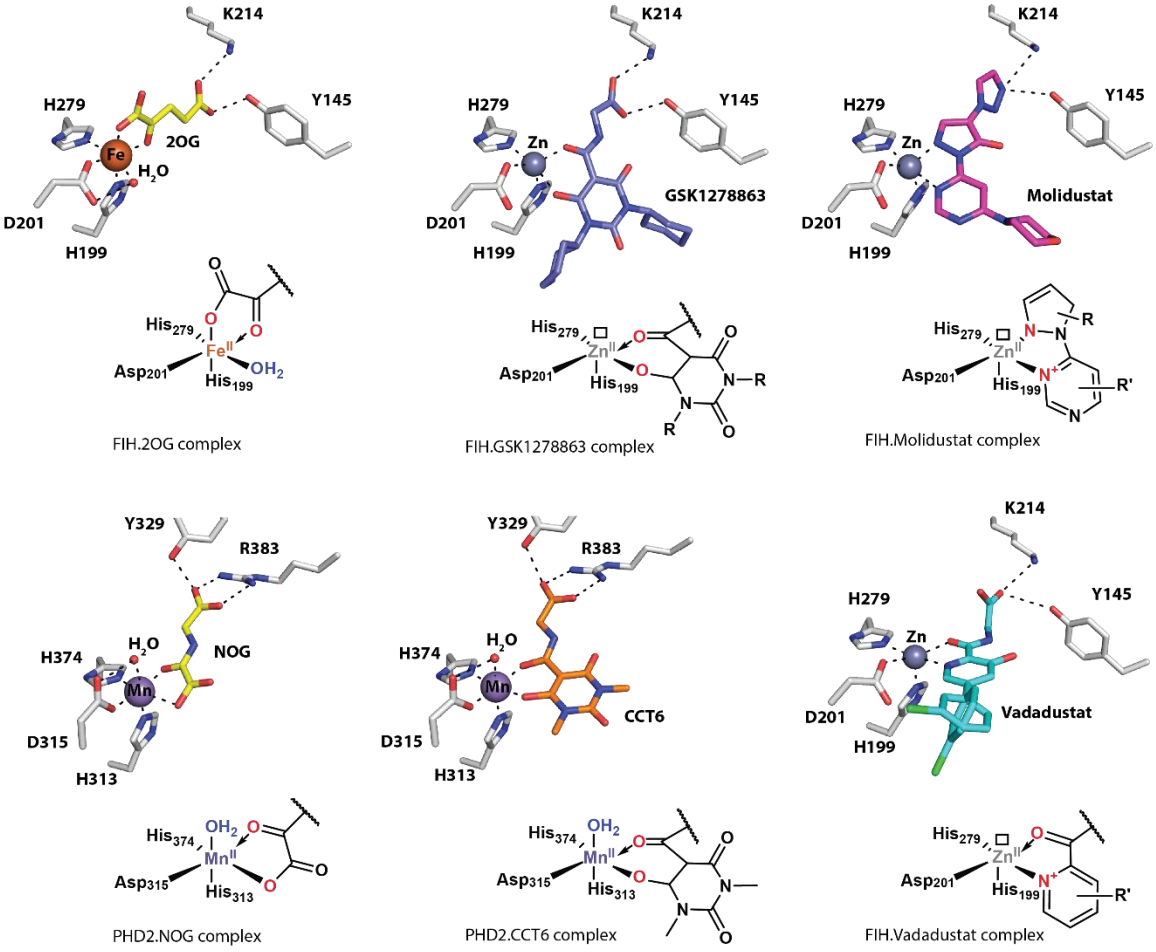


Figure 10 Views from crystal structures and schematic metal coordination of FIH and PHD in complex with 2OG and various inhibitors. Figure modified from Yeh et al[7]. Views from a crystal structure of FIH in complex with 2OG (2.28 Å, PDB code 2Y0I[41]), FIH in complex with GSK1278863 (2.45 Å); FIH in complex with Molidustat (2.30 Å); PHD2 in complex with NOG (2.30 Å, PDB code 3HQR[53]), PHD in complex with CCT (2.30 Å, PDB code 5OX5[7]), FIH in complex with Vadadustat (2.30 Å). Skeletal views of the active site interactions illustrate the metal coordination mode. Unoccupied coordination sites are marked with boxes.

2.2.4 Models for PHD2 binding and mode of inhibition of clinically used PHD inhibitors

Crystal structures of Molidustat, GSK1278863 and Vadadustat in complex with FIH revealed the binding modes of the inhibitors in the active site of FIH, at least in the crystal state, which largely mimics the binding mode of 2OG in the PHD2 active site (see Figure 10). This would suggest that the binding mode of each inhibitor to FIH may represent a possible mode of binding to PHD2. Using the structural information gained from the FIH-inhibitor complex structures as well as crystal structures of PHD2 in complex with structurally related compounds, models of the inhibitors binding to PHD2 were generated (Figure 11).

The models of the potential binding modes of clinically used PHD inhibitors shown in Figure 11 suggest that all of the inhibitors bind to the active site in the 2OG binding pocket. This means that all of the inhibitors could potentially compete with the 2OG binding to PHD2. Besides occupying the 2OG binding pocket by the N-glycinamide group of FG-4592, Vadadustat and GSK1278863 or the pyrazolone and triazole rings of Molidustat, the inhibitors vary significantly in their further interactions with the active site and substrate binding site. Molidustat, on the one hand, appears to occupy very little space around the active site. The cyclohexyl rings of GSK1278863 however, force the compound to bind to FIH at a ~20 ° angle compared to Molidustat; this is most likely due to steric clashes with the DSBH core fold (see Figure 6). Steric clashing with the DSBH fold is also be expected for GSK1278863 binding to PHD2, comparable to that observed for the FIH.GSK1278863 complex. In order to bind in a similar fashion to that observed for GSK1278863 binding to FIH and CCT6 binding to PHD2, the active site of PHD2 would have to rearrange to accommodate space for the sterically demanding cyclohexyl groups. Docking attempts were made but results were not reliable (data not shown). Rearrangements around the active site of PHD2 might be a reason why a stable PHD2.GSK1278863 complex did not crystallise. Indeed, the evidence suggests that small molecule binding to PHD2 may lead to conformational changes around the active site of PHD2 which interferes with efficient crystal packing[7].

Notably, the degree to which the inhibitor projects into the substrate binding site varies significantly. Molidustat shows little to no overlap with the HIF-1α binding site. In contrast, the phenoxy group of FG-4592 is predicted to overlap significantly with HIF-1α (Figure 12).

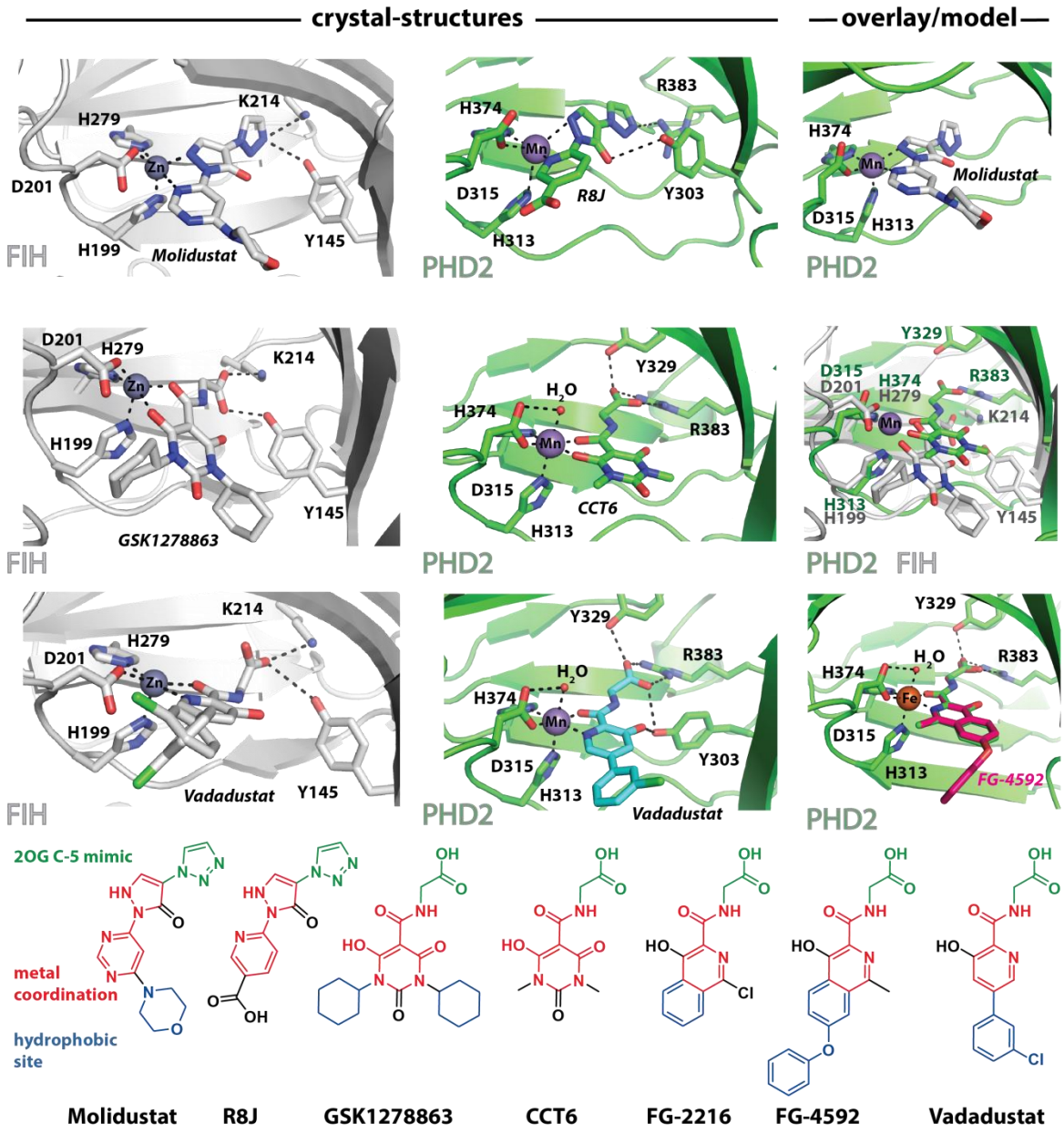


Figure 11 Binding models of clinically used PHD inhibitors to PHD2. Figure modified from Yeh et al[7]. **Top:** View from a crystal structure of FIH in complex with Molidustat; PHD2 in complex with IOX4 derivative R8J (3.30 Å, PDB code 5A3U[16]); Model of PHD2.Molidustat complex generated from overlay of the FIH.Molidustat complex and PHD2.R8J complex. **Middle:** Views from a crystal structure of FIH in complex with GSK1278863; PHD2 in complex with GSK1278863 derivative CCT6 (2.30 Å, PDB code 5OX5[7]); overlay of the FIH.GSK1278863 complex structure and the PHD2.CCT6 complex structure; **Bottom:** Views from a crystal structure of FIH in complex with Vadadustat; PHD2 in complex with Vadadustat (1.99 Å, PDB code 5OX6[7]); Model of PHD.FG-4592 complex generated from overlay of PHD2.FG-2216 complex with FG-4592. The three motifs present in all clinical PHD inhibitors are highlighted in green, red and blue.

To validate the structural findings in PHD2, NMR based substrate displacement studies with the catalytic domain of PHD2 with 2OG and its HIF-1α substrates, performed by Dr. Martine Abboud (University of Oxford), confirmed that Molidustat as well as FG-4952 can fully displace 2OG from the active site of PHD2 (Figure 12).

Notably, Molidustat shows little to no displacement of either HIF-1α NODD or CODD, whereas FG-4592 shows near complete displacement of HIF-1α NODD and considerable displacement of CODD under the experimental conditions, for details see Yeh et al (2017)[7].

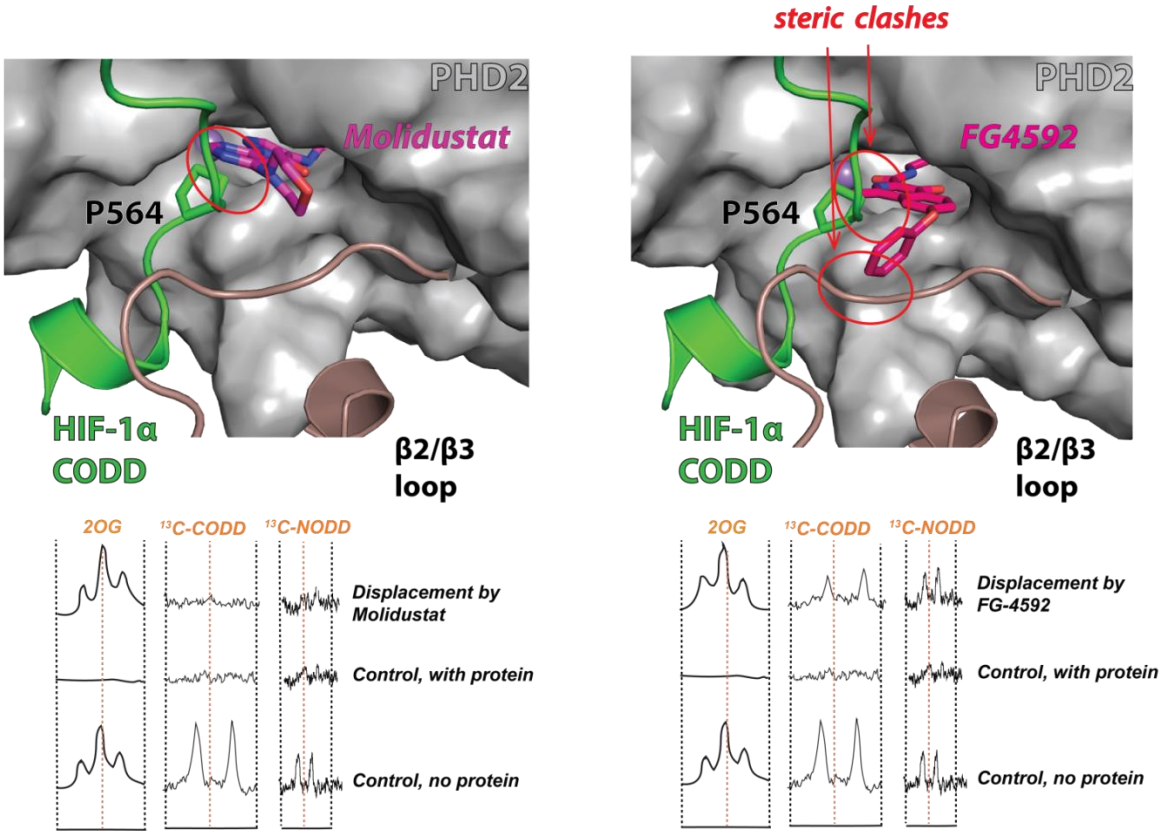


Figure 12 Binding models of clinical candidates to PHD2 suggest variations in their capability to compete with the HIF substrate. Top: Models of PHD inhibitors binding to the active site of PHD2 superimposed with HIF-1α CODD substrate. Bottom: NMR substrate displacement measurement adapted from Yeh et al[7].

In summary, it was shown that all of the clinically used PHD inhibitors bind to the active site of PHD2 and coordinate the central metal in a bidentate manner. Crystal structures of the PHD inhibitors in complex with FIH used as a PHD surrogate, predict the binding mode of known inhibitors to PHD2 and were a useful tool for structural analysis of otherwise non-crystallising protein-ligand complexes. All of the inhibitors occupy the 2OG binding pocket, outcompeting the co-substrate 2OG. Molidustat shows little interactions with, and occupancy of, the HIF-1α substrate binding site and most likely only functions as a 2OG competitor. GSK1278863, Vadadustat and FG-4592 occupy to some degree the HIF-1α substrate binding site and have, at least in some cases, a 2OG and substrate competition mode of action. In solution data with the catalytic domain of PHD2 supports the binding modes observed in the crystalline state.

2.3 FIH in complex with non-clinical PHD inhibitors

2.3.1 Structures of FIH in complex with N-glycinamide test compounds

To further investigate if the observed binding modes for the clinically used PHD inhibitors to FIH are specific to the investigated compounds, a set of N-glycinamide containing 2OG oxygenase inhibitors (kindly provided by Amgen) was further investigated. Three compounds were co-crystallised with FIH according to the previously described methods (see chapter 2.2) and the structures were solved, as previously described. Notably, all the inhibitors are structurally related to GSK1278863 as they incorporate an N-glycineamide group adjacent to a keto-enol ring system (Figure 13).

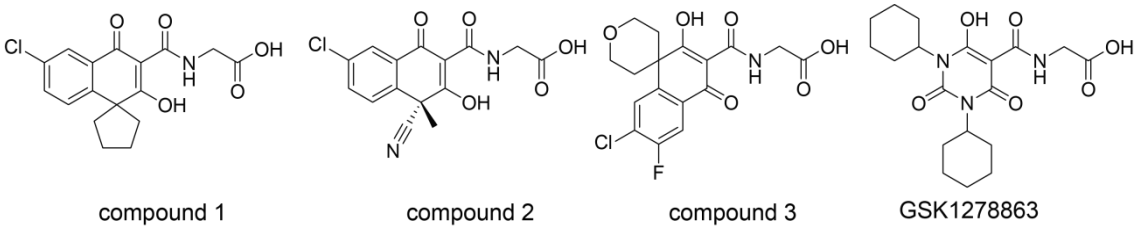


Figure 13 Structures of the compounds crystallized with FIH illustrating structural similarities with GSK1278863.

Upon solving the structures of FIH in complex with the three test compounds, clear electron densities in the active site corresponding to the inhibitors were observed; the inhibitors could be modelled unambiguously (Figure 14).

All of the compounds were found to bind to FIH in a conserved binding mode similar to that previously observed for the FIH.GSK1278863 complex. The central metal was found to be coordinated in a bidentate manner *via* the amide oxygen (*trans* to aspartate-201) and one oxygen corresponding to the pyrimidine-trione ring of GSK1278863 (*trans* to histidine-279) and the carboxylate of the N-glycinamide group is interacting with the 2OG C-5 binding pocket.

A notable difference in the structures, however, is the angular difference of the metal binding plane (as defined by the pyrimindine-trione ring and the glycinamide oxygen of GSK1278863, Figure 14). Compound 1 shows a similar metal binding plane compared to GSK1278863 whereas compound 3 shows a comparable binding plane to Vadadustat. Both compounds show similar steric demands. A key difference between the two compounds it the cyclo-hexyl-substitute of compound 1 and the tetrahydropyran substitute of compound 3. The oxygen of the tetrahydropyran forms a hydrogen bond with arginine-238 which is involved in substrate binding (see chapter 1). This might be of interest for the design of novel FIH inhibitors that can also function as substrate competitors.

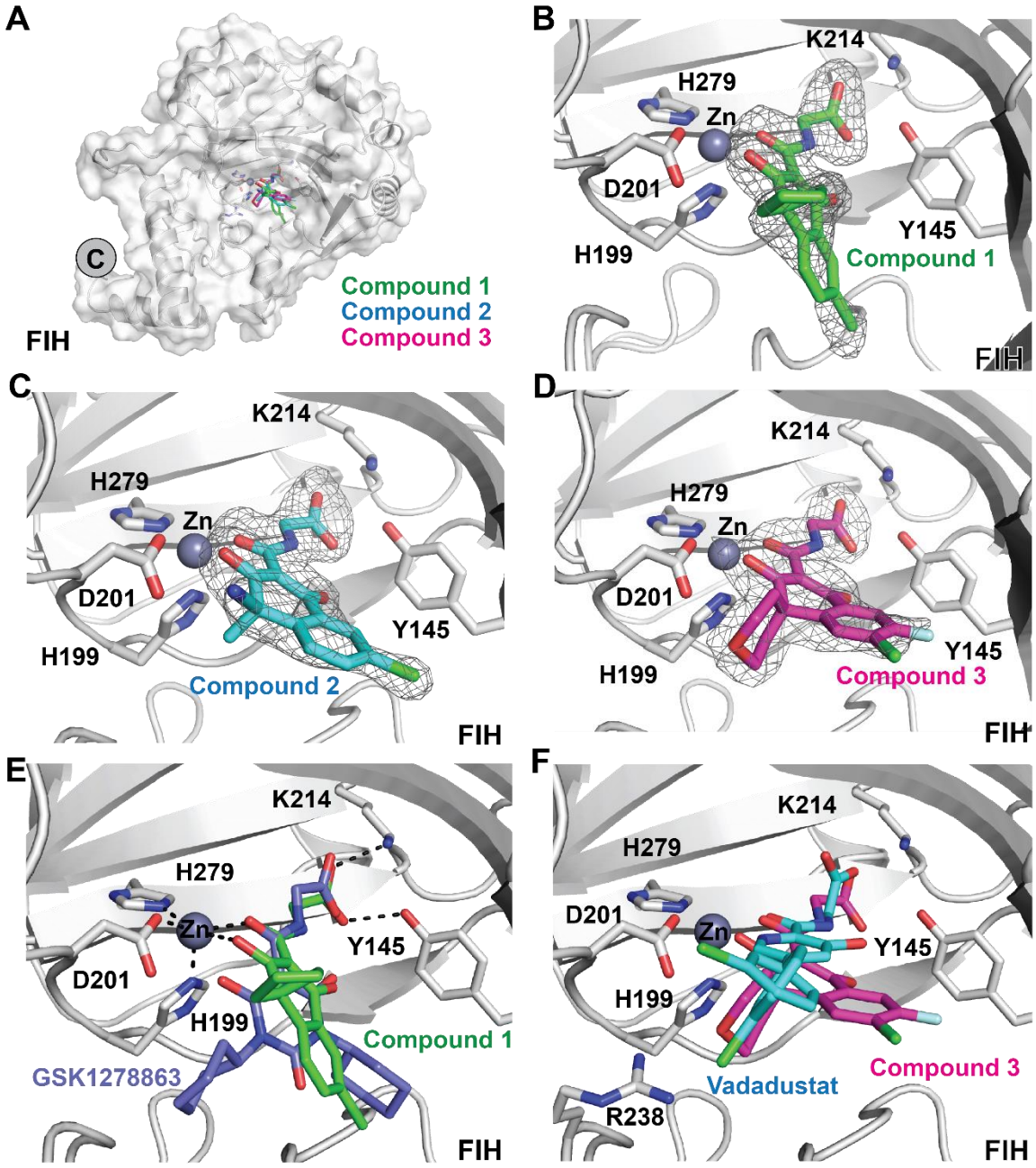


Figure 14 Views from crystal structures of FIH in complex with test compounds 1, 2 and 3. **A** View from a crystal structure of FIH in complex with compounds 1, 2 and 3. **B-D** Views from crystal structures of FIH in complex with compound 1, 2 and 3 illustrate the active site binding modes. The mFo-Fc OMIT maps are contoured to 3σ around the ligands. **E,F** Overlay of compound 1 with GSK1278863 and overlay of compound 3 with Vadadustat illustrate differences in the binding planes of the inhibitors.

Table 4 Data collection and refinement statistics of FIH complexes with non-clinical PHD inhibitors.

Datasets	**FIH**.Zn(II).Compound 1	**FIH**.Zn(II).Compound 2	**FIH**.Zn(II).Compound 3
Data Collection			
Beamline (Wavelength, Å)	In-house (RIGAKU FR-E Superbright) (1.5419)	In-house (RIGAKU FR-E Superbright) (1.5419)	In-house (RIGAKU FR-E Superbright) (1.5419)
Detector	Saturn 944+	Saturn 944+	Saturn 944+
Data Processing	HKL2000	HKL2000	HKL2000
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.1, 86.1, 147.5	86.4, 86.4, 146.4	86.7, 86.7, 147.4
α,β,γ(°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. of molecules/ASU	1	1	1
No. reflections	30955 (3038)*	30897 (3035)*	24318 (2431)*
Resolution (Å)	24.78-2.15 (2.23-2.15)*	23.96-2.15 (2.20-2.15)*	24.92-2.30 (2.38-2.30)*
R_{merge}	0.101 (>1.0)*	0.094 (>1.0)*	0.123 (>1.0)*
I/σI	13.0 (1.3)*	18.9 (1.5)*	8.2 (1.1)*
Completeness (%)	99.6 (100)*	99.7 (100)*	94.7 (97.0)*
Multiplicity	5.6 (6.0)*	8.7 (8.9)*	4.1 (4.2)*
Wilson B value (Å^2)	45.6	45.5	50.2
Refinement			
R_{work}/R_{free}	0.2067/0.2377	0.2251/0.2507	0.2279/0.2416
No. atoms			
-Enzyme (A/B)	2706	2648	2662
-Ligand	23	24	26
-Water	107	64	75
Avg. B-factors			
Enzyme	67.9	54.8	63.6
Ligand	82.6	53.7	70.7
Water	60.8	47.3	60.1
R.m.s deviations			
-Bond lengths (Å)	0.002	0.004	0.003
-Bond angles (°)	0.548	0.690	0.585

* Highest resolution shell in parentheses.

† $R_{factor} = \Sigma_{hkl} | |F_{obs}(hkl)| - k|F_{calc}(hkl)| | / \Sigma_{hkl} |F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

2.3.2 Structure of FIH in complex with IOX4

Following from structural studies on the clinical PHD inhibitors and a set of additional N-glycinamide PHD inhibitors (see chapter 2.2 and 2.3.1), further structural studies were performed to investigate whether IOX4 shows the same binding mode as that predicted from the previously studied inhibitors. As one of the best studied PHD inhibitors, at least in the published literature[7], IOX4 is well suited to inform on potency and selectivity of HIF hydroxylase inhibitors. To date, there is no published structural information on the binding mode of IOX4 to any 2OG oxygenases. A structure with the structurally related inhibitor R8J (see chapter 2.1) is however reported.

Initially, FIH was incubated with IOX4 (from Yeh et al.[7]) and crystallisation trials were performed under optimised conditions as previously described for the crystallisation of FIH with the clinical PHD inhibitors (see chapter 2.2). Surprisingly, no crystals in the characteristic tetragonal space group $P4_12_12$ with bipyramidal morphology (for FIH crystal morphology see also chapter 8) were observed. However, orthorhombic crystals were observed. Processing of the data confirmed that the FIH-IOX4 complex crystallised in the $P2_12_12_1$ space group with 4 molecules per asymmetric unit and diffracted to 2.70 Å resolution (see Table 5).

Crystals of FIH in complex with small molecule inhibitors in the tetragonal crystal form $P4_12_12$ show higher resolution, up to 2.00-2.30 Å under similar conditions. All published FIH structures were crystallised in the tetragonal space group. Consequently, it was tested whether IOX4 could be soaked into tetragonal crystals of the FIH.Zn complex. FIH was crystallised in the absence of small molecule inhibitors (see also chapter 2.2). But, addition of IOX4 to the soaking solution for the FIH crystals let to immediate disintegration of the crystals. Very short soaks (>1 min) did not yield in any detectable density for the ligand in the active site, longer soaks broke the crystal and no data could be collected. The effects were independent of the final DMSO concentration of the soaking conditions.

Processing the data from the FIH-IOX4 complex in the orthorhombic form enabled the structure to be solved and the ligand to be modelled into the active site (see Figure 15).

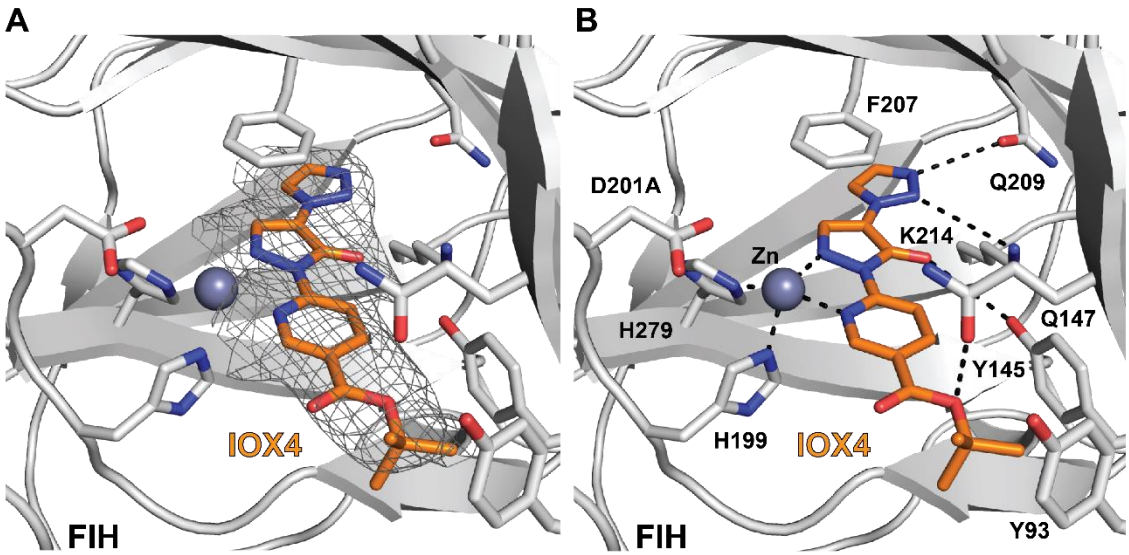


Figure 15 Views from a crystal structure of FIH in complex with IOX4. A The mFo-Fc OMIT map is contoured to 3σ around IOX4. **B** View from the active site of FIH with bound ligand illustrates the interactions of IOX4 with active site residues of FIH.

IOX4 was found to coordinate the active site metal, like Molidustat, in a bidentate manner. Unexpectedly, the overall binding mode was found to be very distinct from that of Molidustat to FIH (Figure 16). The central metal is coordinated by the nitrogens of the pyrazolone and pyridine for IOX4 and histidine-274 and histidine-199 of FIH. Surprisingly, aspartate-201 is not interacting with the central metal, changing the overall Zn coordination to a planar square coordination with two potential coordination sites free. Whether these can be occupied by water cannot be determined unambiguously at the given resolution of the structure. Further, if the observed change in coordination of the central metal is specific for zinc or representative for iron(II) remains to be determined.

The triazole group of Molidustat, which interacts with the 2OG C-5 binding residues lysine-314 and tyrosine-145 of FIH, does not show any interactions with the lysine-314 and tyrosin-145. Instead the triazole group pushes into a newly formed hydrophobic pocket, π-stacking with phenylalanine-207; this presumably cause as series of side and main-chain conformational changes in the active site affecting the structural integrity of the DSBH fold. The β-3 strand is “pushed out” ~1.5 Å which leads to a displacement of the iron-coordinating residue aspartate-201 by approximately the same distance. This leads to the loss of the coordination site of central metal. Large conformational changes can also be observed in the major sheet of the DSBH fold and adjacent beta sheets N-terminal to the DBSH core fold with up to ~6 Å shift. Some helices N-terminal of the β-1 shift which are not involved in crystal packing in the tetragonal crystal form are in contact

with other molecules in the crystal lattice of the orthorhombic crystal. Whether these conformational changes are due to crystal packing or ligand binding is unclear. Structural changes in the DSBH and adjacent domains have not been observed for any other crystal form of FIH in complex with small molecules or peptide substrates (see also chapter 4 and chapter 5).

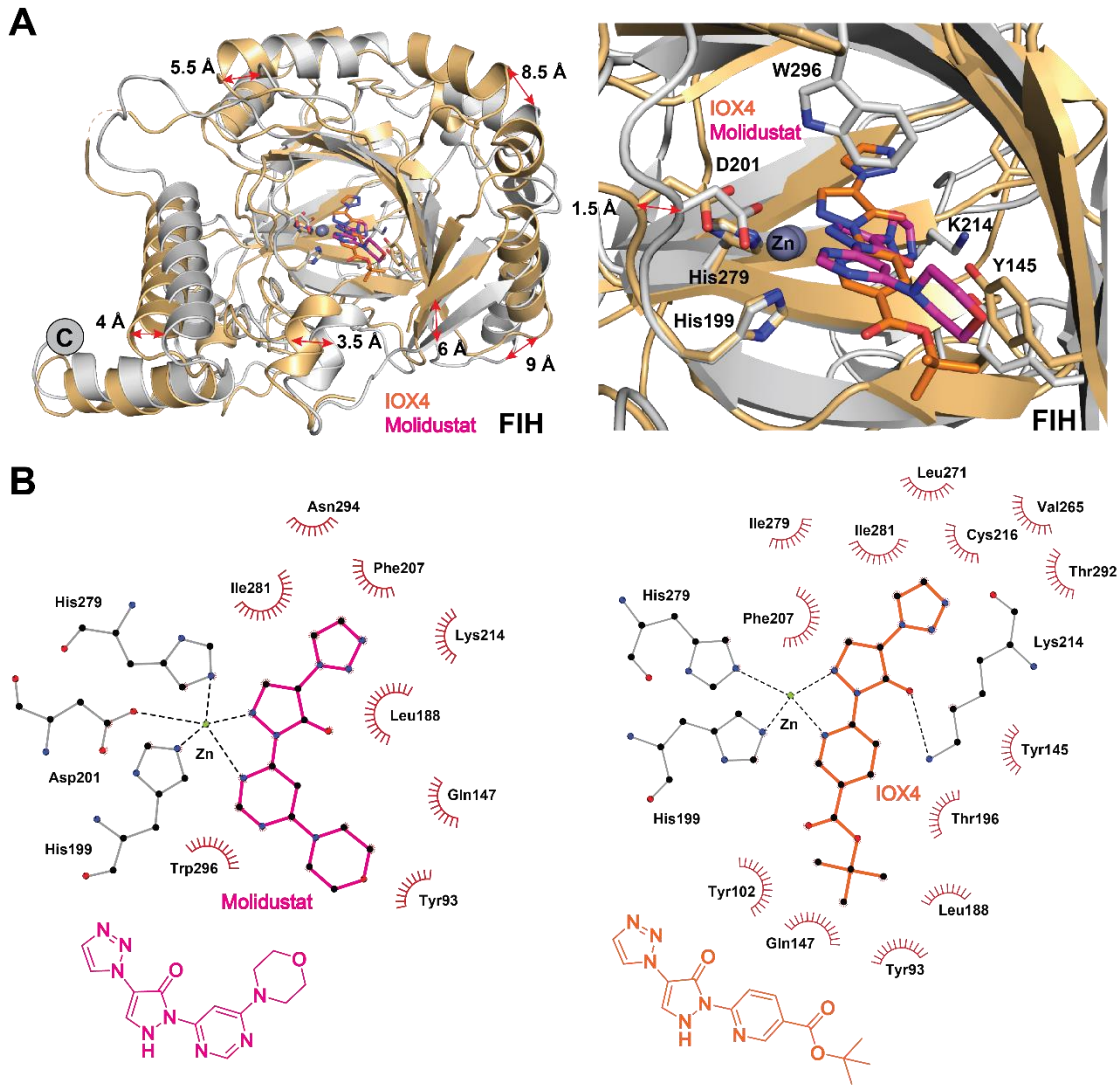


Figure 16 Comparison of Molidustat and IOX4 binding to FIH. A Views from a crystal structure of FIH in complex with IOX4 (light orange) overlaid with a crystals structure of FIH in complex with Molidustat (grey) illustrate conformation changes of the protein between the FIH.Zn.Molidustat and the FIH.Zn.IOX4 complex. **B** LIGPLOT[51] diagrams of active site interactions of Molidustat and IOX4 with the active site of FIH.

2OG C-5 binding residues lysine-314 and tyrosine-145 of FIH apparently do not interact with the triazole ring of IOX4 (Figure 15 and Figure 16). Lysine-314 forms electrostatic interactions with the oxygen of the pyrazolone, with tyrosine-145 also forming hydrophobic interactions with the *tert*.-butyl group and pyrimidine. Tryptophan-296,

located in the β8 strand of the DSBH, is displaced, potentially due to structural changes in the DSBH (C_α-atom of tryptophane-296 is moved ~4 Å, Figure 16).

The *tert.*-butyl ester of IOX4 is pointing in a hydrophobic pocket near the active site, which includes tyrosine-102, tyrosine-93, phenylalanine-100 and leucine-188 (Figure 16). Interaction with this hydrophobic pocket have been observed for other ligands as well, for example in the case of GSK1278863 binding to FIH (see chapter 2.2).

Table 5 Data collection and refinement statistics of an FIH.Zn(II).IOX4 complex.

Data Collection		**Refinement**	
Beamline (Wavelength, Å)	DLS I04 (0.9796)	R_{work}/R_{free}	0.2860/0.3093
Detector	PILATUS 6M-F	No. atoms	
Data Processing		-Enzyme (A/B/C/D)	9878
Space group	$P2_12_12_1$	-Ligand	96
Cell dimensions		-Water	82
a,*b*,*c* (Å)	66.9, 186.9, 187.8	Avg. B-factors	
α,β,γ(°)	90, 90, 90	Enzyme	77.6
No. of molecules/ASU	4	Ligand	48.3
No. reflections	65771 (2299)*	Water	61.5
Resolution (Å)	93.91-2.70 (2.80-2.70)*	R.m.s deviations	
R_{merge}	0.215 (1.738)*	-Bond lengths (Å)	0.004
I/σI	9.2 (1.6)*	-Bond angles (°)	0.821
CC ½	0.996 (0.523)*		
Completeness (%)	100 (100)*		
Multiplicity	11.9 (12.4)*		
Wilson B value ($Å^2$)			

* Highest resolution shell in parentheses.

†$R_{factor} = \Sigma_{hkl} | |F_{obs}(hkl)| - k|F_{calc}(hkl)| | / \Sigma_{hkl} |F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

Overall the observed binding mode of IOX4 differs significantly, at least in the crystal state, from the binding mode of Molidustat to FIH (Figure 16). This is surprising considering the high structural similarity of both compounds with nearly identical iron chelating motifs and 2OG binding pocket interaction sites. Interactions of IOX4 with FIH are mainly hydrophobic and it was previously observed that solution based data did not always correlate with the binding mode observed in the crystal state for some 2OG oxygenases[54]. Further investigations are needed to confirm the binding mode of IOX4 to FIH in solution.

2.4 FIH as a binding model for other 2OG oxygenases

The work performed on the structural characterisation of PHD inhibitors in clinical trials demonstrates the value of a model oxygenase for gaining structural information on otherwise non-crystallising samples. FIH can be crystallised reliably in its tetragonal form, which gives an open active site with a solvent content of >60%, allowing quick diffusion of small molecules and even peptide ligands. This may be of importance if the ligand is not stable in aqueous solution.

Ongoing studies on 2OG derivatives raise the question of whether 2OG analogous could be used to deliver halogen atoms to the active site of 2OG oxygenases. Carbonic acids halogenated in α-position are not stable in aqueous solution for prolonged periods of time which is challenging considering the long time of the FIH crystallisation process.

For this study, the binding mode of 2,2-dicholoroglutarate (DCG) was of particular interest, as it is stable in aqueous solution over several hours (Dr Martine Abboud, University of Oxford). To account for potential structural changes in the active site, DCG was co-crystallised with FIH. FIH crystallises within 24 h; therefore, it was hoped that crystallisation would be fast enough to avoid significant hydrolysis of the ligand.

Crystals were fished ~14 h after the crystallisation trials were set up. The data of the cryo-cooled crystals were collected immediately after fishing, and the structure solved (see Table 6). Electron density suggested that not DCG, but 2OG, was bound to the active site. DCG was used in excess, so only partial hydrolysis would give levels of 2OG high enough to bind to the active site of FIH which then could crystallise in complex with FIH.

To avoid ligand hydrolysis, handling time of dissolved ligand were reduced to the lowest time possible. Solid DCG was dissolved in FIH well solution supplemented with 20% glycerol and the ligand incubated for ~40 sec reducing the time of the ligand in solution to >1 min prior to freeze-cooling in liquid nitrogen. Data were collected within 72 h and the structure solved. Unlike with co-crystallised DCG with FIH, apparent electron density corresponding to the di-chloro group of DCG could be observed in the active site after soaking of the compound (Figure 17).

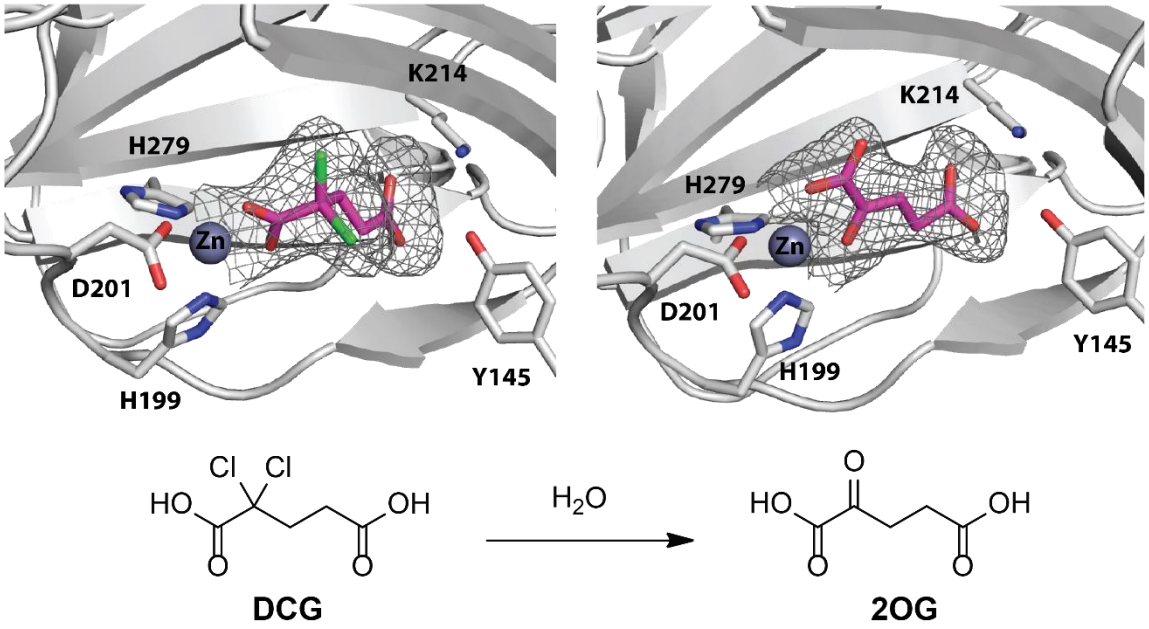


Figure 17 FIH in complex with 2OG and DCG. Views from crystal structures of FIH in complex with DCG (left) and 2OG (right), mFo-Fc OMIT maps contoured to 3σ around the ligands.

Notably, the apparent active site metal binding of DCG reassembles the binding of succinate to FIH with two coordination sites of the metal occupied by the two oxygens of the carboxylic acid (rather than that observed for 2OG binding, see also chapter 2.1). To achieve this coordination mode, DGC is very strained in the active site which suggests a less favourable binding mode compared to succinate or 2OG.

For crystallisation purposes, iron was substituted for zinc to avoid metal oxidation. As the structure gives a potential binding mode of DCG to the active site of FIH, effects of the nature of the central metal on ligand coordination and binding cannot be excluded. Further, whether the hydrolysis of DCG to 2OG occurred in solution or was facilitated by FIH is unclear.

Table 6 Data collection and refinement statistics of FIH complexes with 2OG and DCG.

Datasets	**FIH**.Zn(II).2OG co-crystallised (with DCG)	**FIH.**Zn(II).DCG soaked
Data Collection		
Beamline (Wavelength, Å)	DLS I02 (0.9795)	DLS I04 (0.9796)
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	XIA2 3dii (XDS)	XIA2 3d (XDS)
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	86.3, 86.3, 149.0	86.6, 86.6, 148.1
α,β,γ(°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. reflections	38210 (2758)*	27311 (1980)*
Resolution (Å)	56.48-2.01 (2.06-2.01)*	61.22-2.32 (2.38-2.32)*
R_{merge}	0.129 (1.695)*	0.099 (2.2662)*
I/σI	10.2 (1.4)*	21.3 (1.2)*
CC1/2	1.0 (0.700)*	1.0 (0.611)*
Completeness (%)	99.8 (100)*	100 (99.9)*
Multiplicity	12.4 (12.3)*	17.1 (17.5)*
Wilson B value (Å^2)	42.5	58.0
Refinement		
R_{work}/R_{free}†	0.2175/0.2393	0.2041/0.2346
No. atoms		
-Enzyme (A/B)	2669	2610
-Ligand	10	11
-Water	85	44
Avg. B-factors		
Enzyme	68.3	73.8
Ligand	44.9	81.8
Water	57.1	59.6
R.m.s deviations		
-Bond lengths (Å)	0.005	0.007
-Bond angles (°)	0.617	0.754

* Highest resolution shell in parentheses.

† $R_{factor} = \Sigma_{hkl} | |F_{obs}(hkl)| - k|F_{calc}(hkl)| | / \Sigma_{hkl} |F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

2.5 Conclusions and outlook

With only a very few exceptions, FIH crystallises reliably in space group $P4_12_12$ yielding generally well diffracting crystals. This makes FIH a good model 2OG oxygenase for structural studies.

Crystal structures of FIH in complex with Molidustat, GSK1278863 and Vadadustat were solved. All of the inhibitors were found to bind to the active site of FIH in a conserved manner. Notably, the active site metal coordination of the inhibitors bound to FIH re-assembled 2OG binding to 2OG oxygenases from the DAOCS subfamily rather than the JmjC subfamily of 2OG oxygenases. These observations imply greater flexibility within the active site of 2OG oxygenases than the crystal states might suggest. The observed binding mode is not only limited to the clinical candidates but has also been observed for a small series of other N-glycinamide containing inhibitors. Prediction of the binding mode of inhibitors remains challenging, as observed for the case of IOX4 which caused major conformational changes within the active site of FIH including distortion of the DSBH core fold. Whether this is an artefact from crystallography remains to be determined.

The use of Pan-Dataset Density Analysis has been shown to be very useful for defining multiple conformations of a single bound ligand, as demonstrated for the FIH.Vadadustat complex. However, even with the use of PANDDA maps, unambiguous ligand modelling was not possible for the FIH.FG-4592 complex. Although the exact binding mode of FG-4592 could not be determined, information of the binding site could be obtained with high confidence.

Using FIH as a model for otherwise non-accessible enzyme-inhibitor complex structures might be a very useful tool for inhibition studies of other 2OG oxygenases. In particular, FIH might be useful for the KDMs where the development of specific inhibitors is still challenged by a lack of structural data.

Using structural data of the novel inhibitor binding modes to FIH could guide further structure based drug design projects with the goal of highly selective, potent and cell permeable FIH inhibitors.

2.6 References

1 Tars, K. *et al.* Targeting carnitine biosynthesis: discovery of new inhibitors against γ-butyrobetaine hydroxylase. *J Med Chem* **57**, 2213-2236, doi:10.1021/jm401603e (2014).

2 Schobersberger, W., Dünnwald, T., Gmeiner, G. & Blank, C. Story behind meldonium-from pharmacology to performance enhancement: a narrative review. *Br J Sports Med* **51**, 22-25, doi:10.1136/bjsports-2016-096357 (2017).

3 Henry, L., Leung, I. K., Claridge, T. D. & Schofield, C. J. γ-Butyrobetaine hydroxylase catalyses a Stevens type rearrangement. *Bioorg Med Chem Lett* **22**, 4975-4978, doi:10.1016/j.bmcl.2012.06.024 (2012).

4 Morera, L., Lübbert, M. & Jung, M. Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. *Clin Epigenetics* **8**, 57, doi:10.1186/s13148-016-0223-4 (2016).

5 Wang, Z. & Patel, D. J. Small molecule epigenetic inhibitors targeted to histone lysine methyltransferases and demethylases. *Q Rev Biophys* **46**, 349-373, doi:10.1017/S0033583513000085 (2013).

6 Maes, T., Carceller, E., Salas, J., Ortega, A. & Buesa, C. Advances in the development of histone lysine demethylase inhibitors. *Curr Opin Pharmacol* **23**, 52-60, doi:10.1016/j.coph.2015.05.009 (2015).

7 Yeh, T.-L. *et al.* Molecular and cellular mechanisms of HIF prolyl hydroxylase inhibitors in clinical trials. *Chem. Sci.* **8**, 7651-7668, doi:10.1039/C7SC02103H (2017).

8 Chan, M. C., Holt-Martyn, J. P., Schofield, C. J. & Ratcliffe, P. J. Pharmacological targeting of the HIF hydroxylases - a new field in medicine development. *Mol Aspects Med*, doi:10.1016/j.mam.2016.01.001 (2016).

9 Chandra, M., Clemons, G. K. & McVicar, M. I. Relation of serum erythropoietin levels to renal excretory function: evidence for lowered set point for erythropoietin production in chronic renal failure. *J Pediatr* **113**, 1015-1021 (1988).

10 Walle, A. J., Wong, G. Y., Clemons, G. K., Garcia, J. F. & Niedermayer, W. Erythropoietin-hematocrit feedback circuit in the anemia of end-stage renal disease. *Kidney Int* **31**, 1205-1209 (1987).

11 Babitt, J. L. & Lin, H. Y. Mechanisms of anemia in CKD. *J Am Soc Nephrol* **23**, 1631-1634, doi:10.1681/ASN.2011111078 (2012).

12 Maxwell, P. H. & Eckardt, K. U. HIF prolyl hydroxylase inhibitors for the treatment of renal anaemia and beyond. *Nat Rev Nephrol* **12**, 157-168, doi:10.1038/nrneph.2015.193 (2016).

13 Rossert, J., Casadevall, N. & Eckardt, K. U. Anti-erythropoietin antibodies and pure red cell aplasia. *J Am Soc Nephrol* **15**, 398-406 (2004).

14 Macdougall, I. C. *et al.* Incidence of erythropoietin antibody-mediated pure red cell aplasia: the Prospective Immunogenicity Surveillance Registry (PRIMS). *Nephrol Dial Transplant* **30**, 451-460, doi:10.1093/ndt/gfu297 (2015).

15 Manns, B. J. & Tonelli, M. The new FDA labeling for ESA--implications for patients and providers. *Clin J Am Soc Nephrol* **7**, 348-353, doi:10.2215/CJN.09960911 (2012).

16 Chan, M. C. *et al.* Potent and Selective Triazole-Based Inhibitors of the Hypoxia-Inducible Factor Prolyl-Hydroxylases with Activity in the Murine Brain. *PLoS One* **10**, e0132004, doi:10.1371/journal.pone.0132004 (2015).

17 Feng, C. *et al.* Crystal structures of the human RNA demethylase Alkbh5 reveal basis for substrate recognition. *J Biol Chem* **289**, 11571-11583, doi:10.1074/jbc.M113.546168 (2014).

18 Chowdhury, R. *et al.* Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature*, doi:10.1038/nature13263 (2014).

19 Upadhyay, A. K. *et al.* An analog of BIX-01294 selectively inhibits a family of histone H3 lysine 9 Jumonji demethylases. *J Mol Biol* **416**, 319-327, doi:10.1016/j.jmb.2011.12.036 (2012).

20 Aik, W. *et al.* Structural basis for inhibition of the fat mass and obesity associated protein (FTO). *J Med Chem* **56**, 3680-3688, doi:10.1021/jm400193d (2013).

21 Rydzik, A. M. *et al.* Modulating carnitine levels by targeting its biosynthesis pathway - selective inhibition of γ-butyrobetaine hydroxylase. *Chem Sci* **5**, 1765-1771, doi:10.1039/C4SC00020J (2014).

22 Walport, L. J. *et al.* Human UTY(KDM6C) is a male-specific Nϵ-methyl lysyl demethylase. *J Biol Chem* **289**, 18302-18313, doi:10.1074/jbc.M114.555052 (2014).

23 Johansson, C. *et al.* Structural analysis of human KDM5B guides histone demethylase inhibitor development. *Nat Chem Biol* **12**, 539-545, doi:10.1038/nchembio.2087 (2016).

24 Becker, K. & Saad, M. A New Approach to the Management of Anemia in CKD Patients: A Review on Roxadustat. *Adv Ther* **34**, 848-853, doi:10.1007/s12325-017-0508-9 (2017).

25 Groenendaal-van de Meent, D. *et al.* Effect of Moderate Hepatic Impairment on the Pharmacokinetics and Pharmacodynamics of Roxadustat, an Oral Hypoxia-Inducible Factor Prolyl Hydroxylase Inhibitor. *Clin Drug Investig* **36**, 743-751, doi:10.1007/s40261-016-0422-y (2016).

26 Pergola, P. E., Spinowitz, B. S., Hartman, C. S., Maroni, B. J. & Haase, V. H. Vadadustat, a novel oral HIF stabilizer, provides effective anemia treatment in nondialysis-dependent chronic kidney disease. *Kidney Int* **90**, 1115-1122, doi:10.1016/j.kint.2016.07.019 (2016).

27 Johnson, B. M., Stier, B. A. & Caltabiano, S. Effect of food and gemfibrozil on the pharmacokinetics of the novel prolyl hydroxylase inhibitor GSK1278863. *Clin Pharmacol Drug Dev* **3**, 109-117, doi:10.1002/cpdd.83 (2014).

28 Thevis, M. *et al.* Mass spectrometric characterization of a prolyl hydroxylase inhibitor GSK1278863, its bishydroxylated metabolite, and its implementation into routine doping controls. *Drug Test Anal*, doi:10.1002/dta.1870 (2015).

29 Flamme, I. *et al.* Mimicking hypoxia to treat anemia: HIF-stabilizer BAY 85-3934 (Molidustat) stimulates erythropoietin production without hypertensive effects. *PLoS One* **9**, e111838, doi:10.1371/journal.pone.0111838 (2014).

30 Vogler, M. *et al.* Pre- and post-conditional inhibition of prolyl-4-hydroxylase domain enzymes protects the heart from an ischemic insult. *Pflugers Arch* **467**, 2141-2149, doi:10.1007/s00424-014-1667-z (2015).

31 Sen Banerjee, S. *et al.* HIF-prolyl hydroxylases and cardiovascular diseases. *Toxicol Mech Methods* **22**, 347-358, doi:10.3109/15376516.2012.673088 (2012).

32 Karuppagounder, S. S. *et al.* Therapeutic targeting of oxygen-sensing prolyl hydroxylases abrogates ATF4-dependent neuronal death and improves outcomes after brain hemorrhage in several rodent models. *Sci Transl Med* **8**, 328ra329, doi:10.1126/scitranslmed.aac6008 (2016).

33 Buisson, C. *et al.* Detection by LC-MS/MS of HIF stabilizer FG-4592 used as a new doping agent: Investigation on a positive case. *J Pharm Biomed Anal* **121**, 181-187, doi:10.1016/j.jpba.2016.01.029 (2016).

34 Eichner, D. *et al.* Implementation of the prolyl hydroxylase inhibitor Roxadustat (FG-4592) and its main metabolites into routine doping controls. *Drug Test Anal*, doi:10.1002/dta.2202 (2017).

35 Dib, J. *et al.* Mass spectrometric characterization of the hypoxia-inducible factor (HIF) stabilizer drug candidate BAY 85-3934 (molidustat) and its glucuronidated metabolite BAY-348, and their implementation into routine doping controls. *Drug Test Anal* **9**, 61-67, doi:10.1002/dta.2011 (2017).

36 Hansson, A. *et al.* Investigation of the metabolites of the HIF stabilizer FG-4592 (roxadustat) in five different in vitro models and in a human doping control sample using high resolution mass spectrometry. *J Pharm Biomed Anal* **134**, 228-236, doi:10.1016/j.jpba.2016.11.041 (2017).

37 Thevis, M. *et al.* Mass spectrometric characterization of a prolyl hydroxylase inhibitor GSK1278863, its bishydroxylated metabolite, and its implementation into routine doping controls. *Drug Test Anal* **8**, 858-863, doi:10.1002/dta.1870 (2016).
38 Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806, doi:10.1074/jbc.C200644200 (2003).
39 Chowdhury, R. *et al.* Structural basis for oxygen degradation domain selectivity of the HIF prolyl hydroxylases. *Nat Commun* **7**, 12673, doi:10.1038/ncomms12673 (2016).
40 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, doi:10.1101/gad.924501 (2001).
41 Yang, M. *et al.* Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidinyl residues within ankyrin repeat domains. *FEBS J* **278**, 1086-1097, doi:10.1111/j.1742-4658.2011.08022.x (2011).
42 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, doi:10.1074/jbc.M608337200 (2007).
43 Chowdhury, R. *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep* **12**, 463-469, doi:10.1038/embor.2011.43 (2011).
44 McDonough, M. A. *et al.* Selective inhibition of factor inhibiting hypoxia-inducible factor. *J Am Chem Soc* **127**, 7680-7681, doi:10.1021/ja050841b (2005).
45 Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a Broad Spectrum 2-Oxoglutarate Oxygenase Inhibitor which Causes Iron Translocation. *Chem Sci* **4**, 3110-3117, doi:10.1039/C3SC51122G (2013).
46 Moon, H., Han, S., Park, H. & Choe, J. Crystal structures of human FIH-1 in complex with quinol family inhibitors. *Mol Cells* **29**, 471-474, doi:10.1007/s10059-010-0058-3 (2010).
47 Conejo-Garcia, A. *et al.* Structural basis for binding of cyclic 2-oxoglutarate analogues to factor-inhibiting hypoxia-inducible factor. *Bioorg Med Chem Lett* **20**, 6125-6128, doi:10.1016/j.bmcl.2010.08.032 (2010).
48 Rose, N. R. *et al.* Plant growth regulator daminozide is a selective inhibitor of human KDM2/7 histone demethylases. *J Med Chem* **55**, 6639-6643, doi:10.1021/jm300677j (2012).
49 McDonough, M. A. *et al.* Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2). *Proc Natl Acad Sci U S A* **103**, 9814-9819, doi:10.1073/pnas.0601283103 (2006).
50 Pearce, N. M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat Commun* **8**, 15123, doi:10.1038/ncomms15123 (2017).
51 Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng* **8**, 127-134 (1995).
52 Aik, W. *et al.* Structure of human RNA N^6-methyladenine demethylase ALKBH5 provides insights into its mechanisms of nucleic acid recognition and demethylation. *Nucleic Acids Res* **42**, 4741-4754, doi:10.1093/nar/gku085 (2014).
53 Chowdhury, R. *et al.* Structural basis for binding of hypoxia-inducible factor to the oxygen-sensing prolyl hydroxylases. *Structure* **17**, 981-989, doi:10.1016/j.str.2009.06.002 (2009).
54 Chowdhury, R. *et al.* Selective small molecule probes for the hypoxia inducible factor (HIF) prolyl hydroxylases. *ACS Chem Biol* **8**, 1488-1496, doi:10.1021/cb400088q (2013).

3 High-throughput crystallographic fragment screening

This chapter describes work on the development and application of a high-throughput crystallographic fragment screen protocol for FIH using the XChem pipeline at the Diamond Light Source. 340 fragments were soaked into single crystals; the fragment binding was determined by identifying peaks in the Fo-$Fo_{average}$ density maps using the program PANDDA. Fragments observed to bind to FIH cluster in 4 distinct sites of which at least 3 overlap with the HIF-1α substrate binding site. The binding modes of fragments to FIH and potential follow-up compounds are discussed. Novel methodologies of in-crystal fragment reactions between boronic acids and hydrazines and thiol-amino compounds are described. Some of the compounds identified inhibit FIH activity and some appear to enhance its activity.

3.1 Introduction

3.1.1 Fragment based drug discovery

Fragment based drug discover (FBDD) has evolved into an efficient approach for initial hit discovery in drug design[1-3]. With technical advances in X-ray crystallography, *e.g*. faster detectors, robotic sample mounting, automated data processing[4] and hit detection[5] as well as improved sample preparation pipelines[6,7], crystallographic fragment screen has become a competitive approach for primary screening[8,9]. The advantage of a crystallographic screen over an enzyme inhibition or protein binding screen is that the binding modes of the hits are directly known. The major disadvantage of a crystallographic screen is that neither binding affinity nor the ability to inhibit the enzyme can be concluded from small molecule binding in the crystal state.

3.1.2 Crystallographic fragment screen

Studies on the molecular mechanism of clinically used PHD inhibitors revealed differences in their ability to displace the substrate from the enzyme[10]. However, all of the inhibitors tested compete with 2OG and therefore likely inhibit the turnover of all substrates indiscriminately. A substrate specific inhibitor is of great interest not only for the PHD enzymes but also for other 2OG oxygenases with multiple substrates, including FIH. FIH accepts next to the HIF transcription factors also a large number of ARD proteins (chapter 1) which were shown to be able to compete with HIF for FIH activity[11-13]. Therefore, blocking FIH-ankyrin interactions could “free” endogenous FIH for inhibition of HIF. The downregulation of FIH regulated HIF target gene expression, including CA9 and VEGF, is of potential pharmaceutical interest[14-16]. To achieve substrate specific inhibition, variations in the binding of different substrates at allosteric sites can be exploited. Variations in substrate binding have been shown for FIH between its HIF and ankyrin substrates (see chapter 4).

The objective of this chapter is to develop a crystallographic fragment screen for FIH to screen for small molecules binding to FIH.

3.2 Optimisation of fragment screen conditions for FIH

For crystallographic fragment screen, 100s to 1000s of uniform crystals, *i.e*. crystals with nearly identical cell parameters, are needed. Further, the crystals need to tolerate organic solvents used for soaking and the handling steps involved in the pipeline[9].

FIH crystallises reliably in high ammonium sulphate concentrations (1.1 M to 2.2 M) and low PEG400 concentrations (2.5-6%). The solvent content of an average tetragonal FIH crystals is >60% (see chapter 2). The high solvent content of the crystal and high salt concentration of the crystallisation buffer make FIH crystals very prone to dehydration. Many handling steps involved in the fragment soaking protocol can decrease crystal quality, which turned out to be a major obstacle for the soaking experiments. Initial soaking experiments using the XChem pipeline did not yield diffracting crystals. Degradation of the FIH protein crystals could be observed for all concentrations of DMSO and ethylene glycol (EE) tested (data not shown) showing that the crystals did not survive long enough for the handling steps (soaking, cryo-protection and fishing). Addition of extra well solution to the drops (one volume of the original drop volume) expanded the lifetime of the crystals long enough for the use in the pipeline (Table 1).

Table 1. Summary of solvent characterization using the XChem pipeline.

DMSO %	Diffraction [Å]	soaking time [min]	EE %	Diffraction [Å]	soaking time [min]
0	2.81	14:16	0	3.3	19:17
0	nd	14:46	0	2.68	20:29
5	2.67	15:08	10	2.21	21:42
5	-	-	10	-	-
10	-	-	20	2.31	22:18
10	-	-	20	2.24	22:59
20	nd	16:53	40	-	-
20	-	-	40	-	-
30	nd	17:50	60	nd	23:53
30	nd	18:43	60	-	-
0	2.85	38:38	0	2.71	44:13
0	2.82	39:21	0	-	-
5	3.17	40:32	10	2.39	45:04
5	2.67	41:39	10	-	-
10	-	-	20	2.80	44:13
10	2.73	42:31	20	2.31	45:40
20	-	-	40	-	-
20	-	-	40	-	-
30	-	-	60	-	-
30	-	-	60	-	-

Previous soaking experiments (see chapter 2.2.3) showed a low tolerance of FIH crystals for DMSO. This was also observed during the solvent tolerance test (Table 1). Crystals appeared to tolerate a final concentration of 2.5% DMSO.

Ethylene glycol appears to be tolerated by the FIH crystals to a considerably higher level than DMSO. 10-20% ethylene glycol appears to yield the highest diffracting crystals (Table 1). Further, ethylene glycol acts as cryo-protectant. Limited availability of libraries in ethylene glycol, however, challenge large scale screens.

Most fragment libraries are dissolved in DMSO; the LEEDS 3D library is also available in ethylene glycol. To reduce number of crystal manipulation steps, no further cryo-protectant was added in addition to the ethylene glycol or DMSO used in the fragment library. For low concentrations of DMSO and ethylene glycol, some ice rings were observed but found to be favourable over an additional cryo-protectant step. All volumes added are calculated on the initial volume of the drop. This does not take into account the loss of volume in the crystallisation drop during vapour diffusion as this could not be determined. The summary from the solvent tolerance test is shown in Table 1.

For fragment screens, a finial concentration of ethylene glycol of 10% and 2.5% of DMSO (assuming double of the original drop volume after addition of well solution) were chosen. Soaking times were set to <1 h at room temperature.

FIH crystals are weakly diffracting, thus the exposure time per image was increased from 0.04 sec (standard for fragment screening at DLS I04-1) to 0.08 sec for FIH data collection and the number of images collected reduced to 1,000 (corresponding to 120°) to decrease data collection time. An overview of the data collection strategies is shown in Table 2.

Table 2 Data collection strategies for FIH fragment screen.

Oscillation	0.12°	Exposure	0.08 sec
No. Images	1000	Transmission	100%
Resolution	1.9 Å	Wavelength	0.9159 Å

3.3 Fragment screen against FIH

For the fragment based screening against FIH, three different libraries were chosen: LEEDS 3D[17,18], OxXChem (unpublished) and DSLP[19], avoiding PAINS-like structures. The LEEDS 3D library consists of 125 compounds, the OxXChem library consists of 50 compounds and the DSPL library of 776 compounds. In total, 343 crystals were screened (125 compounds from LEEDS 3D library, 50 fragments from the OxXchem library and fragments 1-168 from the DSPL library). A summary of collected datasets and validation is shown in Figure 1.

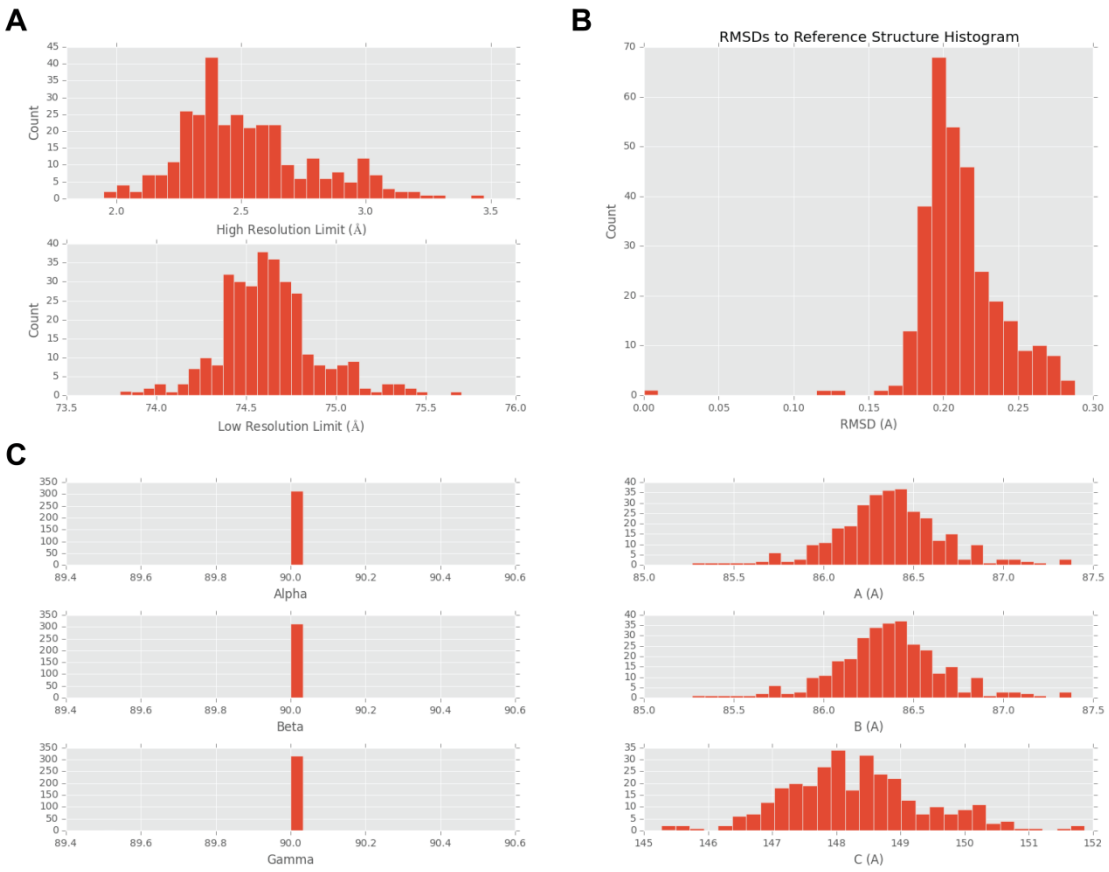


Figure 1 Dataset summary and validation. A Distribution of high and low resolution limits of FIH datasets collected. **B** Distribution of RMSD values of processed dataset compared to a reference structure. **C** Cell parameter distribution of processed FIH datasets.

Analysis of all datasets revealed the binding of 40 fragments from 32 conditions. An overview of identified binding sites is shown in Figure 2 and ligand validation scores are summarised in Table 3.

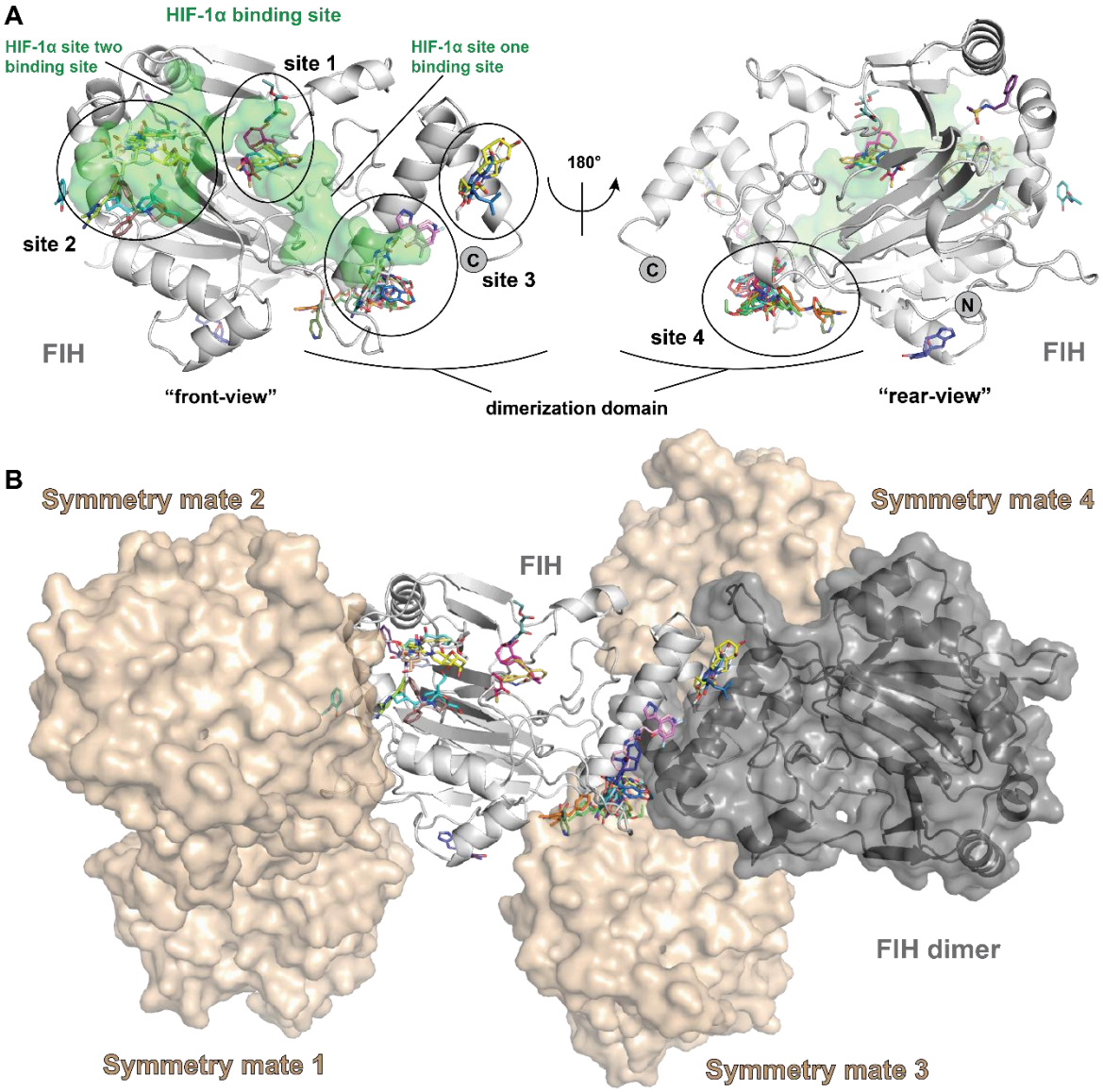


Figure 2 Overview of hits from the crystallographic Fragment screen with FIH. A Views from a crystal structure of FIH with bound fragments as assigned by PANDDA[5]. Note that due to crystallographic symmetry, fragments bound in site 3 are assigned in two binding sites. HIF-1α binding site is illustrated by surface representation (green) generated using PDB code 1H2K[20]. Circles indicate distinct binding sites for potential inhibitors to FIH. **B** Views from a crystal structure of FIH with bound ligands and symmetry related proteins illustrates accessible surfaces of FIH for fragment binding. The dimer partner is shown in grey; all other crystal contacts are shown by surface representation in light orange.

Most of the fragments cluster in 4 distinct sites. Notably, at least three of the fragment binding sites are in the substrate binding grove (also discussed in detail in chapter 1). However, ~40% of the hits bind in site 4. As site 1-3 are characterised for substrate binding, any physiological importance in terms of substrate binding of site 4 remains to be elucidated. Hardly any fragments binding to the "rear" side, when the "top" of the DSBH and substrate binding site is defined as "front".

Table 3. FIH ligand validation scores. Fragments are ordered according to their corresponding FIH crystal ID.

Dataset	Site	resolution (Å)	Occupancy	RSCC	RSZD	B ratio	rmsd (Å)
x0069	3	2.02	1.00	0.65*	0	2.67	0.41
x0075	2	2.37	0.62	0.84	0	1.29	2.64*
x0077	1	2.25	0.72	0.87	0	1.91	2.2*
x0078	2	2.37	0.84	0.91	0	1.73	0.81
x0078	3	2.37	1.00	0.88	0	1.56	2.14*
x0101	4	2.62	0.92	0.89	0	1.37	4.46*
x0107	2	2.25	0.86	0.94	0	1.08	0.25
x0107	3	2.25	0.70	0.90	0	1.30	0.65
x0107	1	2.25	0.64	0.88	0	1.10	1.48*
x0108	n.d.	2.52	1.00	0.90	0	2.58	0.22
x0130	4	2.36	0.66	0.90	0	1.04	1.29*
x0133	4	2.18	0.85	0.72	0	1.52	1.12*
x0180	4	2.02	0.98	0.94	2.4	1.06	1.47*
x0182	1	2.56	0.42	0.90	0	1.14	0.53
x0184	1	2.98	0.70	0.89	0	1.28	4.55*
x0203	n.d.	2.47	0.20	0.9	0.1	0.47	0.51
x0203	4	2.47	0.89	0.89	0	1.70	0.53
x0206	4	2.26	0.63	0.86	0	1.17	0.29
x0206	n.d.	2.26	0.92	0.90	0	0.93	0.55
x0206	1	2.26	0.63	0.82	2.3	0.98	2.17*
x0218	4	2.33	0.71	0.90	0	1.61	0.28
x0222	4	2.15	0.94	0.90	1.7	1.53	6.85*
x0233	n.d.	2.57	0.98	0.82	0	1.65	0.47
x0239	4	2.44	0.77	0.90	0.1	1.55	0.83
x0243	2	2.41	0.98	0.93	0	1.51	5.17*
x0249	2	2.41	0.80	0.91	0	1.49	0.18
x0256	4	2.49	0.84	0.93	0	1.77	0.47
x0275	3	2.26	0.98	0.92	0	1.46	0.88
x0277	4	2.29	0.81	0.92	0.4	1.66	0.97
x0277	2	2.29	0.81	0.92	0.4	1.66	0.97
x0281	2	2.33	0.90	0.88	0.8	1.75	0.96
x0281	4	2.33	0.78	0.86	2.2	1.67	1.96*
x0314	4	2.14	0.91	0.92	2.6	1.65	1.21*
x0315	3	2.30	0.98	0.74	0	1.79	1.73*
x0316	4	2.27	0.83	0.91	0.4	1.30	5.88*
x0322	4	2.23	0.62	0.89	0	1.05	1.66*
x0330	3	2.26	0.82	0.84	0	1.5	2.71*
x0361	2	2.52	0.68	0.84	0	1.59	0.54
x0378	4	2.77	0.82	0.92	0.3	1.67	0.41

*Values exceed the suggested values for “good hits” stated by Pearce et al[5]: RSCC>0.7, RSZD <3, B-factor ratio ~1, RMSD<1.

3.3.1 Fragment screen against FIH: Fragments binding to site 1

In total, 5 fragments were found to bind in the active site or close to the active site of FIH (x0077, x0107, x0182, x0184 and x0206, see also Figure 3). Note that fragment x0107 binds also in site 3, ligand x0206 binds in site 4 and one further unique site (Figure 3).

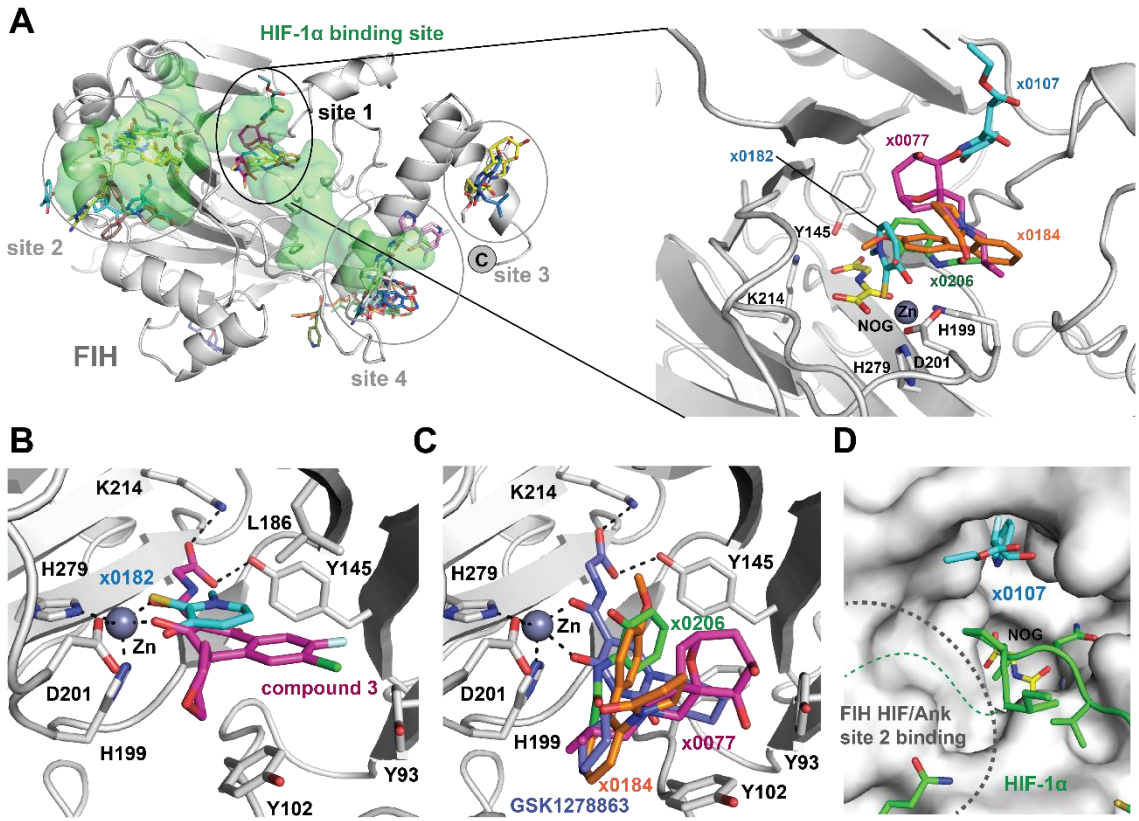


Figure 3 Overview of fragments binding to site 1 of FIH. A Views from a crystal structure of FIH in complex with fragments soaked in to the crystal lattice highlight the specific binding in the active site. **B** View from a crystal structure of FIH in complex with compound 3 (chapter 2) overlayed with x0182 illustrates conserved metal coordination of the two compounds. **C** View from a crystal structure of FIH in complex with GSK1278863 (chapter 2) overlayed with structures of FIH in complex with x0206, x0077 and x0184. **D** View from a crystal structure of FIH in complex with x0107 overlayed with the HIF-1α binding generated using PDB code 1H2K[20].

Out of all compounds screened, only x0182 displaced 2OG (Figure 3A), coordinating the central metal with its hydroxyl-group and its sulphur (Figure 3B). Like previously observed in the case of the clinical PHD inhibitors (see chapter 2.2) and other N-glycinamide test compounds (see chapter 2.3), x0182 prefers the coordination of the active site metal *trans* to histidine-279 and aspartate-201, and not *trans* to histidine-199 and aspartate-201 like 2OG or NOG (see chapter 2). The potency of x0182 as FIH inhibitor is described below. Further modifications could include the extension of the N-methyl group into the 2OG C-5 binding pocket and further out of the active site into the hydrophobic binding pocket build up out of tyrosin-102, tyrosine-93 and leucine-186 occupied by *e.g*. compound 3 or GSK1278863 (Figure 3B).

Compounds x0077 and x0184 bind in the hydrophobic pocket surrounding tyrosine-102, tyrosine-93 and leucine-186 (Figure 3C). Interactions of x0077 and x0184 appear to be mainly hydrophobic. Fragment x0184 and x0206 binds to the active site of FIH occupying the biding pocket of HIF-1α asparagine-803 resembling, at least to some degree, the binding of the FIH target residue. Compound x0206 further interacts with the central metal and the HxD/E…H motif.

Fragment x0107 shows an unprecedented binding mode (Figure 3D), *i.e*. at a site different to those observed in previous structures of FIH in complex with small molecules (see chapter 2). Comparison with the HIF-1α binding site reveals that x0107 does not overlay with the substrate binding site (Figure 3D). Interestingly, x0107 points into the substrate binding site two of FIH. This is the site where most differences in binding between different substrates were observed (see chapter 4).

Detailed protein-ligand binding analysis and structures of the hits are show in Figure 4. High RMSD values are often observed for FIH crystals and appear to be an intrinsic feature of this crystal form. Therefore, fragments with high RMSD values were taken into the analysis if the other validation parameters and electron density maps were acceptable (Figure 4).

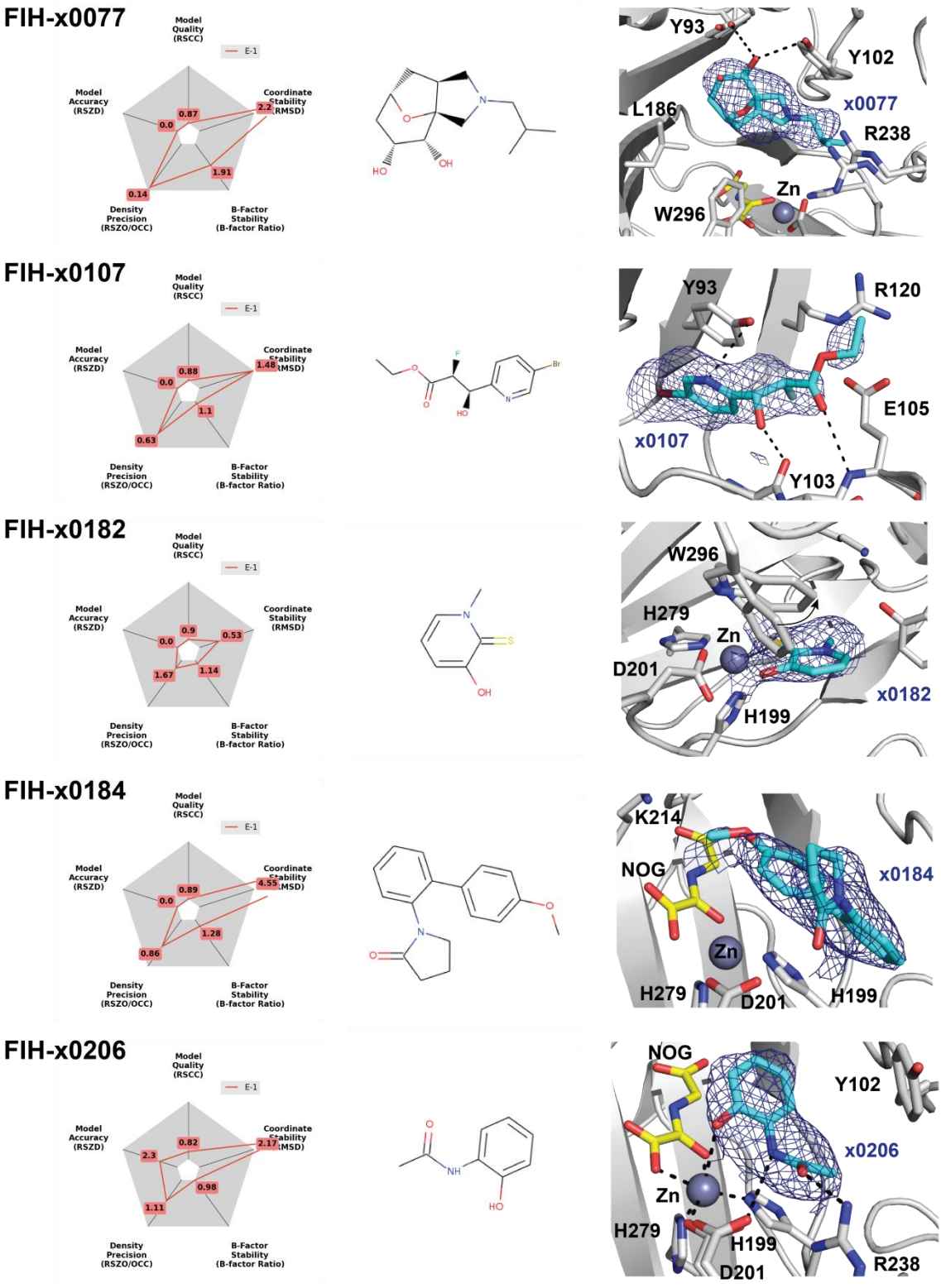


Figure 4 Analysis of fragments binding to the active site of FIH (site 1). Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows.

3.3.2 Fragment screen against FIH: Fragments binding to site 2

8 fragments were found to bind to site 2 of FIH and appear to overlap with the binding of HIF-1α binding to FIH (Figure 5). Binding in this site is of interest as differences in the binding mode of different substrates to FIH have been observed at this site (differences between HIF-α and ankyrin binding modes are discussed in detail in chapter 4). To date, no small molecules interacting with FIH in this site have been reported.

The binding site of the inhibitors can roughly be separated according to their overlap with the FIH-HIF-1α binding site (Figure 5A). Compounds x0078, x0249 and x0361 overlap in their binding site with L818 and L822 of HIF-1α on FIH, compounds x0177, x0277 and x0281 overlap with the binding of L813 and L818 of HIF-1α to FIH and compounds x0075 and x0243 overlap with the binding site of L819 of HIF-1α on FIH (Figure 5). Notably, x0277 and x0281 have different binding modes despite differing in only one stereocenter.

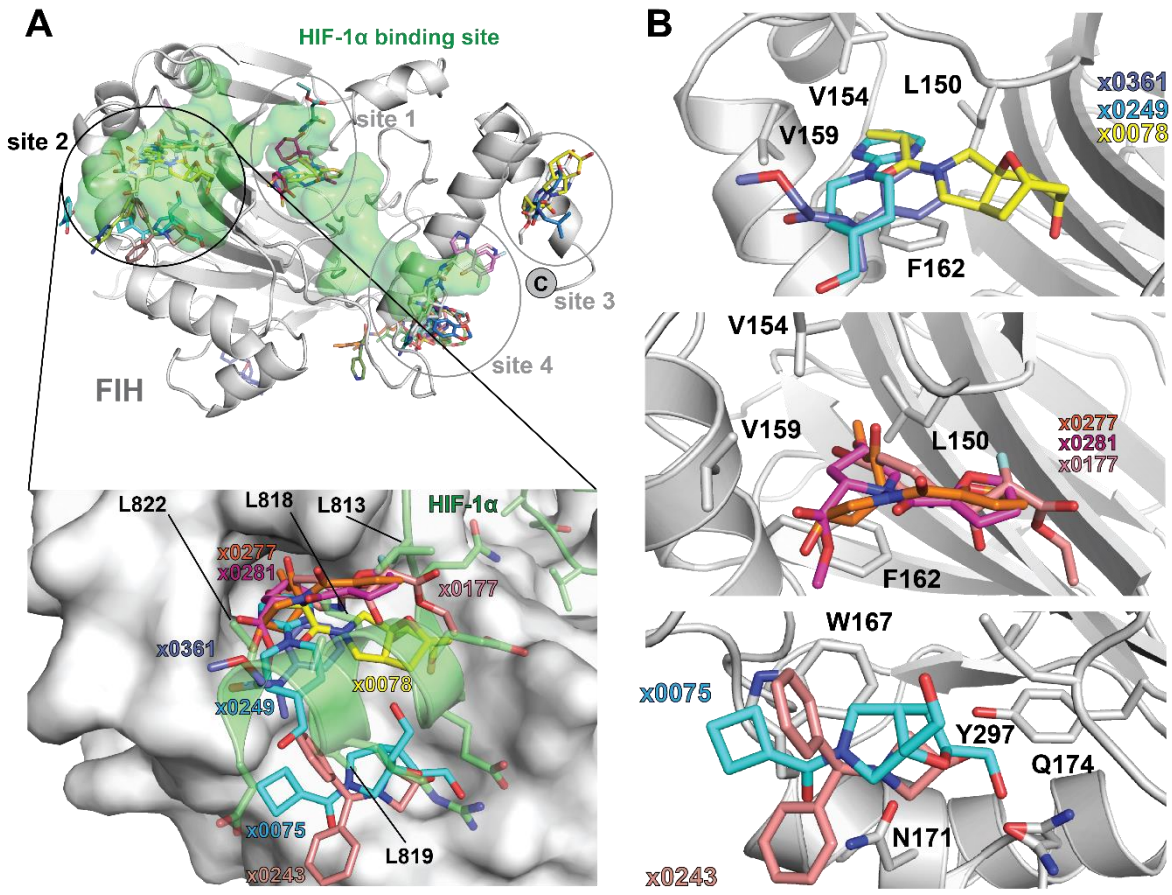


Figure 5 Overview of fragments binding to site 2 of FIH. A Views from a crystal structure of FIH in complex with fragments soaked in to the crystal overlayed with the HIF-1α binding generated from PDB code 1H2K[21] **B** View from a crystal structure of FIH in complex with **top:** fragments binding in the HIF-1α L818/L822 binding pocket; **middle:** fragments binding overlapping with the binding sites of HIF-1α L818 and L813; **bottom:** fragments binding in the HIF-1α L819 binding site.

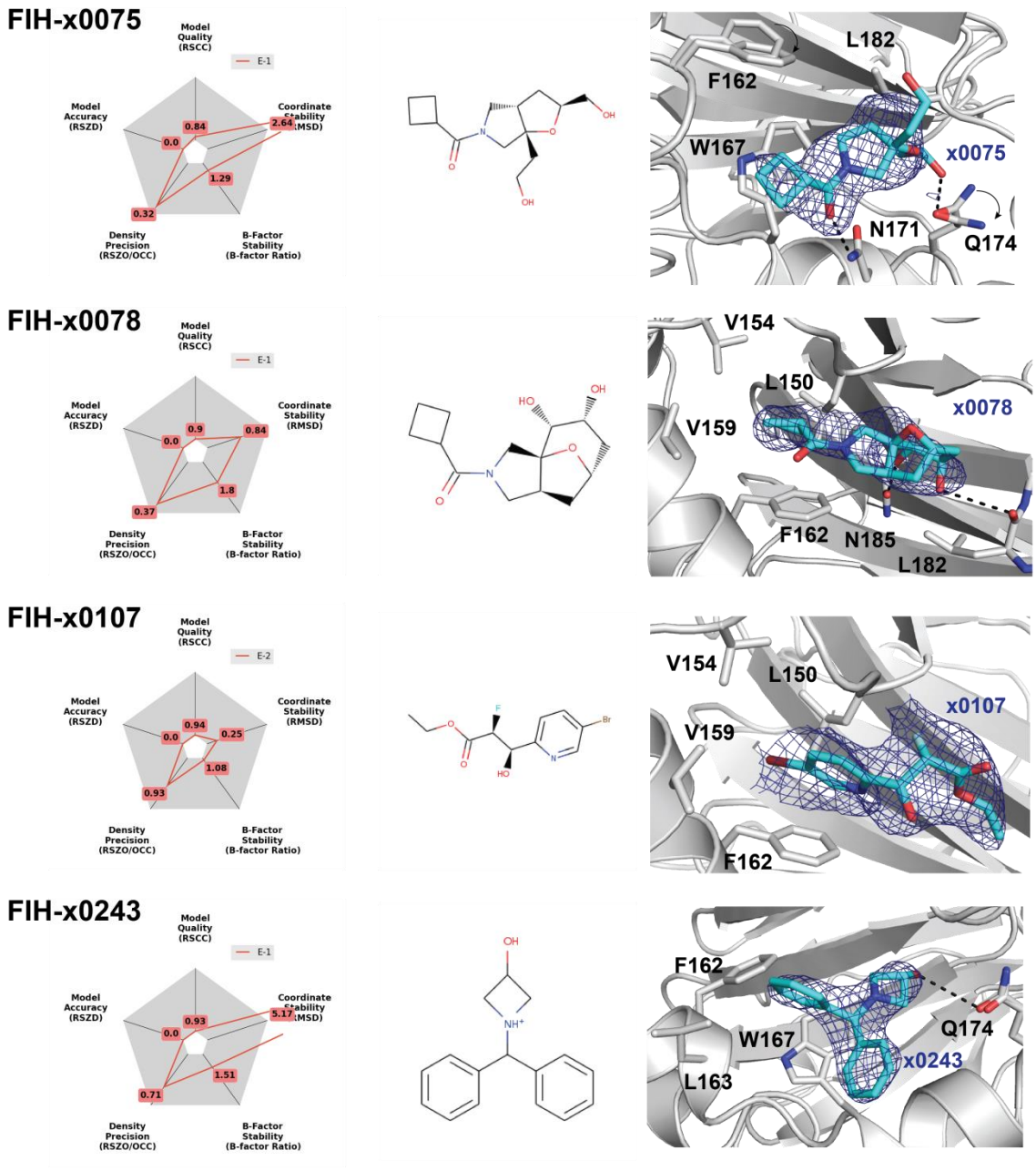


Figure 6 Analysis of fragments binding to site 2 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows.

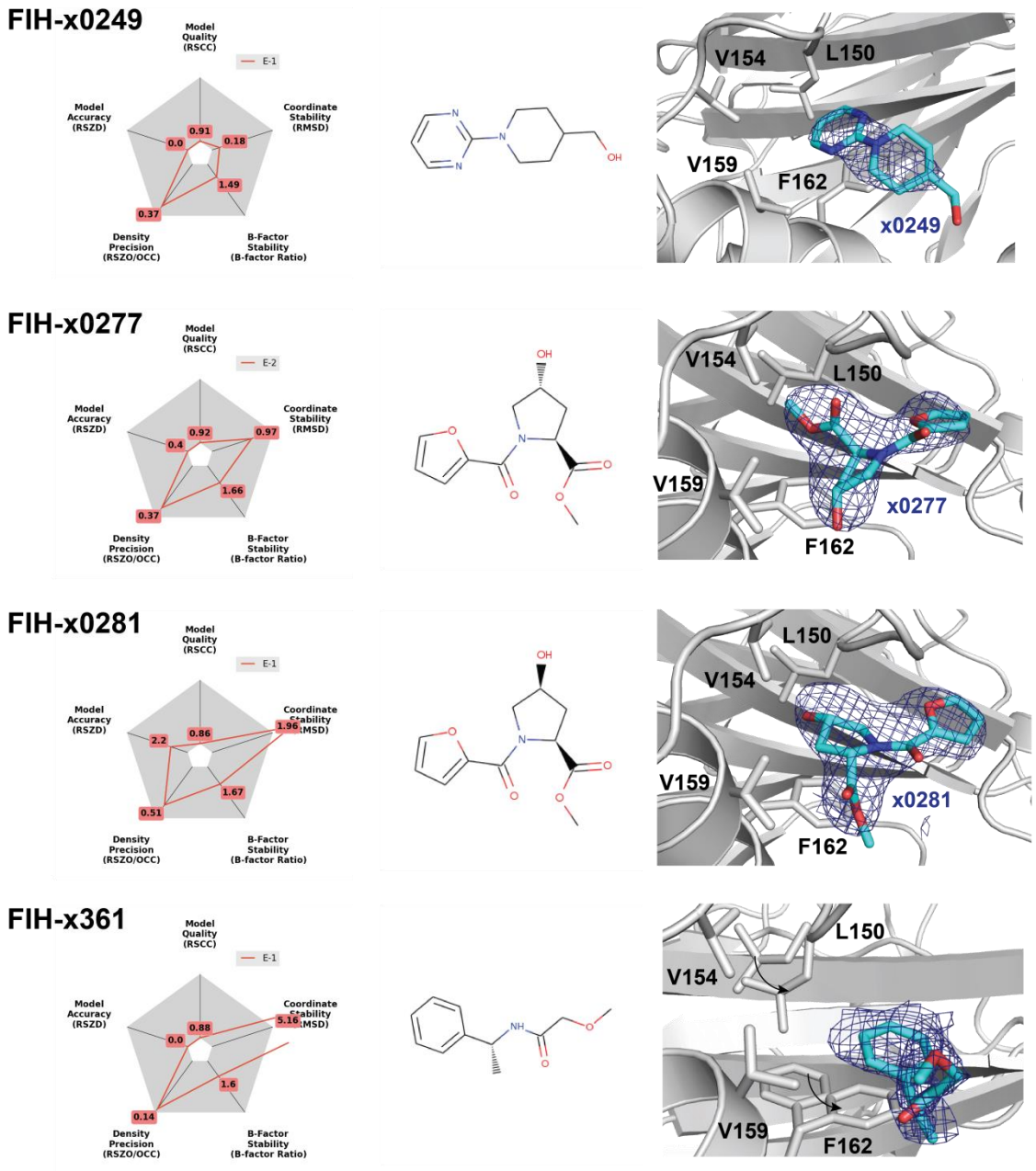


Figure 6 (continued) Analysis of fragments binding to site 2 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with an arrow.

3.3.3 Fragment screen against FIH: Fragments binding to site 3

6 fragments were found to bind to the dimerisation domain of FIH in the binding site of HIF-1α leucine-795[20] (Figure 7). This site, referred to as the L^{-8} binding site, is crucial for substrate binding (chapter 1). The leucine in -8 position relative to the hydroxylated residue is conserved in all known FIH substrates (see also chapter 1). Compounds binding to this site of FIH could potentially compete with substrate binding in an allosteric manner. This could be used for the development of an allosteric inhibitors. Whether substrate selectivity can be achieved by small molecules binding to the L^{-8} binding pocket remains to be tested.

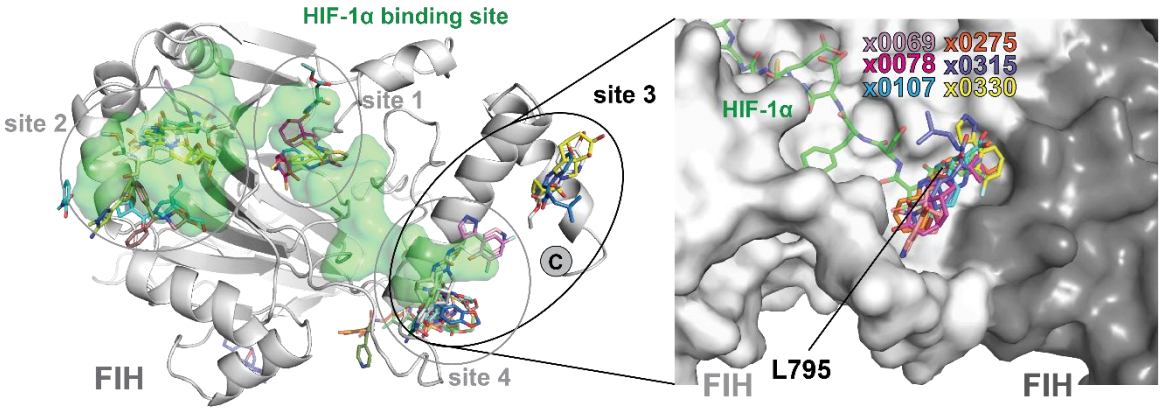


Figure 7 Overview of fragments binding to site 3 of FIH. Left: Views from a crystal structure of FIH in complex with fragments soaked in to the crystal overlayed with the HIF-1α binding generated from PDB code 1H2K[20] **Right:** View from a crystal structure of FIH in complex with fragments binding to site 3 of FIH. Note that compounds from the symmetry related proteins have been overlayed with the fragments assigned by PANDDA in the L^{-8} binding pocket. FIH is coloured in light grey and its dimer dark grey.

It is noteworthy that, due to the proximity to the crystallographic symmetry axes, all compounds binding in site 3 are potentially close to each other. Fragments x0275, x0330 and x0107 are in close proximity to their symmetry related binding partners, potentially interacting with themselves and the protein. Detailed protein-fragment interactions are shown in Figure 8.

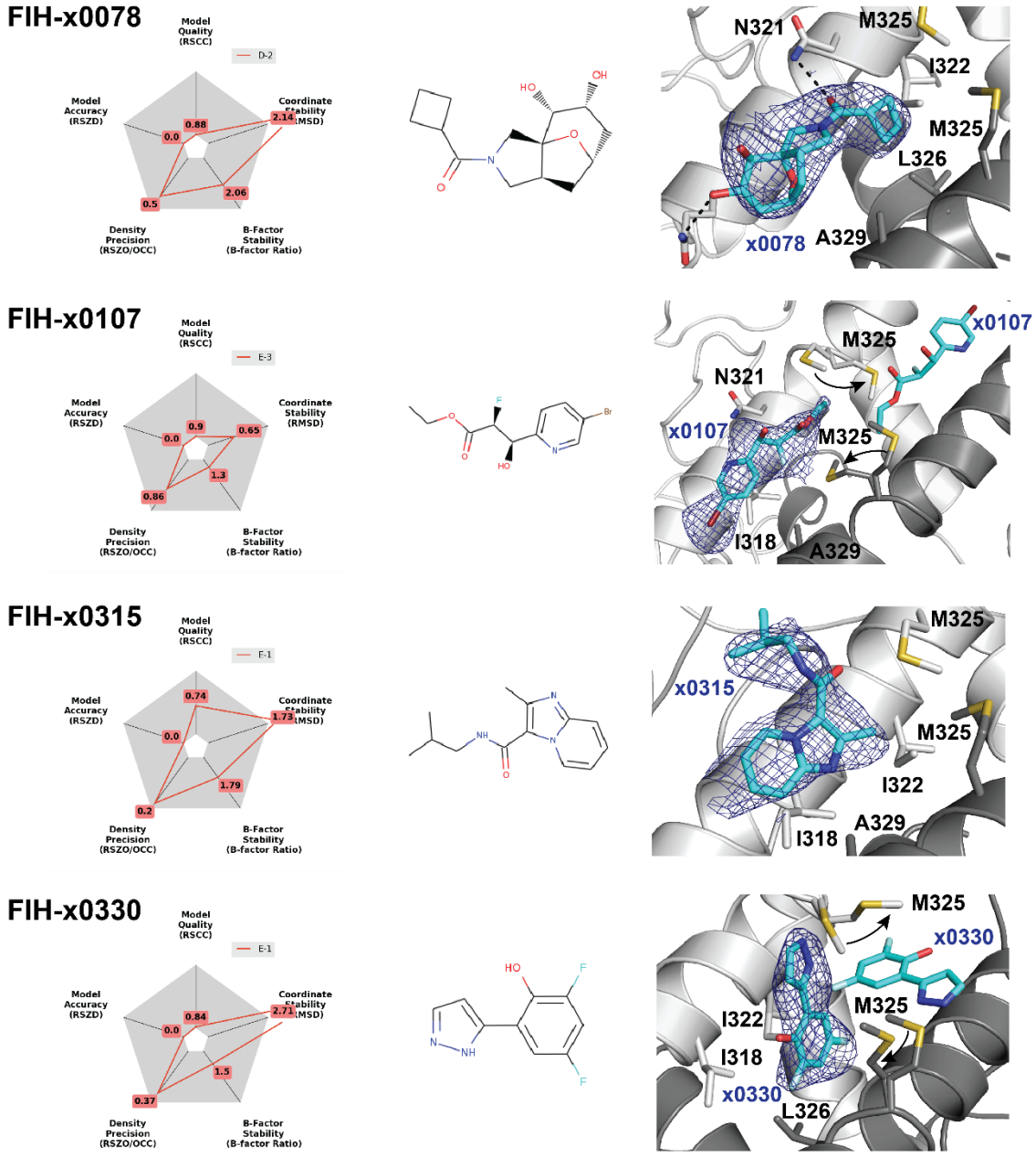


Figure 8 Analysis of fragments binding to site 3 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows. FIH is coloured in light grey and its dimer dark grey.

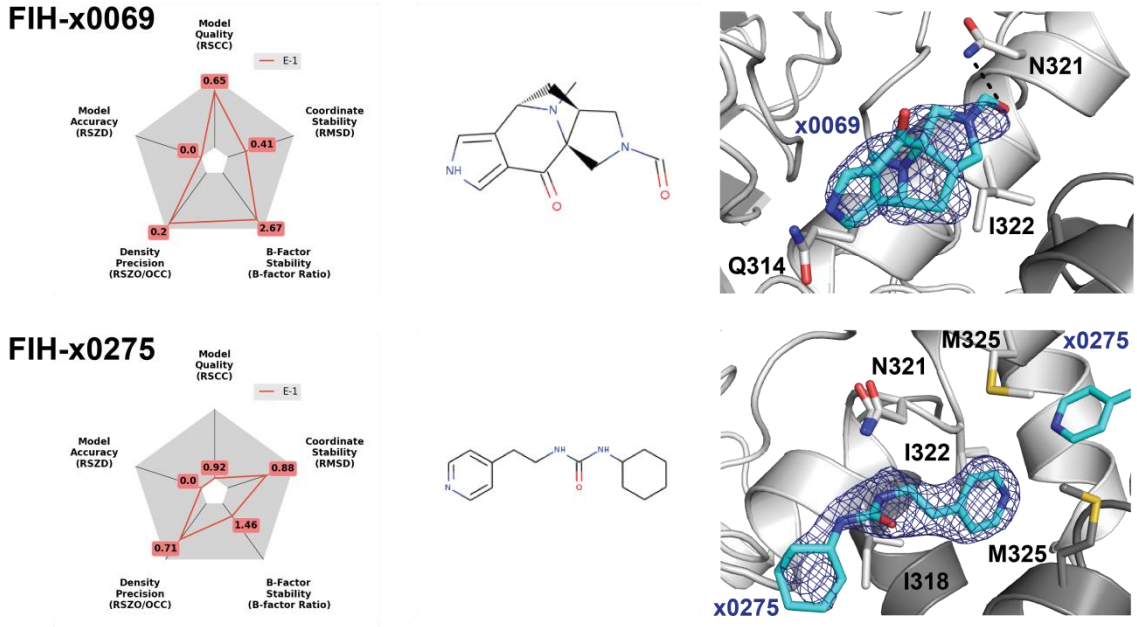


Figure 8 (continued) Analysis of fragments binding to site 3 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows. FIH is coloured in light grey and its dimer dark grey.

3.3.4 Fragment screen against FIH: Fragments binding to site 4

In addition to sites 1-3, which are all part of the substrate binding site, another apparently preferred binding site (site 4), for fragments to FIH was observed. 15 fragments were found to bind to site 4 of FIH.

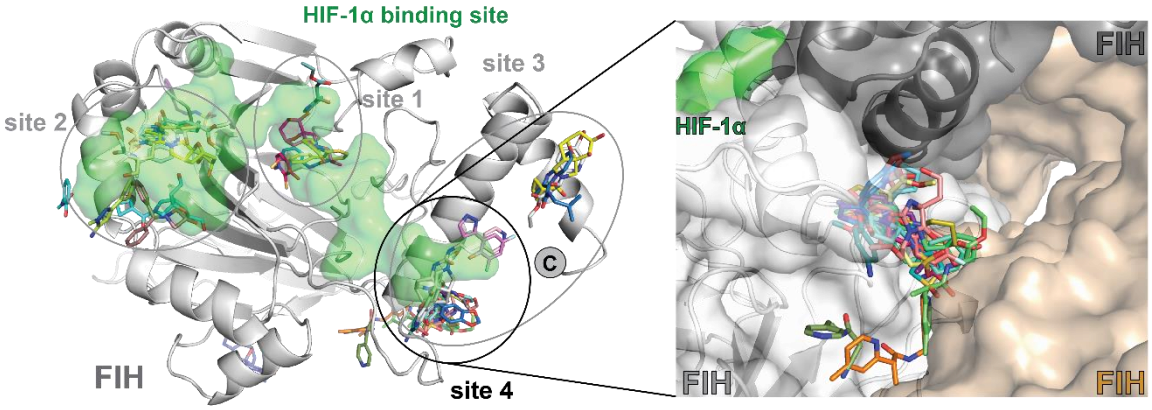


Figure 9 Overview of fragments binding to site 4 of FIH. Left: Views from a crystal structure of FIH in complex with fragments soaked in to the crystal overlayed with the HIF-1α binding generated from PDB code 1H2K[20] **Right:** View from a crystal structure of FIH in complex with fragments binding to site 4 of FIH. FIH is coloured in light grey and its dimer dark grey and further symmetry mates in the crystal packing are coloured in light orange.

Up to date, peptide binding to the active site of FIH has only been observed in the crystalline state until the L^{-8} binding site (see chapter 1). Studies with longer HIF substrates describe the importance of an up-stream "RLL" motif on HIF-1α for FIH mediated hydroxylation of HIF proteins[22] which could indicate that further protein-substrate interactions could exist. However, these interactions are not essential for FIH substrate binding, thus the chances of an allosteric inhibitor binding at site 4 are limited.

Fragments x0180, x0203, x0206, x0218, x0222, x0239, x0256 interact with only one protein of FIH or with one protein of FIH and its corresponding dimer. Fragments x0101, x0130, x0133, x277, x0281, x0314, x0316, x0378 bind to FIH at the interfaces with other FIH enzymes in the crystal lattice (Figure 10), which makes them more likely to be crystallographic artefacts.

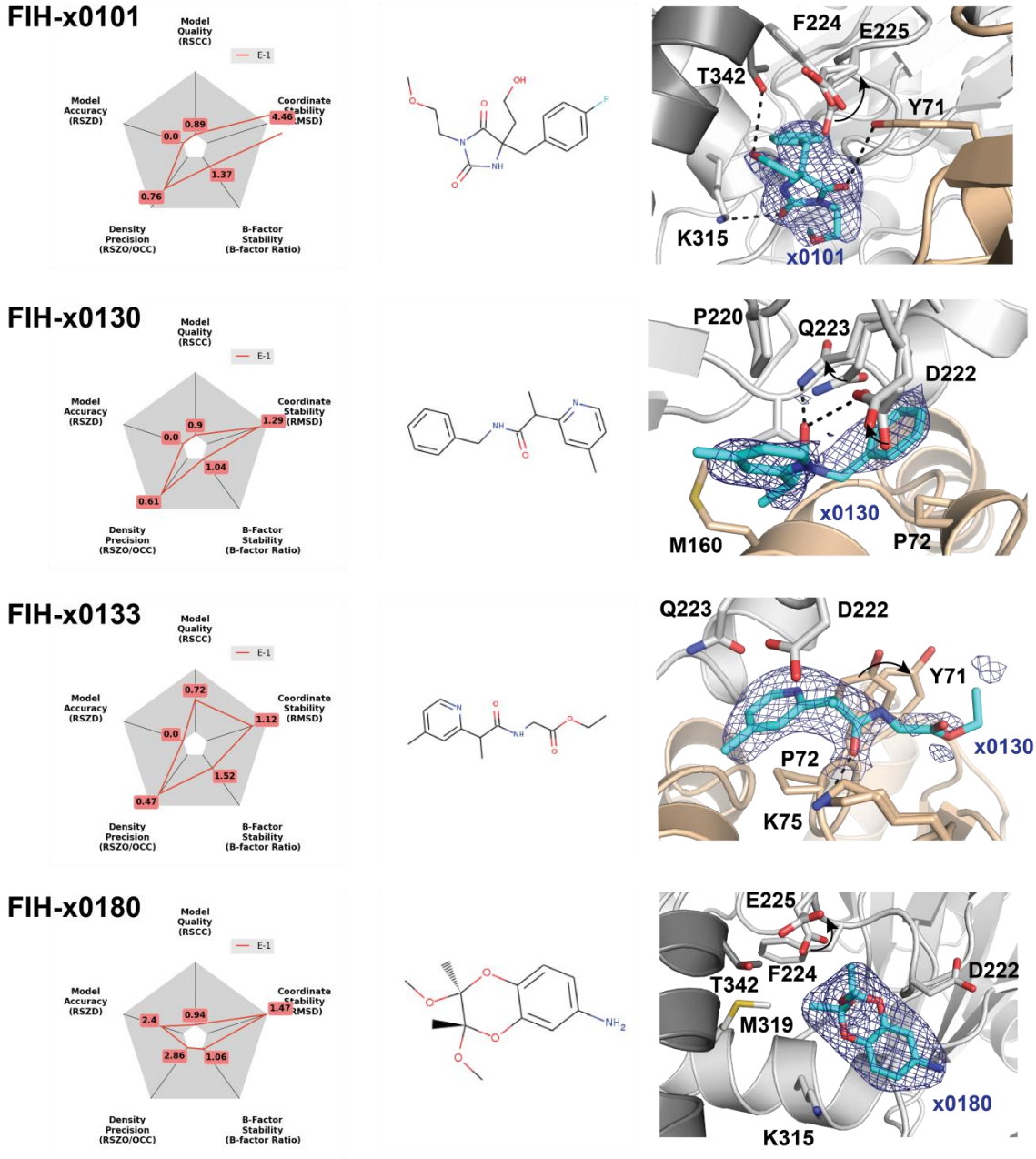


Figure 10 Analysis of fragments binding to site 4 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows. FIH is coloured in light grey and its dimer dark grey and further symmetry mates in the crystal packing are coloured in light orange.

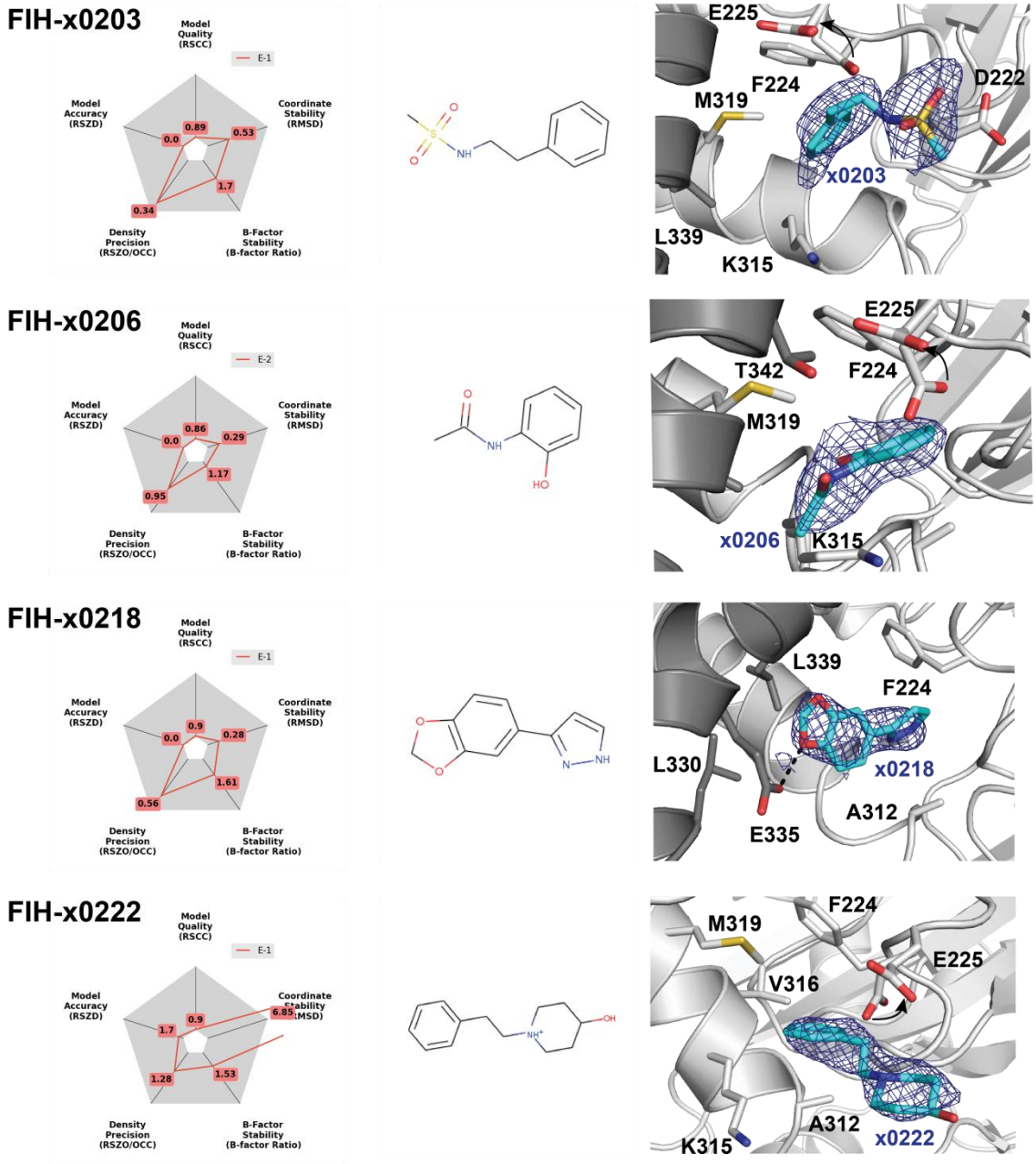


Figure 10 (continued) Analysis of fragments binding to site 4 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows. FIH is coloured in light grey and its dimer dark grey and further symmetry mates in the crystal packing are coloured in light orange.

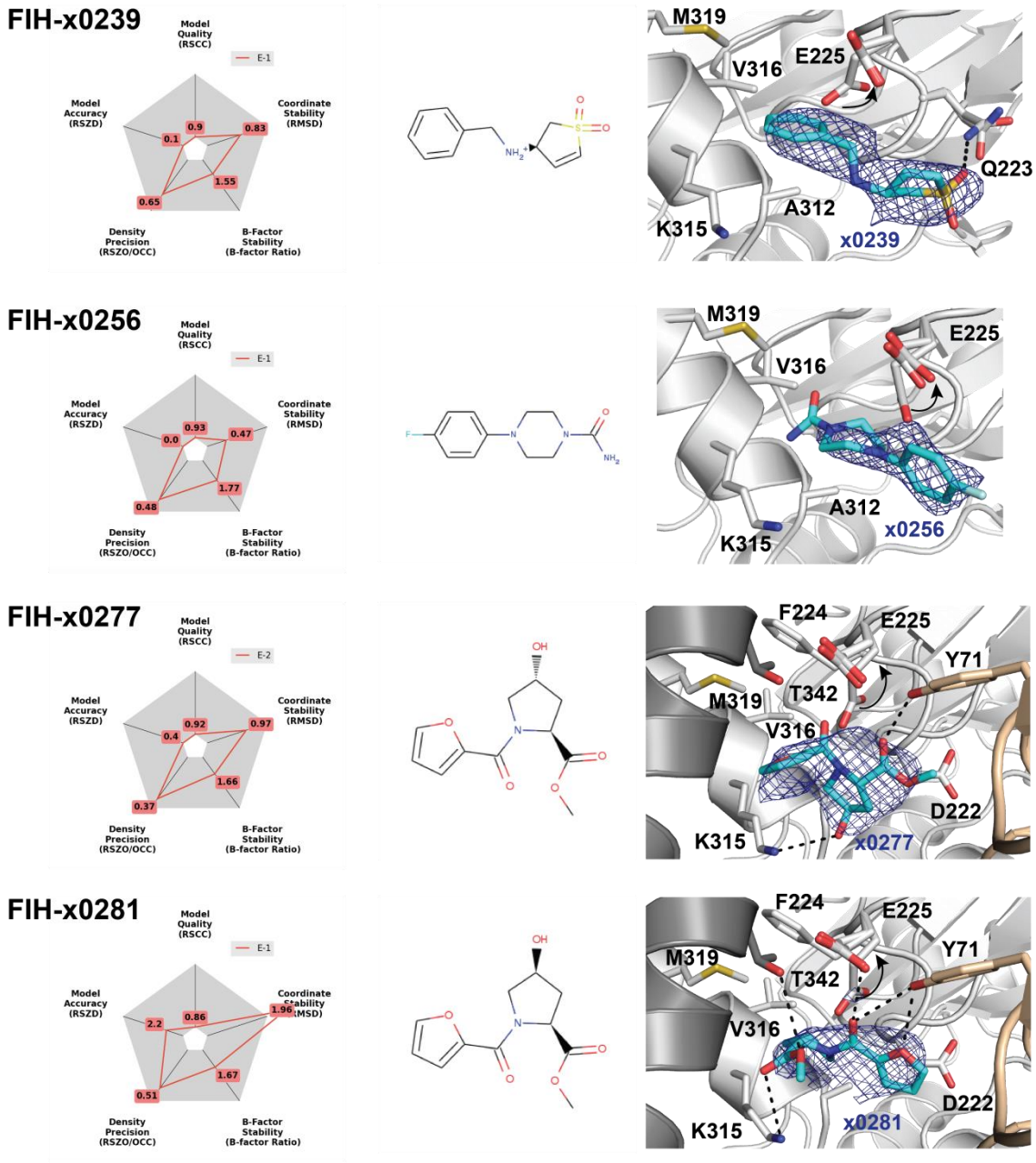


Figure 10 (continued) Analysis of fragments binding to site 4 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows. FIH is coloured in light grey and its dimer dark grey and further symmetry mates in the crystal packing are coloured in light orange.

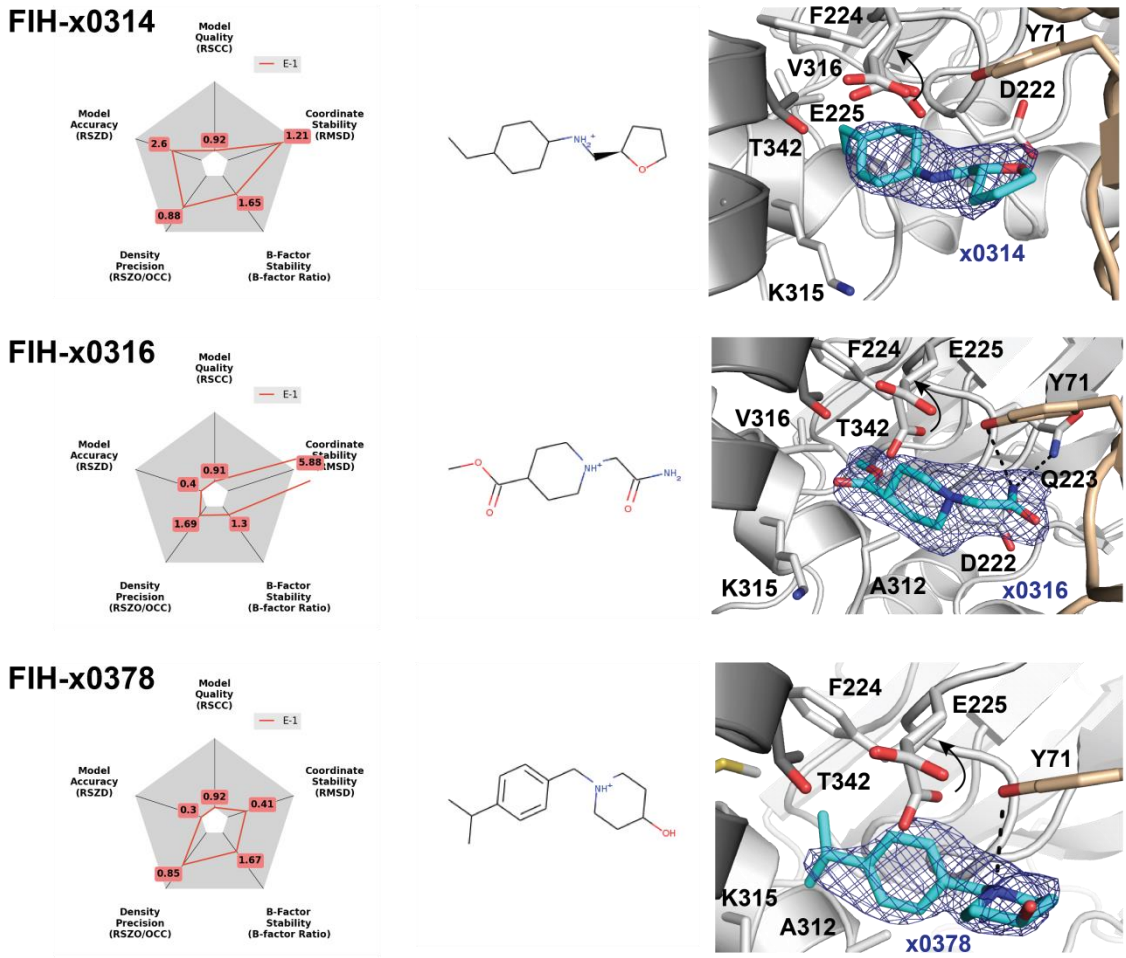


Figure 10 (continued) Analysis of fragments binding to site 4 of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows. FIH is coloured in light grey and its dimer dark grey and further symmetry mates in the crystal packing are coloured in light orange. Note that the dataset FIH-x0314 is a low resolution dataset (>3 Å). The orientation of the ligand could not be unambiguously determined.

3.3.5 Fragment screen against FIH: Fragments binding further binding sites of FIH

In addition to the 4 observed "hot-spots" for fragment biding to FIH, four fragments were found to bind to surface of FIH at other sites. Fragments x0108, x0203 and x0233 bind to the surface of FIH, whereas x0206 binds at the crystal packing interface. Fragment-protein interactions and binding statistics are shown in Figure 11.

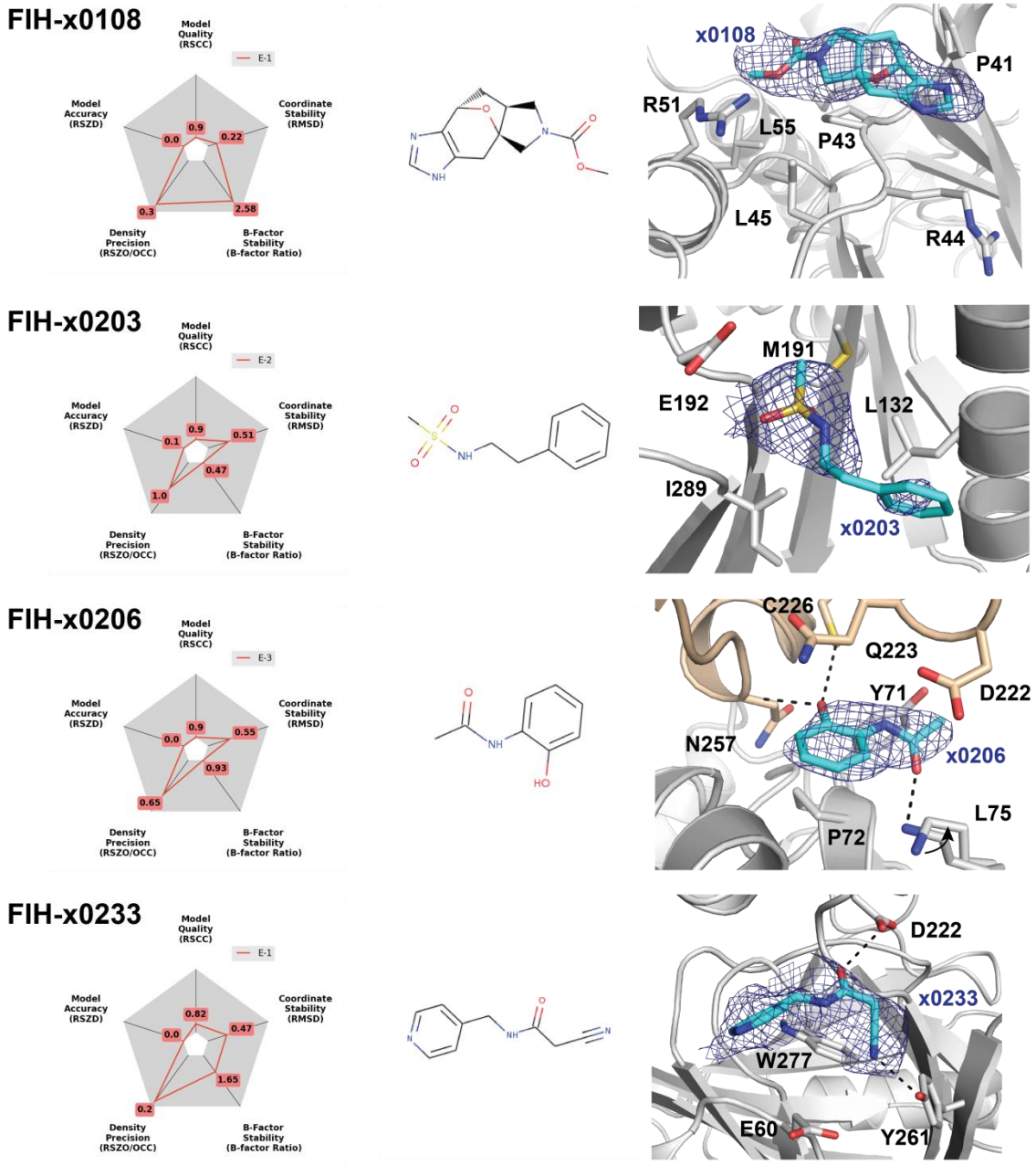


Figure 11 Analysis of fragments binding to additional sites of FIH. Ligand validation plots, 2D structure and ligand-protein interactions given for each ligand. PANDDA event maps contoured to 2σ around the ligand. Side chain movements induced by ligand binding are indicated with black arrows. FIH is coloured in light grey and its dimer dark grey and further symmetry mates in the crystal packing are coloured in light orange.

3.3.6 Validation and potential follow-up on fragments hits

The hits identified from the crystallographic fragment screen were tested for their ability to inhibit FIH activity using a LC-MS based activity assay (Table 4). From all fragment tested, x0182 was found to strongly inhibit FIH activity, an IC_{50} =1.1 μM for x0182 was measured in collaboration with Dr Anthony Tumber (University of Oxford). Potential follow-up compounds are discussed in Figure 12 and Figure 13.

Note that fragment x0107 does not bind to the HIF-1α binding site. Overlap with the ankyrin binding might be of interest for the development of a substrate specific FIH inhibitor (Figure 14).

Table 4 Preliminary FIH inhibition data from the hits of the fragment screen. Percentage FIH inhibition given in bold numbers, normalised to 2,4 PDCA inhibition. Substrate hydroxylation measured by LC-MS. Reaction conditions: 100 nM FIH, 10 μM HIF-1α 39mer peptide, 10 μM 2OG, 100 μM L-ascorbic acid, 10 μM iron(II) ammonium sulphate, 50 mM Tris-HCl pH = 7.5, 50 mM NaCl, 40 min at ambient temperature.

100 μM		200 μM	
6.6 (x0322)	**7.1** (x0206)	**5.2** (x0322)	**10.5** (x0206)
100 (x0182)	**34.5** (x0218)	**100** (x0182)	**37.5** (x0218)
30.7 (x0184)	**2.3** (x0233)	**29.2** (x0184)	**3.0** (x0233)
1.6 (x0069)	**17.3** (x0239)	**2.2** (x0069)	**17.7** (x0239)
0.8 (x0077)	**4.2** (x0243)	**0.7** (x0077)	**4.5** (x0243)
0.5 (x0078)	**3.7** (x0245)	**-1.3** (x0078)	**6.0** (x0245)
1.4 (x0101)	**0.3** (x0249)	**-0.3** (x0101)	**1.5** (x0249)
1.2 (x0107)	**0.0** (x0256)	**1.6** (x0107)	**0.1** (x0256)
-2.8 (x0108)	**0.7** (x0275)	**-2.2** (x0108)	**-0.8** (x0275)
1.3 (x0130)	**0.6** (x0281)	**0.7** (x0130)	**0.9** (x0281)
3.5 (x0133)	**4.2** (x0315)	**3.1** (x0133)	**3.0** (x0315)
8.1 (x0203)	**1.8** (x0316)	**4.9** (x0203)	**2.2** (x0316)
13.6 (x0322)	**2.3** (x0330)	**8.7** (x0322)	**3.5** (x0330)
3.1 (x0361)	**1.1** (x0378)	**3.9** (x0361)	**2.0** (x0378)
1.0 DMSO	**100** 2,4-PDCA	**-1.0** DMSO	**100** 2,4-PDCA

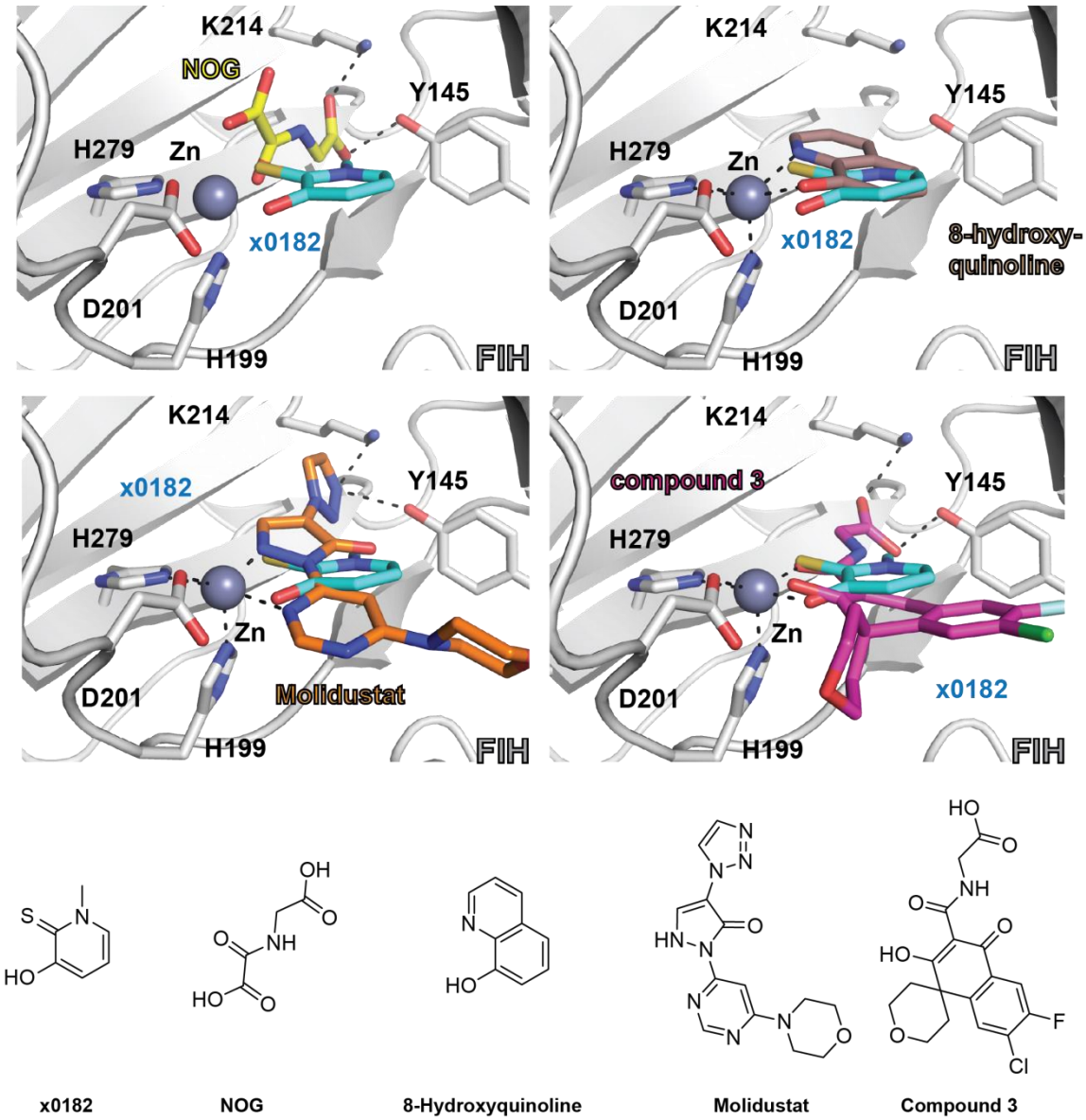


Figure 12 Structure based design of potential follow-up compounds derived from x0182. Views the crystal structures of FIH in complex with x0182, 8-hydroxyquinoline (2.59 Å, PDB code 3KCY[23]), Molidustat and compound 3 (see chapter 2).

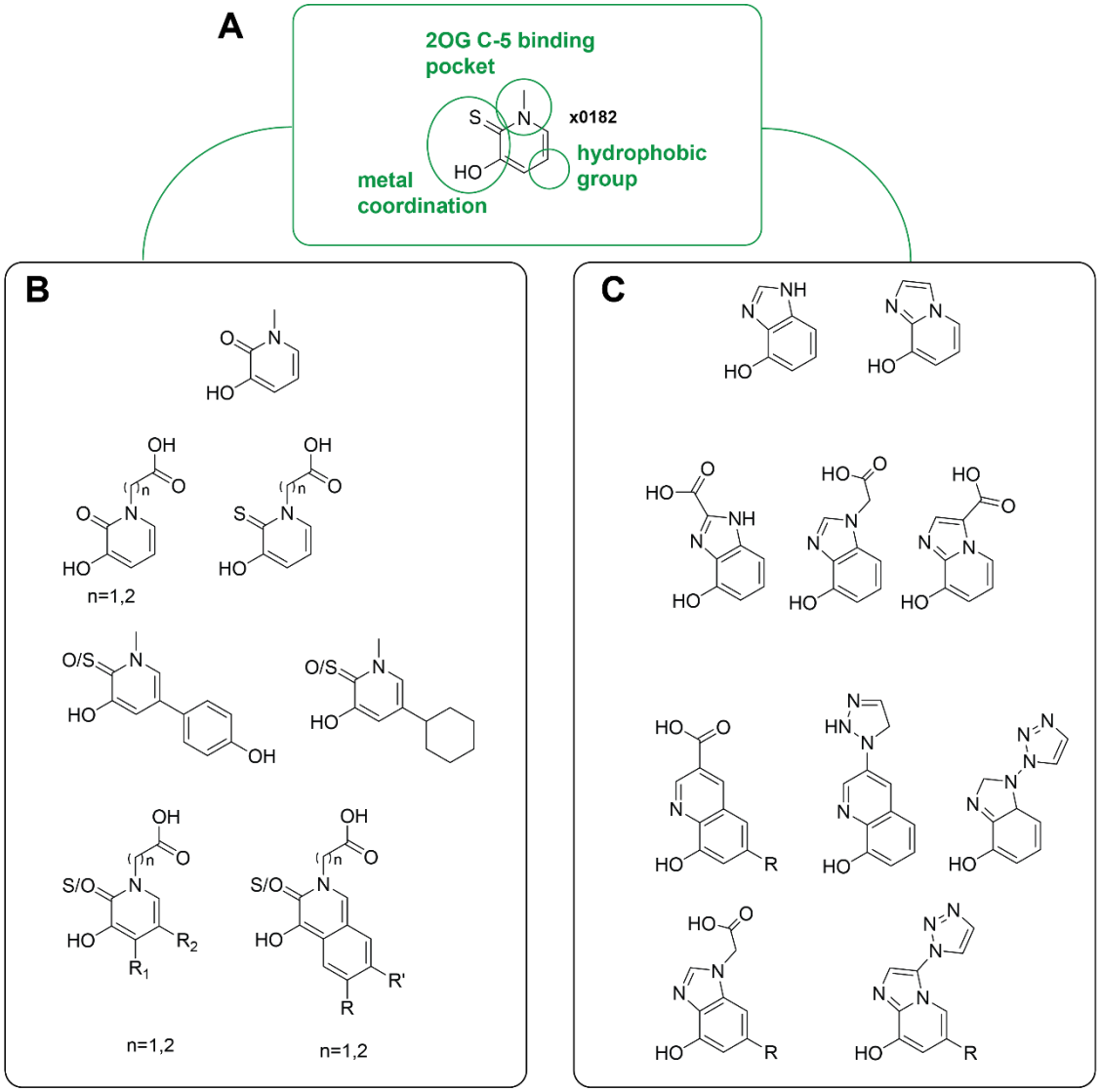


Figure 13 Potential follow-up compounds based on x0182. Structure of x0182 with marked potential expansion sites. **B** Potential follow-up compounds based on the x0182 scaffold. Extensions designed on the basis on the structure of FIH in complex with Molidustat and compound 3 (see above). **C** Potential follow-up compounds based on the structural comparison of x0182 binding with hydroxyl-quinoline, Molidustat and compound 3.

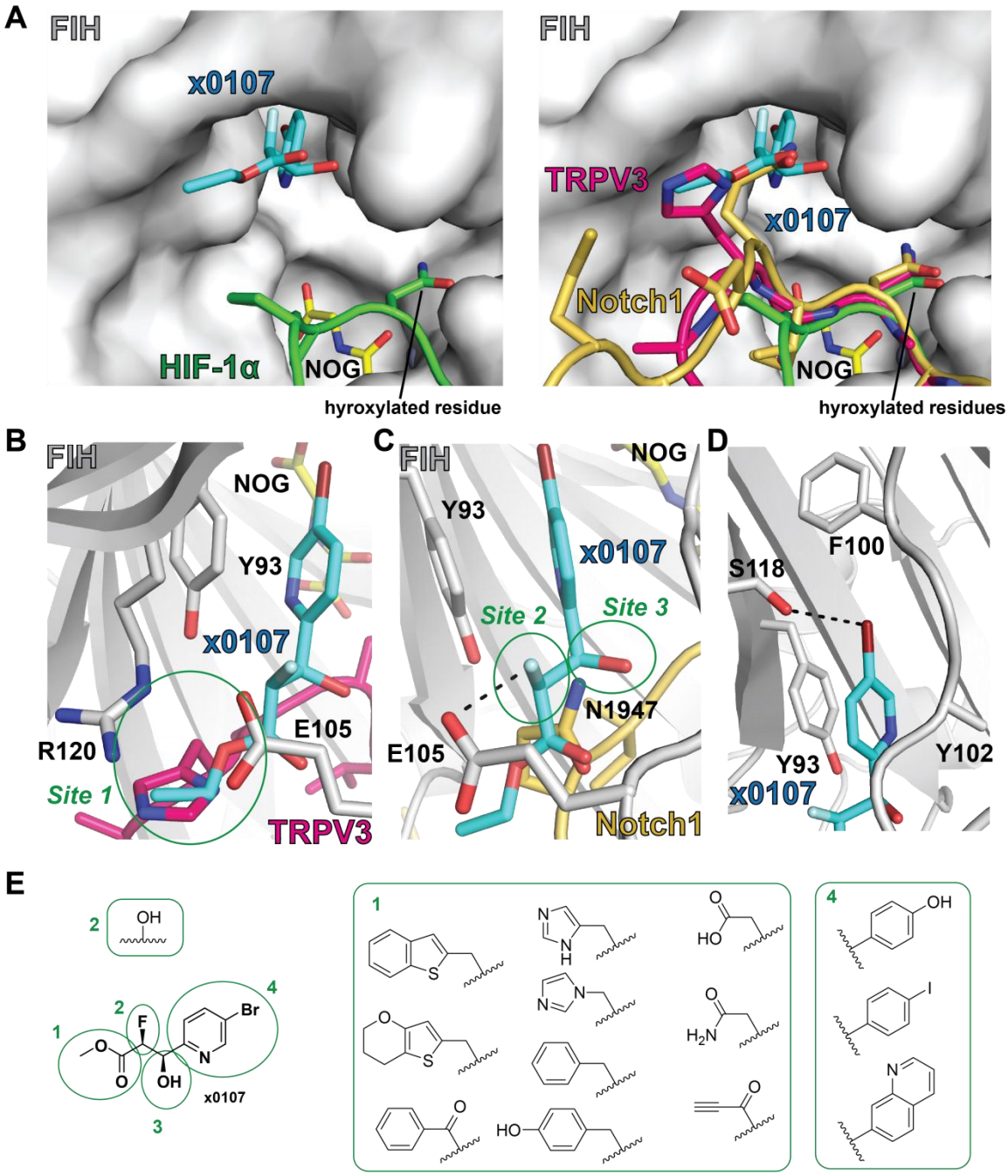


Figure 14 Structure based design of potential follow-up compounds derived from x0107. **A** Views from a crystal structure of FIH in complex with x0107 overlayed with HIF-1α (generated from PDB code 1H2K[20]), TRPV3 and NOTCH1 (see chapter 4 and chapter 5). **B** Overlay of the crystal structures of FIH in complex with x0107 and TRPV3 suggested potential follow-up modifications in from of aromatic groups interacting with arginine-120 of FIH. **C** Overlay of a crystal structure of FIH in complex with x0107 and mNOTCH1 suggests further modifications to the initial fragment. The fluorine atom of site 2 could be replaced by a hydroxy- group and the hydroxyl group of site 3 removed. **D** Interactions of the bromo-pyridinyl group of x0107 with FIH, note that tryrosine-102 is disordered, possibly due to steric clashes with the hydroxyl group in site 3 of x0107. **E** Suggested modification to the initial hit fragment. Easy preparation of larger libraries can be achieved by compounds with carboxylic acid, amides or alkynes in site one allowing for easy derivatization.

3.3.7 Data collection statistics

Table 5 Data collection statistics of the FIH fragment screen.

Datasets	**FIH-x0069**	**FIH-x0075**	**FIH-x0077**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-DIALS	Xia2-DIALS	Xia2-3d (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.4, 86.4, 148.8	86.3, 86.3, 148.2	86.6, 86.6, 148.0
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	37629 (2720)*	23549 (1710)*	27553 (2005)*
Resolution (Å)	74.39-2.02 (2.07-2.02)*	86.33-2.37 (2.43-2.37)*	56.60-2.25 (2.31-2.25)*
R_{sym} or R_{merge}**	0.066 (1.68)*	0.106 (2.265)*	0.054 (1.729)*
I/σI	13.3 (1.3)*	12.1 (1.3)*	19.7 (1.3)*
CC 1/2	0.999 (0.695)*	0.999 (0.564)*	0.999 (0.563)*
Completeness (%)	99.8 (99.5)*	100 (100)*	99.9 (100)*
Multiplicity	8.5 (8.9)*	8.5 (8.9)*	8.6 (8.3)*
Datasets	**FIH-x0078**	**FIH-x0101**	**FIH-x0107**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	fast-dp	Xia2-DIALS	Xia2-DIALS
Space group	*P*422	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.3, 86.3, 148.2	86.2, 86.2, 150.4	86.3, 86.3, 150.5
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	23549 (1710)*	21963 (1611)*	31280 (2259)*
Resolution (Å)	86.33-2.37 (2.43-2.37)*	56.66-2.41 (2.47-2.41)*	74.88-2.16 (2.22-2.16)*
R_{sym} or R_{merge}**	0.94 (2.019)*	0.86 (1.063)*	0.082 (1.645)*
I/σI	12.1 (1.3)*	11.5 (1.8)*	10.8 (1.1)*
CC 1/2	0.999 (0.564)*	0.999 (0.752)*	0.999 (0.578)*
Completeness (%)	100 (100)*	97.8 (98.9)*	100 (99.9)*
Multiplicity	8.5 (8.9)*	8.8 (9.1)*	8.4 (8.5)*

* Highest resolution shell in parentheses.

Table 5 (continued)

Datasets	**FIH-x0108**	**FIH-x0130**	**FIH-x0133**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-3d (XDS)	Xia2-3d (XDS)	Xia2-3d (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	87.1, 87.1, 150.2	86.0, 86.0, 151.2	86.3, 86.3, 148.9
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	20215 (1462)*	23470 (1679)*	33163 (2404)*
Resolution (Å)	43.39-2.52 (2.59-2.52)*	38.47-2.38 (2.44-2.38)*	26.83-2.11 (2.16-2.11)*
R_{sym} or R_{merge}**	0.081 (2.287)*	0.059 (1.490)*	0.045(1.628)*
I/σI	13.5 (0.8)*	15.8 (1.3)*	20.5 (1.3)*
CC 1/2	0.999 (0.444)*	0.998 (0.645)*	1.00 (0.625)*
Completeness (%)	99.8 (100)*	99.9 (99.7)*	99.9(100)*
Multiplicity	8.3 (9.0)*	8.6 (9.1)*	8.6 (8.9)
Datasets	**FIH-x0180**	**FIH-x0182**	**FIH-x0184**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia-3d (XDS)	Xia2-DIALS	Xia2-3dii (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	85.6, 85.6, 149.5	86.3, 86.3, 147.7	86.4, 86.4, 147.6
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	41226 (2974)*	23667 (1719)*	16937 (1233)*
Resolution (Å)	27.07-1.95 (2.00-1.95)*	73.87-2.35 (2.41-2.35)*	73.82-2.65 (2.72-2.65)*
R_{sym} or R_{merge}**	0.054 (1.341)*	0.071 (1.234)*	0.092 (1.723)*
I/σI	16.5 (1.5)*	12.5 (1.7)*	15.2 (1.3)*
CC 1/2	0.999 (0.689)*	0.997 (0.699)*	0.999 (0.549)*
Completeness (%)	99.8 (99.9)*	99.0 (99.5)*	99.8 (100)*
Multiplicity	8.6 (9.2)*	8.5 (8.8)*	8.5 (8.9)*

* Highest resolution shell in parentheses.

Table 5 (continued)

Datasets	**FIH-x0203**	**FIH-x0206**	**FIH-x0218**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-DIALS	Xia2-3dii (XDS)	Xia2-3dii (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.5, 86.5, 148.4	85.6, 85.6, 149.4	86.6, 86.6, 147.1
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	23646 (1725)*	26771 (1850)*	23931 (1763)*
Resolution (Å)	86.46-2.37 (2.43-2.37)*	74.72-2.26 (2.32-2.26)*	86.55-2.33 (2.39-2.33)*
R_{sym} or R_{merge}**	0.071 (1.033)*	0.070 (1.753)*	0.060 (1.838)*
I/σI	13.1 (1.5)*	14.5 (0.6)*	21.1 (1.2)*
CC 1/2	0.999 (0.767)*	0.997 (0.527)*	0.998 (0.482)*
Completeness (%)	100 (100)*	99.5 (95.8)*	97.9 (99.3)*
Multiplicity	8.5 (8.9)*	8.5 (8.4)*	8.7 (8.8)*
Datasets	**FIH-x0222**	**FIH-x0233**	**FIH-x0239**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-3d (XDS)	Xia2-3dii (XDS)	Xia2-3d (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.1, 86.1, 150.8	86.0, 86.0, 151.6	86.2, 86.2, 149.7
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	30223 (2302)*	18596 (1379)*	21104 (1541)*
Resolution (Å)	56.43-2.15 (2.21-2.15)*	85.96-2.57 (2.64-2.57)*	38.62-2.46 (2.52-2.46)*
R_{sym} or R_{merge}**	0.080 (3.792)*	0.090 (1.910)*	0.107 (1.299)*
I/σI	15.2 (0.7)*	11.1 (0.9)*	9.4 (1.3)*
CC 1/2	0.999 (0.548)*	0.996 (0.680)*	0.996 (0.729)*
Completeness (%)	95.6 (100)*	98.9 (100)*	99.6 (100)*
Multiplicity	8.2 (8.7)*	8.0 (8.9)*	7.9 (9.1)*

* Highest resolution shell in parentheses.

Table 5 (continued)

Datasets	**FIH-x0243**	**FIH-x0245**	**FIH-x0249**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-3dii (XDS)	Xia2-3d (XDS)	Xia2-3d (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.4, 86.4, 147.7	86.2, 86.2, 148.4	86.5, 86.5, 148.4
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	25047 (1804)*	25315 (1834)*	23615 (1722)*
Resolution (Å)	74.6-2.31 (2.37-2.31)*	56.41-2.31 (2.37-2.31)*	42.93-2.37 (2.43-2.37)*
R_{sym} or R_{merge}**	0.062 (2.674)*	0.076 (3.006)*	0.056 (1.333)*
I/σI	17.1 (1.1)*	16.2 (1.0)*	14.4 (1.6)*
CC 1/2	0.999 (0.494)*	0.999 (0.533)*	0.999 (0.751)*
Completeness (%)	99.3 (99.2)*	99.8 (100)*	99.9 (99.9)*
Multiplicity	8.5 (8.6)*	8.3 (8.4)*	8.5 (9.1)*
Datasets	**FIH-x0256**	**FIH-x0275**	**FIH-x0277**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-3dii (XDS)	fast-dp	fast-dp
Space group	$P4_12_12$	*P*422	*P*422
Cell dimensions			
a,*b*,*c* (Å)	86.0, 86.0, 147.2	86.2, 86.2, 150.3	86.4, 86.4, 150.2
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	18946 (1395)*	27231 (1944)*	27517 (1901)*
Resolution (Å)	56.61-2.49 (2.55-2.49)*	29.86-2.25 (2.31-2.25)*	29.93-2.25 (2.31-2.25)*
R_{sym} or R_{merge}**	0.061 (0.980)*	0.062 (1.215)*	0.080 (1.412)*
I/σI	18.6 (2.1)*	16.4 (1.0)*	12.9 (1.2)*
CC 1/2	0.999 (0.699)*	0.999 (0.831)*	0.999 (0.861)*
Completeness (%)	95.5 (97.3)*	98.7 (97.2)*	99.6 (95.2)*
Multiplicity	8.8 (9.4)*	8.6 (7.9)*	8.4 (7.7)*

* Highest resolution shell in parentheses.

Table 5 (continued)

Datasets	**FIH-x0281**	**FIH-x0314**	**FIH-x0315**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-3d (XDS)	Xia2-3dii (XDS)	Xia2-3dii (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.4, 86.4, 148.9	85.9, 85.9, 149.4	86.3, 86.3, 148.9
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	24822 (1811)*	10325 (771)*	30301 (2249)*
Resolution (Å)	61.07-2.33 (2.39-2.33)*	74.72-3.10 (3.18-3.10)*	74.68-2.16 (2.22-2.16)*
R_{sym} or R_{merge}**	0.056 (1.648)*	0.238 (1.276)*	0.055 (2.015)*
I/σI	20.1 (1.3)*	9.9 (1.7)*	18.1 (1.1)*
CC 1/2	0.999 (100)*	0.989 (0.535)*	0.999 (0.545)*
Completeness (%)	99.9 (100)*	96.4 (100)*	97.7 (100)*
Multiplicity	8.6 (8.7)*	6.7 (8.7)*	8.5 (8.8)*
Datasets	**FIH-x0316**	**FIH-x0322**	**FIH-x0330**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-3dii (XDS)	Xia2-DIALS	Xia2-3dii (XDS)
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.4, 86.4, 150.0	86.1, 86.1, 147.52	86.2, 86.2, 148.0
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	30569 (2207)*	37232 (2690)*	26888 (1935)*
Resolution (Å)	75.00-2.17 (2.23-2.17)*	60.89-2.02 (2.07-2.02)*	74.49-2.26 (2.32-2.26)*
R_{sym} or R_{merge}**	0.057 (2.261)*	0.087 (1.422)*	0.059 (2.700)*
I/σI	18.6 (1.3)*	9.8 (1.3)*	18.4 (0.7)*
CC 1/2	0.999 (0.613)*	0.998 (0.730)*	0.999 (0.759)*
Completeness (%)	99.1 (99.0)*	100 (100)*	99.8 (99.2)*
Multiplicity	8.4 (8.7)*	8.4 (8.9)*	8.5 (8.4)*

* Highest resolution shell in parentheses.

Table 5 (continued)

Datasets	**FIH-x0361**	**FIH-x0378**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-3dii (XDS)	Xia2-DIALS
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	86.6, 86.6, 147.0	85.7, 85.7, 150.4
α,β,γ(°)	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1
No. reflections	19603 (1439)*	18129 (1334)*
Resolution (Å)	86.62-2.52 (2.59-2.52)*	60.58-2.57 (2.64-2.57)*
R_{sym} or R_{merge}**	0.065 (1.514)*	0.200 (1.708)*
I/σI	17.2 (1.4)*	6.7 (1.3)*
CC 1/2	0.999 (0.753)*	0.737 (0.627)*
Completeness (%)	99.8 (100)*	98.4 (99.1)*
Multiplicity	8.3 (8.9)*	8.4 (8.9)*

* Highest resolution shell in parentheses.

3.4 In-crystal fragment reactions – initial experiments

Fragment based drug discovery is based on the screening of small-molecules with less complex nature compared to traditional HTS. Fragment hits are usually weakly binding and not potent bioactive molecules[1-3]. A key step following from initial hits in fragment screening is the extension of the fragments to a lead compound.

In the course of the optimisation of crystallographic fragment screening for FIH, a novel approach for in-crystal reactions of boronates with hydrazines or thiol-amino compounds was developed for metallo-β-lactamases (Dr Jurgen Brem, University of Oxford, unpublished data).

Boronic esters have been shown to be 2OG oxygenase inhibitors[24]: Small libraries of boronic esters were screened against PHD2 using dynamic-combinatorial mass spectrometry (DCMS)[24]. The approach was based on the equilibrium of the reaction of diols with boronic acids with the resulting boronic esters. Unlike DCMS, reactions of boronates with hydazines are based on the equilibrium between different products of the reaction (Figure 15)[25-28]. The protein crystal environment can shift the reaction equilibrium to different products than in solution. (Dr Jurgen Brem, University of Oxford, unpublished data).

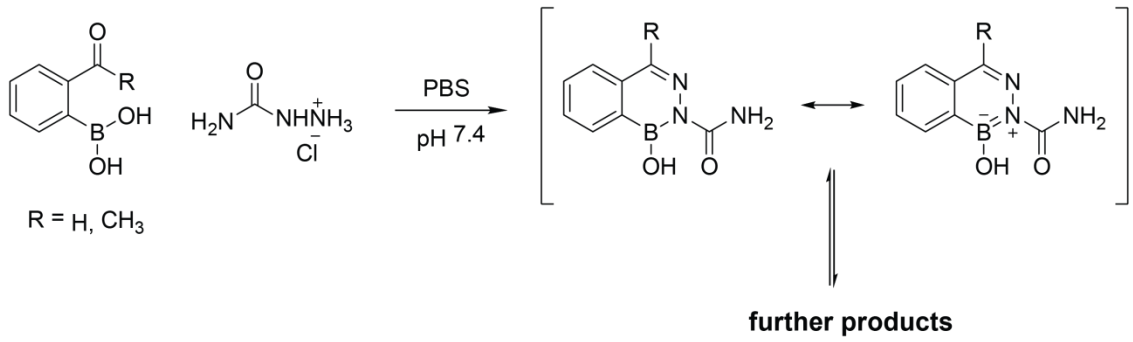


Figure 15 Reaction of boronates with hydrazines. Products from the reaction are dependent on the reactant concentrations and pH values of the solution. Different products of the reactions were observed in the presence of crystallised proteins (Dr Jurgen Brem, University of Oxford, unpublished data).

Following from the successful establishment of a crystallographic fragment screen for FIH, it was further tested if the fragments could directly react within the crystal to form larger compounds during soaking of the fragments.

FIH was incubated with either 6 boronic acids or 14 hydrazides or thiol-amino compounds (Table 6) or an equimolar mixture of two of these reactants at concentrations between 2.5 - 12.5 mM. In total 104 crystals were tested.

Table 6 List of reactants used for in-crystal reactions.

#	boronic acids	#	hydrazides and thio-amino compounds
BA_01	2-Formylphenylboronic acid	RG_06	DL-Cysteine hydrochloride
BA_02	2-Acetylphenylboronic acid	RG_07	D-Cysteine hydrochloride
BA_03	2-Formyl-3-thienylboronic acid	RG_08	L-Cysteine methyl ester hydrochloride
BA_04	2-Acetyl-3-thiopheneboronic acid	RG_09	L-Homocysteine
BA_05	3-Formyl-2-thienylboronic acid	RG_10	L-Penicillamine
BA_06	2-Fluoro-5-acetylphenylboronic acid	RG_11	D-Penicillamine
		RG_12	Carbohydrazide
		RG_13	Semicarbazide hydrochloride
		RG_14	Thiosemicarbazide
		RG_15	Methanesulfonohydrazide
		RG_16	Benzenesulfonyl hydrazide
		RG_17	4-Phenylthiosemicarbazide
		RG_18	2-(4-Fluorophenoxy)acetohydrazide
		RG_19	4-(4-Chlorophenyl)semicarbazide hydrochloride

Electron density was observed for the proposed product of BA_2 and RG_12 in the -8 leucine binding pocket at the dimer interface of FIH (Figure 16). This site has been found to be one of the fragment binding hot-spots on FIH (see above). Of note is the overlap of the compound with the FIH substrate binding site (Figure 16).

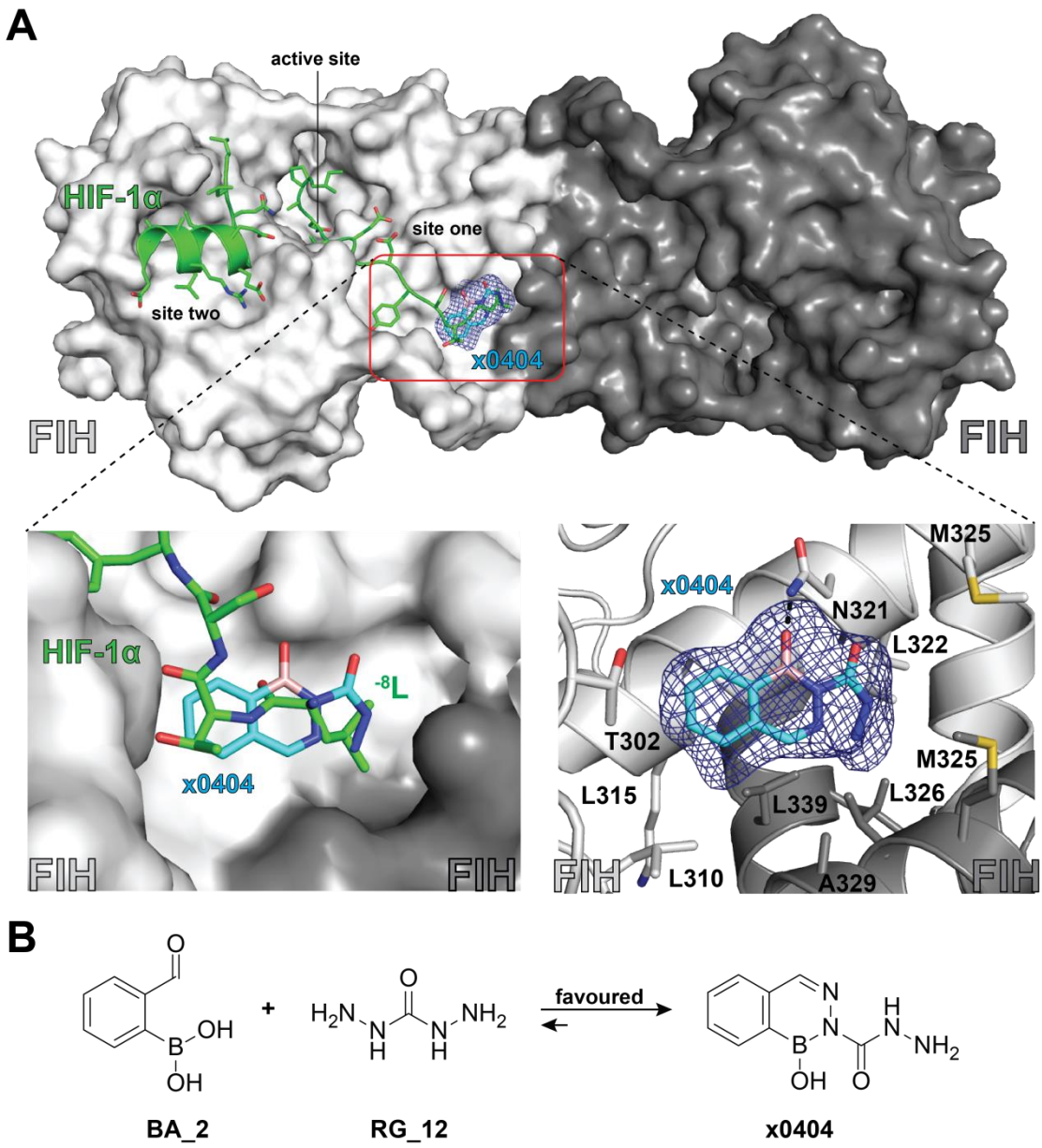


Figure 16 View from a crystal structure of FIH in complex with x0404. A Views from a crystal structure of the FIH in complex with x0404 illustrates the binding site of the product formed from the in-crystal reaction of BA_2 and RG_12. The binding site overlaps with the HIF-1α binding site (HIF-1α binding overlayed from PDB code 1H2K[20]). Note, the symmetry related fragment bound to the dimer is not shown. Interactions in the -8 leucine binding pocket are predominantly hydrophobic interactions. PANDDA event maps are contoured to 2σ around the ligand. **B** Proposed reaction of BA_2 and RG_12 to form the product x0404.

To test for potential inhibition, x0404 and related compounds were screened for their potential to inhibit FIH mediated HIF-1α (39mer) hydroxylation (Table 7). Compounds were synthetized by Mr Asheley Dey under supervision of Dr Jurgen Brem (both University of Oxford).

None of the tested compounds were found to inhibit FIH activity using HIF-1α at 100 µM or 200 µM compound concentration (Table 7). Surprisingly, x0404 and AR_015 showed increased hydroxylation of the HIF-1α substrate (Table 7). At 100 µM concentration of x404 and AR_15 nearly complete hydroxylation of HIF-1α was observed compared to ~60% in the DMSO control. Compounds AR_002 and AR_015, which differ from x0404 and AR_015 only by one methyl group, did not show any effect on FIH activity. Assuming a similar binding mode between AR_015 and x0404, this could be due to steric clashes in the binding pocket. Crystallisation trials with both compounds in the presence and absence of HIF substrate are ongoing.

In addition to x0404 binding to FIH further notable sites were identified as potential reactant binding sites of interest. In the case of the reactions of 2-formyl-3-thienylboronic acid (BA_3) and 3-formyl-2-thienylboronic acid (BA_5) both with thiosemicarbazide (RG_14) large negative density was observed in the PANDDA z-maps (a Fo-$Fo_{average}$ difference map) in the active site around the zinc ion and 2OG (data not shown). Boronic acids have been found to covalently react with specific histidine residues under the tested conditions (Dr Jurgen Brem, University of Oxford, unpublished data) but poor data quality and bad map alignment can lead to false positives in the hit detection. Further validation is ongoing.

Several boronates appear to react with glutamate-225 of FIH (analysis ongoing, data not shown) as indicated by large peaks in the PANDDA z-maps (>8 rmsd). The exact nature of the reactions is unknown as low map resolution limits the interpretation of the density observed. The analysis is ongoing using complementary methods including denaturing and non-denaturing mass-spectrometry of protein-ligand complexes.

Table 7 Preliminary inhibition tests of boronates against FIH mediated HIF-α hydroxylation. Boronates derived from x0404 were tested for FIH inhibition as previously described. Inhibition is normalized to 2,4 PDCA. Note, at 100 µM x0404 and AR_015 showed near complete HIF hydroxylation. The percentage inhibition is the average of two measurements and stated below the compound. Compound names are given in brackets. Reaction conditions as stated in Table 4.

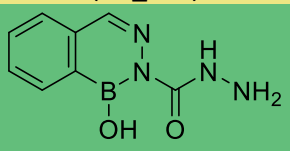
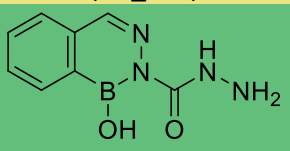
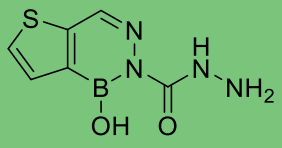
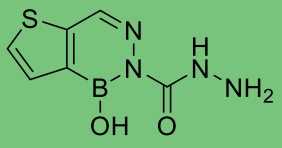
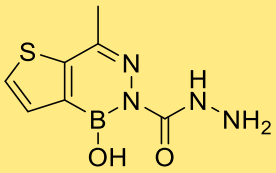
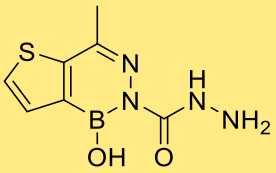
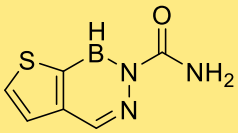
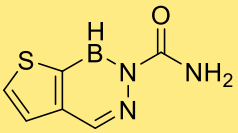
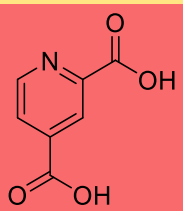
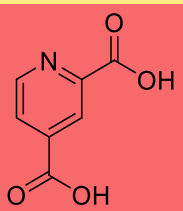
100 µM	200 µM
 -7.6 (AR_002)	 -8.3 (AR_002)
 -57.5 (x0404)	 -57.1 (x0404)
 -49.3 (AR_015)	 -49.3 (AR_015)
 -0.7 (AR_017)	 -0.7 (AR_017)
 -1.8 (AR_018)	 -1.5 (AR_018)
 1.0 (DMSO)	 -1.0 (DMSO)
 100.0 (2,4 PDCA)	 100.0 (2,4 PDCA)

Table 8 Data collection statistics of the FIH.Zn.NOG.x0404 complex.

Datasets	**FIH-x0404**
Beamline (Wavelength, Å)	DLS I04-1 (0.9759)
Detector	PILATUS 6M-F
Data Processing	Xia2-3d (XDS)
Space group	$P4_12_12$
Cell dimensions	
a,*b*,*c* (Å)	86.2, 86.2, 147.7
α,β,γ(°)	90, 90, 90
No. of molecules/ASU	1
No. reflections	25141 (1224)*
Resolution (Å)	38.55-2.31 (2.35-2.31)*
R_{sym} or R_{merge}**	0.055 (1.626)*
I/σI	22.1 (1.1)*
CC 1/2	1.0 (0.6)*
Completeness (%)	100 (99.5)*
Multiplicity	8.6 (8.6)*

* Highest resolution shell in parentheses.

3.5 Conclusions and outlook

In summary, a workflow for high-throughput crystallographic fragment screening was optimised for FIH and ~340 fragments were screened against FIH. In total, 32 fragments were found to bind to FIH in the crystal state, mainly clustering in 4 distinct sites. Three of the identified fragment binding sites of FIH overlap with the HIF-1α substrate binding and are potential starting points for allosteric inhibitors. One fragment, x0182, displaced 2OG in the crystal state and was found to inhibit FIH with an IC_{50}=1.1 μM, presumably by competing with 2OG. Based on further FIH inhibitor structures, potential follow-up compounds were designed. Further fragment optimisation is ongoing.

Using an in-crystal reaction methodology which is based on the reaction of fragments in solution, a proof-of-principle in crystal fragment linking approach was successfully tested for FIH. A novel class of compounds interacting with FIH based on cyclic boronates is described. Surprisingly, some of these compounds appear to enhance FIH activity. The exact mechanism of activation is unknown and subject to ongoing investigations including concentration dependence of the effects observed on FIH activity and crystallographic studies.

3.6 References

1 Hajduk, P. J. & Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat Rev Drug Discov* **6**, 211-219, doi:10.1038/nrd2220 (2007).

2 Hesterkamp, T. & Whittaker, M. Fragment-based activity space: smaller is better. *Curr Opin Chem Biol* **12**, 260-268, doi:10.1016/j.cbpa.2008.02.005 (2008).

3 Erlanson, D. A., Wells, J. A. & Braisted, A. C. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* **33**, 199-223, doi:10.1146/annurev.biophys.33.110502.140409 (2004).

4 Krojer, T. *et al.* The XChemExplorer graphical workflow tool for routine or large-scale protein-ligand structure determination. *Acta Crystallogr D Struct Biol* **73**, 267-278, doi:10.1107/S2059798316020234 (2017).

5 Pearce, N. M. *et al.* A multi-crystal method for extracting obscured crystallographic states from conventionally uninterpretable electron density. *Nat Commun* **8**, 15123, doi:10.1038/ncomms15123 (2017).

6 Collins, P. M. *et al.* Gentle, fast and effective crystal soaking by acoustic dispensing. *Acta Crystallogr D Struct Biol* **73**, 246-255, doi:10.1107/S205979831700331X (2017).

7 Ng, J. T., Dekker, C., Kroemer, M., Osborne, M. & von Delft, F. Using textons to rank crystallization droplets by the likely presence of crystals. *Acta Crystallogr D Biol Crystallogr* **70**, 2702-2718, doi:10.1107/S1399004714017581 (2014).

8 Bradley, A. R. *et al.* The SGC beyond structural genomics: redefining the role of 3D structures by coupling genomic stratification with fragment-based discovery. *Essays Biochem* **61**, 495-503, doi:10.1042/EBC20170051 (2017).

9 Patel, D., Bauman, J. D. & Arnold, E. Advantages of crystallographic fragment screening: functional and mechanistic insights from a powerful platform for efficient drug discovery. *Prog Biophys Mol Biol* **116**, 92-100, doi:10.1016/j.pbiomolbio.2014.08.004 (2014).

10 Yeh, T.-L. *et al.* Molecular and cellular mechanisms of HIF prolyl hydroxylase inhibitors in clinical trials. *Chem. Sci.* **8**, 7651-7668, doi:10.1039/C7SC02103H (2017).

11 Cockman, M. E. *et al.* Posttranslational hydroxylation of ankyrin repeats in IkappaB proteins by the hypoxia-inducible factor (HIF) asparaginyl hydroxylase, factor inhibiting HIF (FIH). *Proc Natl Acad Sci U S A* **103**, 14767-14772, doi:10.1073/pnas.0606877103 (2006).

12 Coleman, M. L. *et al.* Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038, doi:10.1074/jbc.M704102200 (2007).

13 Schmierer, B., Novák, B. & Schofield, C. J. Hypoxia-dependent sequestration of an oxygen sensor by a widespread structural motif can shape the hypoxic response--a predictive kinetic model. *BMC Syst Biol* **4**, 139, doi:10.1186/1752-0509-4-139 (2010).

14 Tan, E. Y. *et al.* Cytoplasmic location of factor-inhibiting hypoxia-inducible factor is associated with an enhanced hypoxic response and a shorter survival in invasive breast cancer. *Breast Cancer Res* **9**, R89, doi:10.1186/bcr1838 (2007).

15 Giatromanolaki, A. *et al.* Expression of prolyl-hydroxylases PHD-1, 2 and 3 and of the asparagine hydroxylase FIH in non-small cell lung cancer relates to an activated HIF pathway. *Cancer Lett* **262**, 87-93, doi:10.1016/j.canlet.2007.11.041 (2008).

16 Tan, E. Y. *et al.* The key hypoxia regulated gene CAIX is upregulated in basal-like breast tumours and is associated with resistance to chemotherapy. *Br J Cancer* **100**, 405-411, doi:10.1038/sj.bjc.6604844 (2009).

17 Aimon, A. *et al.* Realisation of small molecule libraries based on frameworks distantly related to natural products. *Org Biomol Chem* **16**, 3160-3167, doi:10.1039/c8ob00688a (2018).

18 Foley, D. J. *et al.* Synthesis and Demonstration of the Biological Relevance of sp. *Chemistry* **23**, 15227-15232, doi:10.1002/chem.201704169 (2017).

19 Cox, O. B. *et al.* A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chemical Science* **7**, 2322-2330, doi:10.1039/C5SC03115J (2016).
20 Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806, doi:10.1074/jbc.C200644200 (2003).
21 Caines, M. E., Elkins, J. M., Hewitson, K. S. & Schofield, C. J. Crystal structure and mechanistic implications of N2-(2-carboxyethyl)arginine synthase, the first enzyme in the clavulanic acid biosynthesis pathway. *J Biol Chem* **279**, 5685-5692, doi:10.1074/jbc.M310803200 (2004).
22 Wilkins, S. E. *et al.* Factor inhibiting HIF (FIH) recognizes distinct molecular features within hypoxia-inducible factor-α (HIF-α) versus ankyrin repeat substrates. *J Biol Chem* **287**, 8769-8781, doi:10.1074/jbc.M111.294678 (2012).
23 Moon, H., Han, S., Park, H. & Choe, J. Crystal structures of human FIH-1 in complex with quinol family inhibitors. *Mol Cells* **29**, 471-474, doi:10.1007/s10059-010-0058-3 (2010).
24 Demetriades, M. *et al.* Dynamic combinatorial chemistry employing boronic acids/boronate esters leads to potent oxygenase inhibitors. *Angew Chem Int Ed Engl* **51**, 6672-6675, doi:10.1002/anie.201202000 (2012).
25 Bandyopadhyay, A., Cambray, S. & Gao, J. Fast Diazaborine Formation of Semicarbazide Enables Facile Labeling of Bacterial Pathogens. *J Am Chem Soc* **139**, 871-878, doi:10.1021/jacs.6b11115 (2017).
26 Bandyopadhyay, A., Cambray, S. & Gao, J. Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation. *Chem Sci* **7**, 4589-4593, doi:10.1039/C6SC00172F (2016).
27 Bandyopadhyay, A. & Gao, J. Iminoboronate Formation Leads to Fast and Reversible Conjugation Chemistry of α-Nucleophiles at Neutral pH. *Chemistry* **21**, 14748-14752, doi:10.1002/chem.201502077 (2015).
28 Kanichar, D. *et al.* Synthesis, characterization, and antibacterial activity of structurally complex 2-acylated 2,3,1-benzodiazaborines and related compounds. *Chem Biodivers* **11**, 1381-1397, doi:10.1002/cbdv.201400007 (2014).

4 FIH substrate recognition

This chapter describes the structural characterisation of Transient Receptor Potential (TRP) ion channels as FIH substrates. Structures of FIH in complex with peptides derived from TRPA1 and TRPV3 of different lengths were solved and compared to FIH-HIF-1/2α complex crystal structures. All peptides analysed show conserved binding in the site one but vary in site two binding. Potential use of differential substrate binding to FIH as strategy for a substrate specific inhibitor of FIH is discussed.

4.1 Introduction

Transient Receptor Potential (TRP) channels are a family of cation channels related to the voltage-gated ion channel superfamily[1]. The TRP channels are conserved from yeast to humans[2,3]. All TRP channels are involved in signal transduction, including sensing changes in temperature and pH; TRP channels are also activated by small molecule ligands[4-8].

The TRP channels are categorised into seven sub-families: TRPC (canonical), TRPV (vanilloid), TRPM (melastin), TPRA (ankyrin), TRPP (polycystin), TRPML (mucolipin) and TRPN (from *Drosophila melanogaster* NOMPC, not found in mammals but they are expressed in some vertebrates)[9]. With the exception of TRPM4 and TRPM5, all TRP channels are calcium channels and manifest varying selectivity over sodium cations[10].

To date, 3 TRP channels have been reported to interact with the mammalian oxygen sensors, the PHDs and FIH: TRPV4 peptides were found to be accepted by FIH as substrates[11] and, more recently, TRPV3 has also been shown to be accepted by FIH as a substrate[12]. TRPA1 has been reported to be hydroxylated by the PHDs[13-15]; however, PHD mediated hydroxylation of TRPA1 could not be reproduced in this laboratory (Kerstin Lippl, University of Oxford, unpublished data). Reported effects on TRPA1 activity in cells upon treatment with nonspecific 2OG inhibitors[13] might result from the inhibition of other 2OG oxygenases, potentially FIH.

Of the 7 TRP subfamilies, 4 incorporate ankyrin repeat domains (TRPC, TRPV, TRPN and TRPA1) in their cytosolic domains[9]. Considering the broad substrate range of FIH, any TRP channel containing ARDs could potentially interact with FIH. Further TRP hydroxylation sites modified by FIH are therefore possible, and consequently, the potential of FIH to catalyse hydroxylation of TRPA1 was investigated, as described below.

4.2 TRPA1 hydroxylation by FIH

TRPA1, the only member of the TRPA family in higher organsims[9], holds a central role in the induction of pain stimulated by endogenous inflammatory mediators and exogenous irritants such as mustard oil and tear gas. It is also activated by small molecule ligands like menthol or delta$_9$-tetrahydrocanabinol (THC)[16-20].

The cytosolic domains of TRPA1 incorporate 16 ARDs (Figure 1). A recently solved cryo-EM structure of TRPA1 revealed that TRPA1 forms a tetramer with ankyrins 12-16 forming a bundle, ankyrins 1-11 are flexible and were not modelled (Figure 1B)[21]. The sequence alignment of all ankyrin repeats of TRPA1 implies three ankyrins (ankyrin repeat 2,3,8) could potentially be accepted by FIH as substrates (for FIH substrate recognition and specificity see chapter 1).

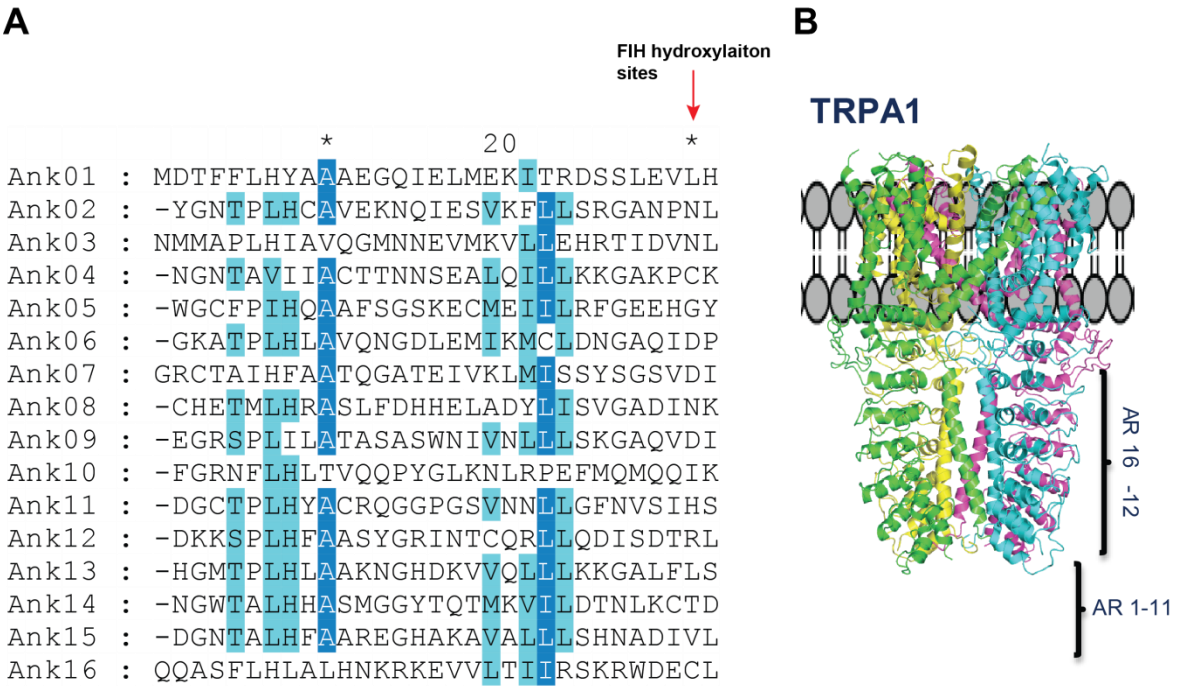


Figure 1 Sequence and structure of TRPA1 ARD domains. A Sequence alignment of the TRPA1 ankyrin repeats. The sequence alignment was generated used ClustalX[22], sequence for the ankyrins repeats are as annotated in UniProt[23]. **B** Modified view from the cryo-EM structure of TRPA1 (4.24 Å, PDB code 3J9P[21]). Note Ankyrin repeats 1-11 are not visible in the cryo-EM structure.

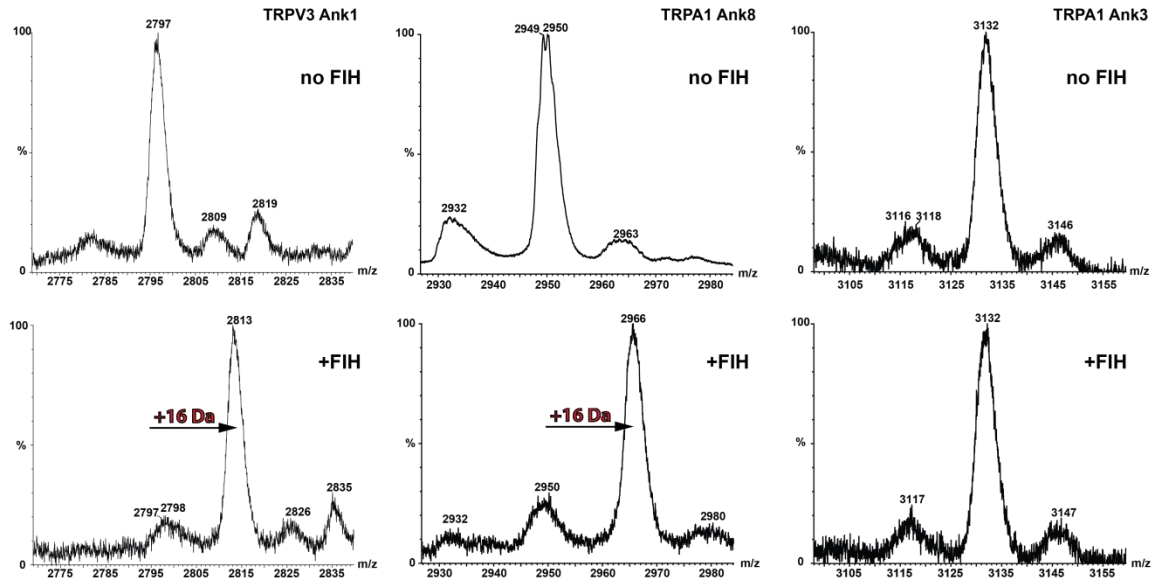


Figure 2 Hydroxylation assays of TRPA1 and TRPV3 derived peptides. End point MALDI-TOF based hydroxylation assays of TRPV3 and TRPA1 derived peptides show FIH mediated hydroxylation of TRPV3 at the asparagine corresponding to asparagine-242 and TRPA1 at the asparagine corresponding to asparagine-336. Reaction conditions: 50 µM peptide, 50 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), 50 mM Tris-HCl pH = 7.5, incubated for 30 min at 37 °C.

To test for potential FIH mediated hydroxylation of TRPA1, peptides derived from all 16 ankyrin repeats of TRPA1 were synthesised and screened as substrate for FIH (Dr Richard Hopkinson and Mr Benjamin Saward, University of Oxford). Out of the 16 ankyrins tested, only ankyrin repeat 8 was found to be hydroxylated (Figure 2). However the peptides were synthesised according to the UniProt sequence annotation[23] of the ankyrins (see Figure 1). The FIH hydroxylation site is therefore at the C-terminus of these peptides, suggesting they are likely to be poorer substrates than the peptides containing a site two sequence (further discussed below).

4.3 Crystal structures of FIH in complex with TRPA1 and TRPV3 derived peptides

4.3.1 Crystal structures of FIH in complex with TRPA1$_{313-339}$ and TRPV3 $_{220-246}$

To investigate the binding mode of TRPA1 ankyrin-8 to FIH, crystallisation trials with FIH were performed. In addition to the TRPA1 peptide, a peptide corresponding to ankyrin repeat 1 of TRPV3 was separately trialled (Figure 3).

For initial co-crystallisation trials, the TRP channel derived peptides (5-10x access) were incubated with FIH in the presence of NOG and zinc(II) for 1-2 h at 4 °C prior to crystallisation trials (for detailed protocols see chapter 10). This methodology was successfully applied for several FIH-peptide complexes (see also chapter 6) but no crystals were obtained for either of the FIH-TRP channel complexes. In the for FIH characteristic tetragonal crystal form (see Chapter 2 and Chapter 8) the active site of FIH is solvent exposed allowing for substrate binding. Whether the peptides can interact N-terminal to the known HIF interaction sites, which may interfere with efficient crystal packing, is unknown.

To overcome challenges caused by impaired crystal packing of the FIH-TRP channel complexes, FIH was crystallised in the presence of zinc and NOG; the peptide was then soaked into the crystals. Solving the crystal structure (by molecular replacement, search model PDB code 1H2K[24]) reveals strong electron density in the substrate binding site of FIH (Figure 3). The by FIH targeted asparagine residues of TRPA1 and TRPV3 bind to the active site analogously to previously observed substrates (chapter 1), stabilised by interactions with tyrosine-102, glutamin-239, histidine-199 and aspartate-201 of FIH (Figure 3).

The density for the N-terminal binding site (site one) of the FIH binding grove reveals conserved binding between both of the TRP channels and HIF-1α to FIH (further discussed below). Notably, no further interactions N-terminal of the -8 leucine binding site (see also chapter 1) could be observed for either of the substrates. This could mean that either no further interactions occur between the substrate and FIH or further interactions between FIH and the substrate are hindered by the crystal packing (see also chapter 5), which also could explain the problem in crystallising the peptide substrate complex. Crystallisation trials with FIH N-terminal truncations (discussed in detail in chapter 5), crystallising in different crystal forms, could potentially reveal further insights into FIH substrate interactions.

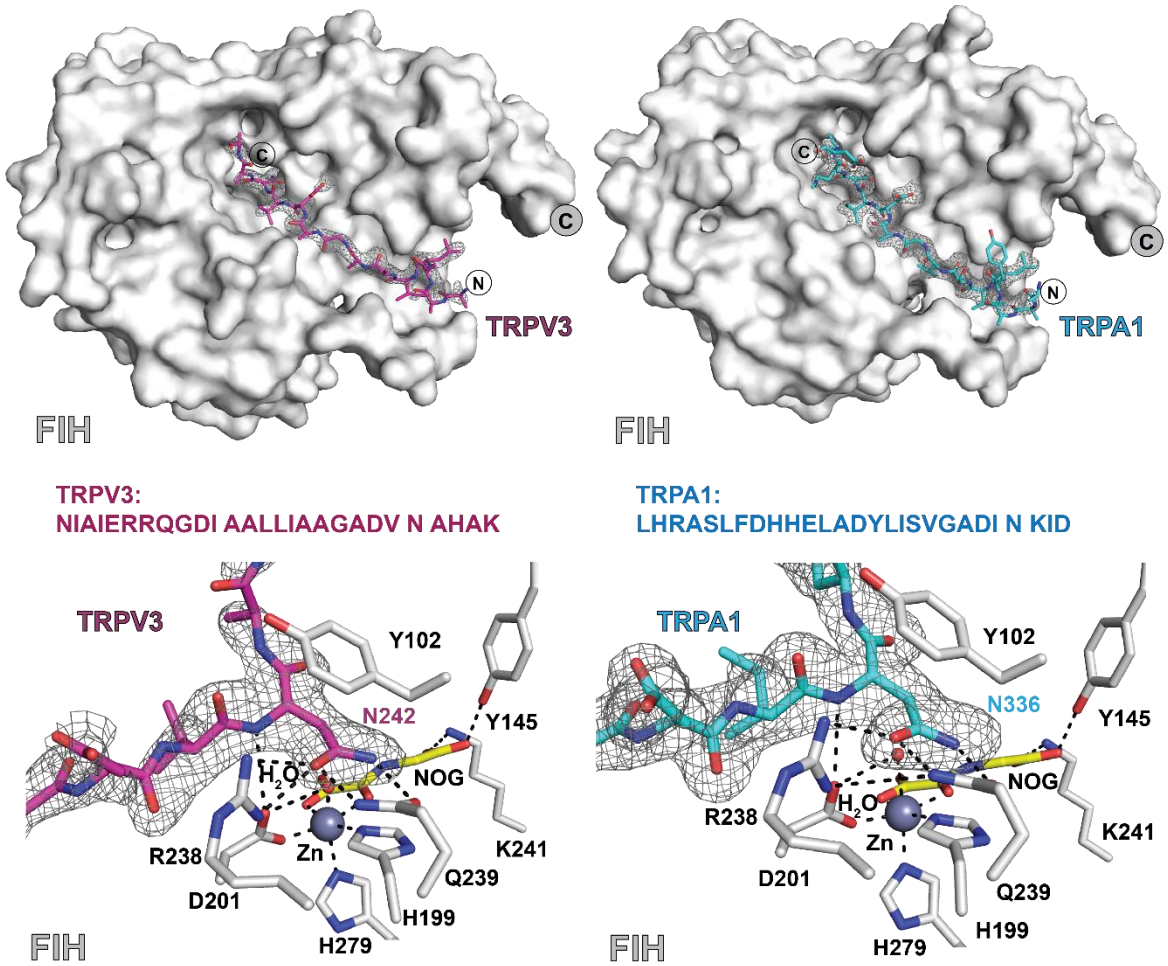


Figure 3 Views from crystal structures of FIH in complex with peptides derived from TRP channels (soaked). Top: Views from crystal structures of FIH in complex with TRPV3 and TRPA1 derived peptides illustrates the overall binding mode. **Bottom:** Views from the active site of crystal structures of FIH in complex with TRPV3 derived peptide and TRPA1 derived peptides. The mFo-Fc OMIT maps contoured to 3σ around the peptide.

Table 1 Data collection and refinement statistics of FIH in complex with TRP channel derived peptides.

Datasets	**FIH**.Zn(II).NOG TRPV3 Ank1 soaked	**FIH**.Zn(II).NOG TRPA1 Ank8 soaked
Data Collection		
Beamline (Wavelength, Å)	DLS I03 (0.9795)	DLS I03 (0.9795)
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2 3dii (XDS)	Xia2 DIALS
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	87.1, 87.1, 149.4	87.2, 87.2, 148.5
α,β,γ(°)	90, 90, 90	90, 90, 90
No. reflections	40834 (2001)*	46707 (2286)*
Resolution (Å)	74.68-1.98 (2.01-1.98)*	87.22-1.89 (1.92-1.89)*
R_{merge}	0.085 (2.821)*	0.084 (3.089)*
I/σI	18.0 (1.4)*	14.8 (1.2)*
CC1/2	1.0 (0.6)*	1.0 (0.5)*
Completeness (%)	100 (100)*	100 (100)*
Multiplicity	19.0 (20.2)*	19.0 (19.0)*
Wilson B value (Å^2)	49.7	38.8
Refinement	PHENIX	PHENIX
R_{work}/R_{free}†	0.1799 / 0.2040	0.1776 / 0.2087
No. atoms		
-Enzyme (A/B)	2755	2813
-Substrate	83	111
-Ligand	NOG (10)	NOG (10)
-Water	138	202
Avg. B-factors		
Enzyme	66.2	59.9
Substrate	112.0	86.5
Ligand	53.4	43.5
Water	64.9	63.2
R.m.s deviations		
-Bond lengths (Å)	0.004	0.006
-Bond angles (°)	0.749	0.794

* Highest resolution shell in parentheses.
†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

4.3. 2 Crystal structures of FIH in complex with TRPA1$_{321-348}$ and TRPV3$_{229-255}$

The above described crystal structures of FIH in complex with TRPV3 (residues 220-244) and TRPA1 (residues 313 to 339) show a conserved binding mode in the substrate binding grove N-terminal of the hydroxylation site (site one). Notably, no further enzyme-substrates interactions N-terminal of the -8 leucine binding pocket were observed (Figure 3). To date, site two (C-terminal of hydroxylated residue) binding is only structurally characterised for HIF-1α peptides[24]. Solution based studies on HIF-1/2α and Notch1 reveal the importance of longer constructs to improve FIH binding and hydroxylation[25-27]. This raises the question of whether site two binding may be important for FIH binding of TRP channels as well.

Peptides spanning site one and site two binding residues were designed and synthesised (Dr Richard Hopkinson and Mr Benjamin Saward, University of Oxford) and crystallisation trials performed as previously described. FIH crystallised well in complex with TRPV3 and a structure was determined as previously described, yielding electron density corresponding to the substrate for site one and site two binding residues of the substrate (Figure 4).

Upon co-crystallisation of FIH with TRPA1$_{321-348}$ little to no density of the substrate was observed in the active site. Subsequent soaking approaches, analogously to previously successfully applied in the cases of the TRPV3$_{220-244}$ and TRPA1$_{313-339}$ peptides, did not improve substrate occupancy in the crystals (data not shown). Only co-crystallisation of FIH in complex with TRPA1$_{321-348}$ followed by substrate soaking yielded satisfactory density for the substrate enabling the modelling of the substrate (Figure 4).

Comparison of the FIH-TRP complex structures revealed that the active site binding mode is nearly identical for both peptides of TRPA1 and TRPV3 (see Figure 3 and Figure 4) showing no effect on the active site interaction depending on the substrate length, at least in the crystal state (Figure 5).

TRPV3 phenylalanine-249 and phenylalanine-250 interact with a hydrophobic pocket of FIH in its site two binding site (see also chapter 1 and chapter 3). Density corresponding to three amino acids backbones could be observed at this site for the TRPA1$_{321-348}$ peptide (Figure 4). Comparable to the binding mode of the HIF-α transcription factors (Figure 7) not all residues of site two are visible in the crystal structure. Resolution of the datasets were not regarded high enough to confidently assign the amino acids and to model the side chains based on the electron density. Amino acids were modelled as alanine to illustrate the binding site (Figure 4).

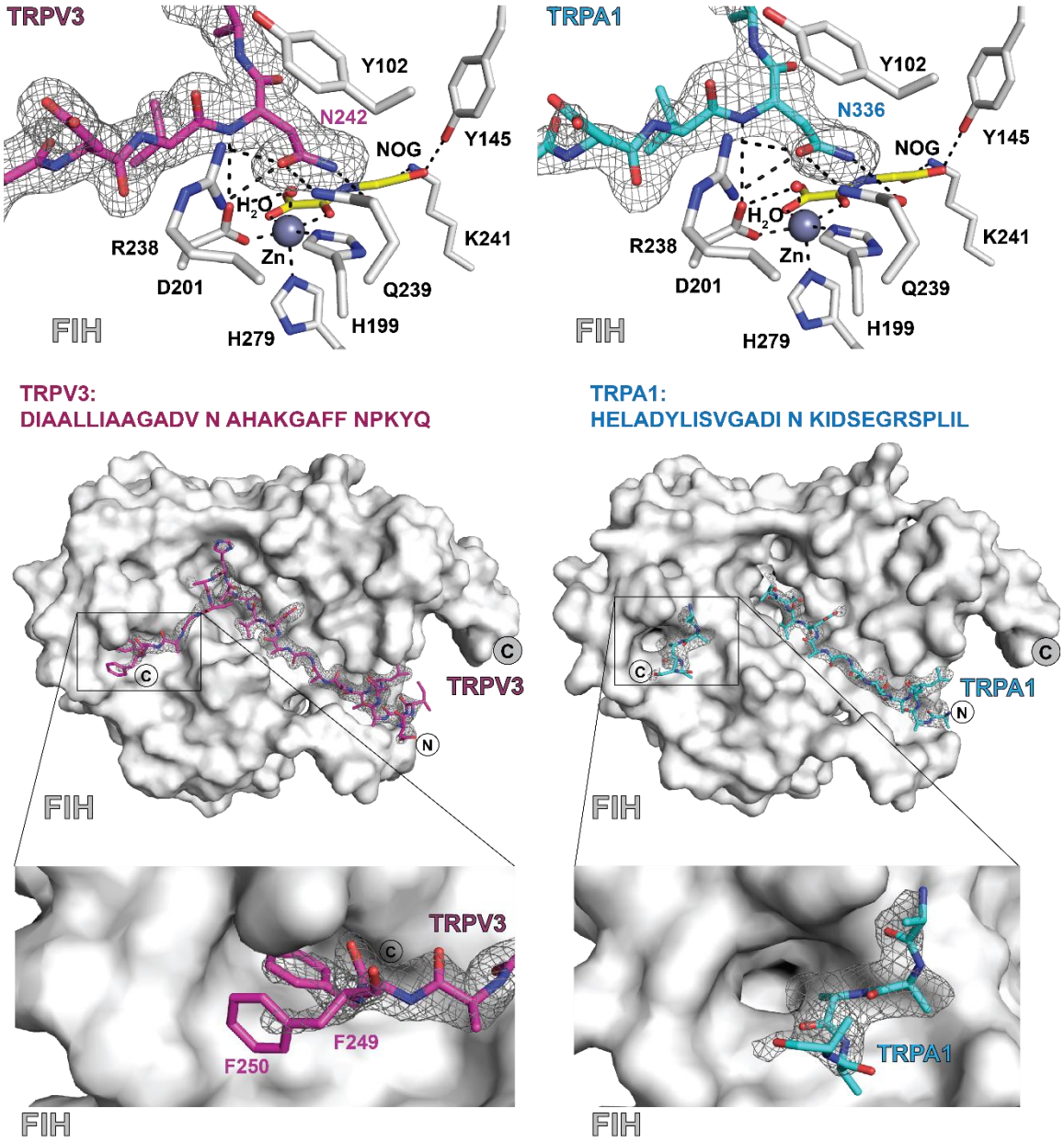


Figure 4 Views from crystal structures of FIH in complex with peptides derived from TRP channels. Top: Views from the active site of a crystal structure of FIH in complex with TRPV3 derived peptides and TRPA1 derived peptides. **Middle:** Overview of the binding sites of TRPV3 and TRPA1 derived peptides to FIH. **Bottom:** Close-up view of site two binding modes of TRPV3 and TRPA1 derived peptides. The mFo-Fc OMIT maps are contoured to 3σ around the peptides. Note, the site two binding of the TRPA1 is ambiguous and residues 336-347 could not be assigned.

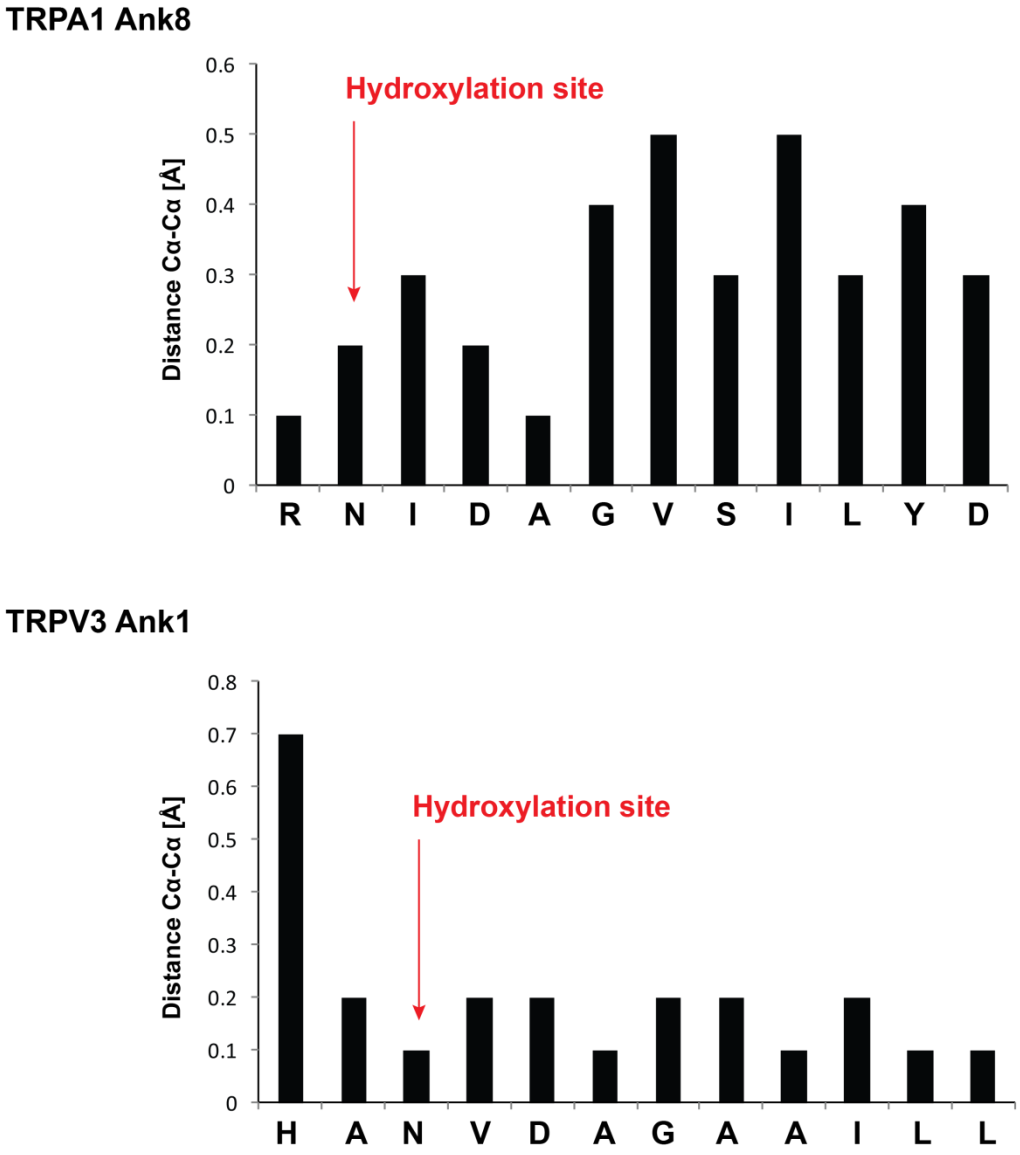


Figure 5 Comparison of TRPA1 and TRPV3 derived peptides. Distances between Cα-atoms of corresponding amino acids of TRPA1 Ank8 and TRPV3 Ank1 from different peptides.

Table 2 Data collection and refinement statistics of FIH in complex with TRPV3 and TRPA1 derived peptides.

Datasets	**FIH**.Zn(II).NOG TRPV3 Ank1 Peptide 2, co-crystallised	**FIH**.Zn(II).NOG TRPA1 Ank8 Peptide 2, co-crystallilsed+soaked
Data Collection		
Beamline (Wavelength, Å)	DLS I03 (0.9793)	DLS I04-1 (0.9282)
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2 3d (XDS)	Xia2 3dii (XDS)
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	86.4, 86.4, 149.6	87.7, 87.7, 151.7
α,β,γ(°)	90, 90, 90	90, 90, 90
No. reflections	50768 (2483)*	27781 (1358)*
Resolution (Å)	38.63-1.83 (1.86-1.83)*	31.01-2.28 (2.32-2.28)*
R_{merge}	0.108 (2.928)*	0.091 (2.058)*
I/σI	13.7 (1.2)*	17.5 (1.6)*
CC1/2	1.0 (0.5)*	1.0 (0.6)*
Completeness (%)	100 (100)*	100 (100)*
Multiplicity	19.4 (20.1)*	16.7 (18.2)*
Wilson B value (Å^2)	41.7	60.3
Refinement	PHENIX	PHENIX
R_{work}/R_{free}†	0.1701/0.1934	0.1932/0.2164
No. atoms		
-Enzyme (A/B)	2733	2693
-Substrate	146	76
-Ligand	NOG (10)	NOG (10)
-Water	205	76
Avg. B-factors		
Enzyme	53.5	73.9
Substrate	75.8	106.0
Ligand	41.6	57.5
Water	68.3	65.5
R.m.s deviations		
-Bond lengths (Å)	0.012	0.003
-Bond angles (°)	1.027	0.2164

* Highest resolution shell in parentheses.
†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

4.3.3 Crystal structures of FIH in complex with TRPA1$_{144-171}$ and HIF-2α

The first FIH substrate complex structurally characterised was for the FIH.HIF-1α complex[24]. Following from there, several complexes with short peptides derived from natural and synthetic ankyrins have been reported[11,28-30] (see chapter 1). Notably, no structure of FIH in complex with HIF-2α has been reported to date. In the light of differences in the ankyrin binding mode observed in crystal structures of FIH in complex with TRPA1 and TRPV3 peptides, it was tested whether the HIF-α transcription factors share a conserved binding mode to FIH. A HIF-2α derived 35mer peptide (see also Tarhonskaya et al[26]) was co-crystallised with FIH, NOG and zinc as previously described, and the structure solved (Figure 7). The overall binding mode of HIF-1α and HIF-2α to FIH is conserved with near identical active site interactions (Figure 7). Notable differences include the side chain orientations of phenylalanine-798 of HIF-1α and the corresponding phenylalanine-842 of HIF-2α within the site one binding site. These are most likely due to crystal to crystal variation (this is further supported by additional crystallographic data on FIH.HIF-1α complexes showing flexibility of pheylalanine-798, data not shown) but the implications for this are further discussed chapter 5.

Both, HIF-1α and HIF-2α, manifest in complex with FIH only partial electron density for their site two binding to FIH. A proline residue +2 compared to the hydroxylation site on HIF-1α and HIF-2α “bends” the peptide chain away from the enzyme predisposing the substrate to form a flexible loop with an additional binding site at the hydrophobic pocket in site two (Figure 7). Notably, TRPA1$_{310-339}$ also shows only little density for its residues binding in site two of FIH (see above) but does not contain a proline reside in the +2 position of its hydroxylation site.



Figure 6 Sequence alignment of HIF-α transcription factors and potential FIH-ankyrin substrates.

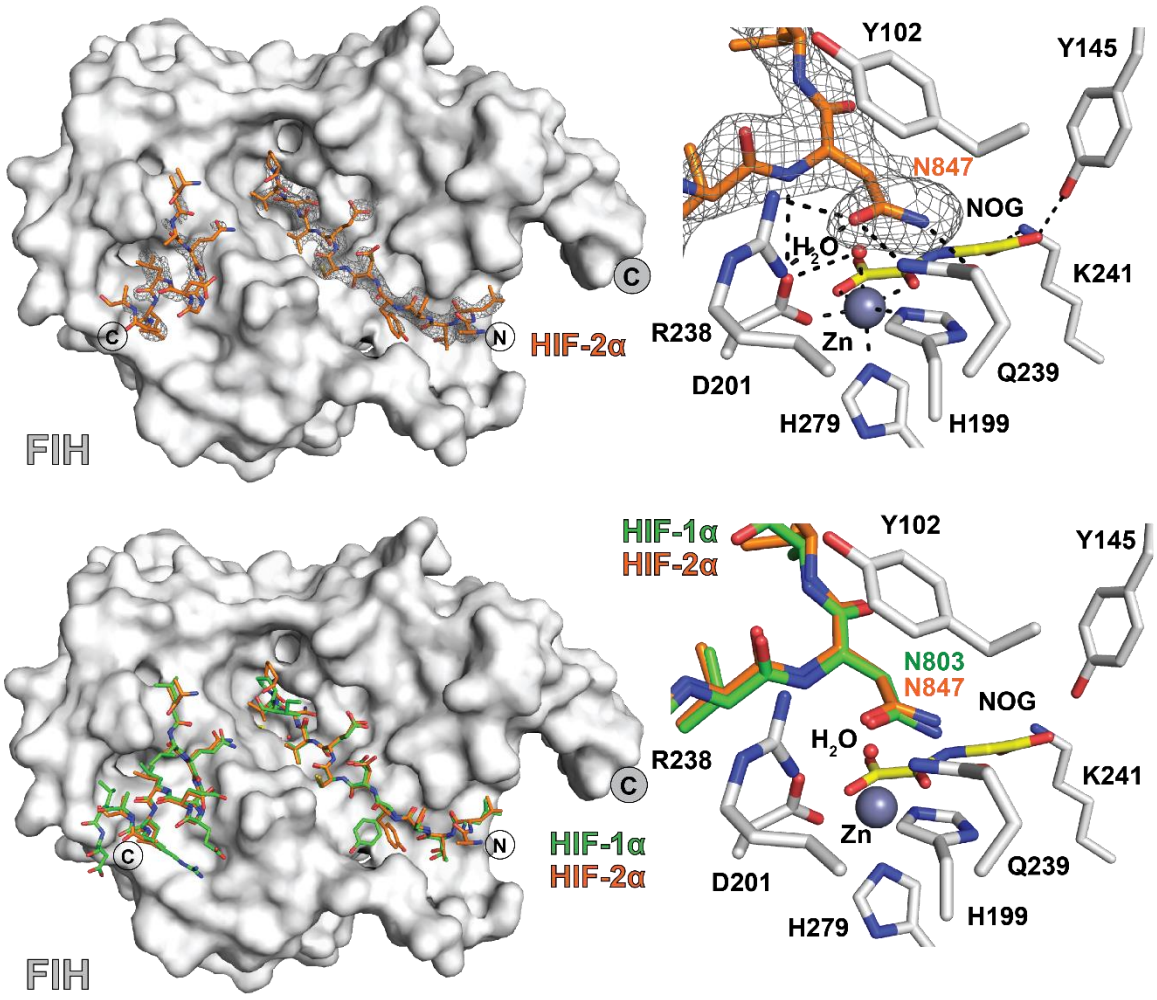


Figure 7 Views from a crystal structure of FIH in complex with HIF-2α. Top View from a crystal structure of FIH in complex with HIF-2α illustrates overall binding mode of the 35mer peptide and active site interactions. The mFo-Fc OMIT map is contoured to 3σ around the peptide. **Bottom:** Comparison of the binding mode of HIF-1α and HIF-2α binding to FIH regarding the overall binding and the active site interactions. FIH.HIF-2α is overlayed with a structure of FIH.HIF-1α (2.15 Å, PDB code 1H2K[24]).

Further, preliminary data suggests that TRPA1 ankyrin-3 can interact with FIH (Dr Richard Hopkinson and Mr Benjamin Saward, University of Oxford). At the peptide level however, no hydroxylation of TRPA1 ankyrin-3 was observed (Figure 2). To investigate the potential binding mode of TRPA ankyrin-3, a peptide corresponding to $TRPA1_{144\text{-}171}$ was synthesised (Dr Richard Hopkinson and Mr Benjamin Saward, University of Oxford) and crystallisation trials were performed as previously described. Surprisingly, no crystals in the FIH-characteristic tetragonal crystal form were observed (see also chapter 8) but rather a hexagonal crystal morphology was observed (data not shown). Processing the data and solving the structure as previously described (see above) confirmed the hexagonal crystal form $P6_422$ (3.04 Å, 1 molecule per asymmetric unit, Table 3).

Relatively low resolution (>3 Å) of the crystal structures in this hexagonal form challenges the interpretation of the electron density. Attempts to soak the $TRPA1_{144-171}$ peptide into the higher diffracting tetragonal crystal form, analogously to previously described methods for TRPV3 and TRPA1 ankyrin-8 derived peptides (see above), did not yield any density corresponding to the bound peptide in the crystal state (data not shown).

Solving a crystal structure of FIH in complex with $TRPA1_{144-171}$ from the hexagonal crystal form using molecular replacement (see above) revealed clear density for the peptide substrate bound in the site one and site two (Figure 8). Strikingly, the N-terminus of the peptide extends towards the substrate binding groove of the FIH dimer partner. This is unlike other peptide substrate binding modes observed for FIH-substrate complexes where the N-terminus characteristically folds into a helix pointing away from the -8 leucine binding site (Figure 9). Notably, all N-terminal residues of the peptide are visible in the structure. Further studies with longer peptides may reveal further interactions between the potential substrate and FIH. In the crystal state, the FIH dimer binds two substrates. The two substrates bound to the FIH dimer come into close proximity (~6 Å) at the dimerisation interface (Figure 8B). Notable is the fact that the functional TRPA1 ion channel contains four TRPA1 proteins forming a homo-tetramer (see also Figure 1). The proximity of several ankyrin repeat domains could enable an FIH dimer to bind two TRPA1 ankyrins simultaneously in a cellular context. Interactions between different ankyrins appear likely due to the spatial proximity of the ankyrins.

Overlaying the structures of FIH in complex with TRPA1 ankyrin-3 and FIH in complex with TRPV3 ankyrin-1 do not reveal major conformation changes in the overall protein fold or in the active site (Figure 8). The difference in resolution of the both complexes (3.06 Å for the FIH.TRPA1 ankyrin-3 complex and 1.86 Å for the FIH.TRPV3 complex), however, challenges the exact determination of small conformational changes. Therefore, conclusions about exact binding modes cannot be drawn. Notably, there appears to be no obvious reason why the TRPA1 ankyrin-3 peptide impaired crystal packing in the tetragonal crystal.

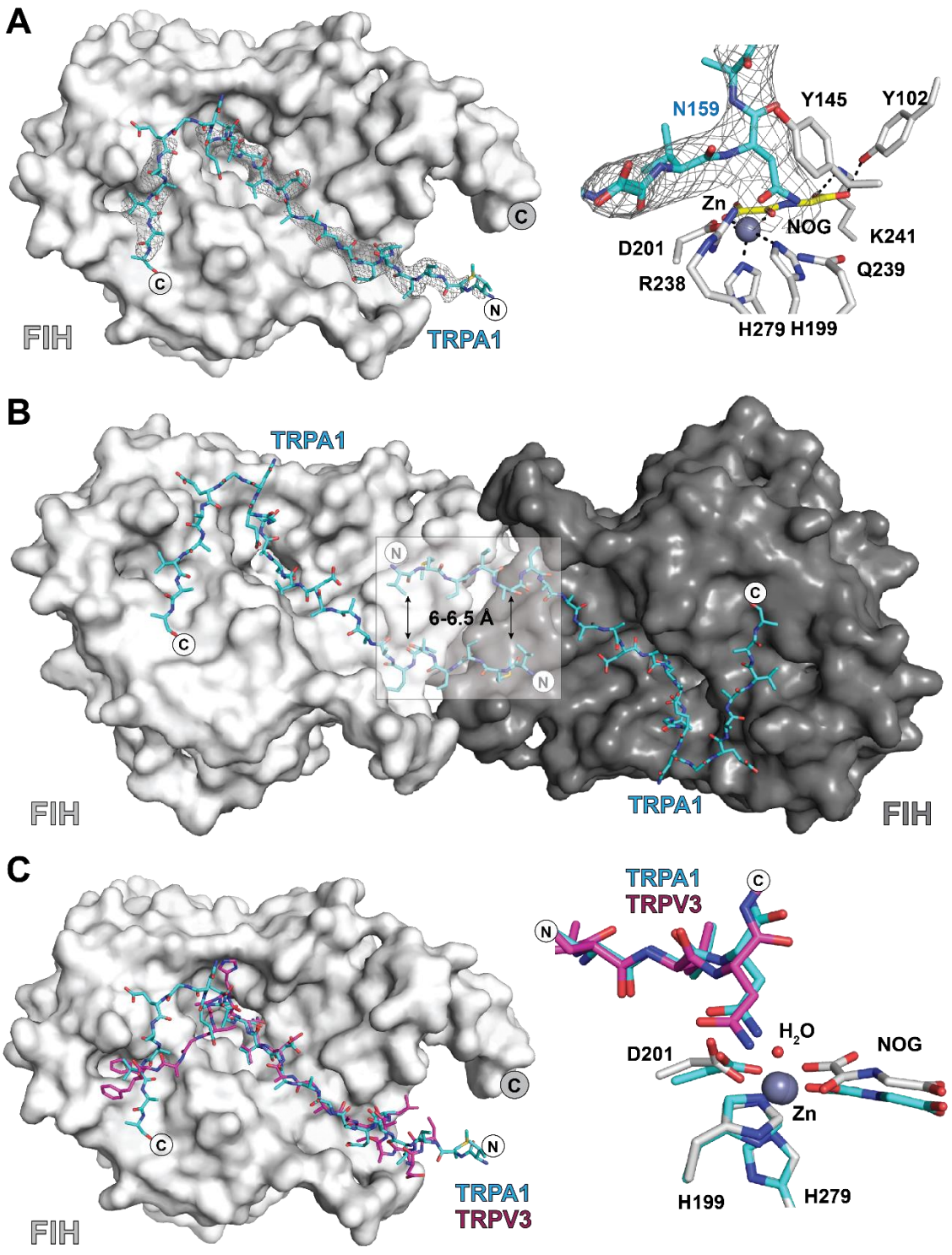


Figure 8 Views from a crystal structure of FIH in complex with a TRPA1 ankyrin-3 derived peptide. A Views from a crystal structure of FIH in complex with $TRPA1_{144-171}$ illustrates overall binding mode of the peptide and active site interactions. The mFo-Fc OMIT map is contoured to 3σ around the peptide. **B** View from a crystal structure of FIH in complex with TRPA1 ankyrin-3 illustrates the substrate binding to an FIH dimer and potential substrate-substrate interactions. **C** Comparison of the binding mode of peptides derived from $TRPA1_{144-171}$ and $TRPV3_{220-244}$ binding to FIH regarding the overall binding and the active site interactions.

Table 3 Data collection and refinement statistics for FIH-HIF-2α and FIH TRPA1-Ank3 complexes.

Datasets	**FIH**.Zn(II).NOG HIF-2α 35mer	**FIH**.Zn(II).NOG TRPA1 Ank3
Data Collection		
Beamline (Wavelength, Å)	DLS I02 (0.9800)	ESRF ID30A-1/MASSIF-1 (0.9600)
Detector	PILATUS 6M-F	PILATUS3 2M
Data Processing	Xia2 3dii (XDS)	MOSFLM/SCALA
Space group	$P4_12_12$	$P6_422$
Cell dimensions		
a,*b*,*c* (Å)	86.3, 86.3, 148.7	129.6, 129.6, 243.4
α,β,γ(°)	90, 90, 90	90, 90, 120
No. reflections	37696 (2732)*	24134 (3430)*
Resolution (Å)	61.04-2.02 (2.07-2.02)*	82.52-3.04 (3.20-3.04)*
R_{sym} or R_{merge}**	0.129 (1.902)*	0.120 (0.898)*
I/σI	12.0 (1.7)*	8.0 (1.3)*
CC1/2	1.0 (0.6)*	0.997 (0.887)*
Completeness (%)	100 (100)*	99.8 (100)*
Multiplicity	14.8 (15.0)*	6.5 (6.9)*
Wilson B value ($Å^2$)	42.5	92.1
Refinement	PHENIX	PHENIX
R_{work}/R_{free}†	0.1838/0.2042	0.2795/0.2993
No. atoms		
-Enzyme (A/B)	2701	2510
-Substrate	157	161
-Ligand	NOG (10)	NOG (10)
-Water	118	---
Avg. B-factors		
Enzyme	57.9	127.2
Substrate	96.0	147.6
Ligand	45.2	182.5
Water	57.9	---
R.m.s deviations		
-Bond lengths (Å)	0.003	0.003
-Bond angles (°)	0.607	0.593

* Highest resolution shell in parentheses.
†$R_{factor} = \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

4.4 Comparison of different ankyrin binding modes to FIH

Solving crystal structures of FIH-ankyrin complexes enables the comparison of different substrate binding modes. In particular, differences observed for different substrates in site two are of interest, as these differences could be exploited to develop substrate-specific inhibitors. Differences in the binding site in direct proximity to the active site are discussed in chapter 3.

An overview of differential binding modes of ankyrins to FIH is shown in Figure 9. This includes a cross-linked structure of mouse mNotch1 to FIH (see chapter 5). The cross-linking site is located in the site one binding site of FIH, which is assumed to have no/little effect on the site two binding (chapter 5).

The two HIF-α transcription factors show near identical binding modes to FIH with limited local differences, most likely resulting from crystal to crystal differences rather than differences in binding modes. The mNotch1 construct shows similar binding to the TRPA1 ankyrin-3 peptide. This could result from a similar distance between the hydroxylation site and the hydrophobic residue (+10 to the hydroxylation site) interacting with the hydrophobic pocket of the FIH site two binding site.

TRPV3 ankyrin-1 shows a distinct binding mode from mNotch1 and TRPA1 ankyrin-3. The hydrophobic residues interacting with the hydrophobic pocket are located +7 and +8 to the hydroxylation site. No hydrophobic residues are found in the +10 position which could interact with this pocket in a manner comparable to mNotch1. This could cause the observed “stretch” in the peptide binding. Systematic studies with TRP derived peptides are needed to confirm the importance of certain residues for distinct binding modes.

Notably, a crystal structure of 3CA cross-linked to FIH did not show any density for the substrate in site 2 (see also chapter 5). Systematic studies using longer peptides are needed to investigate the potential interaction of a consensus ankyrin with FIH site two.

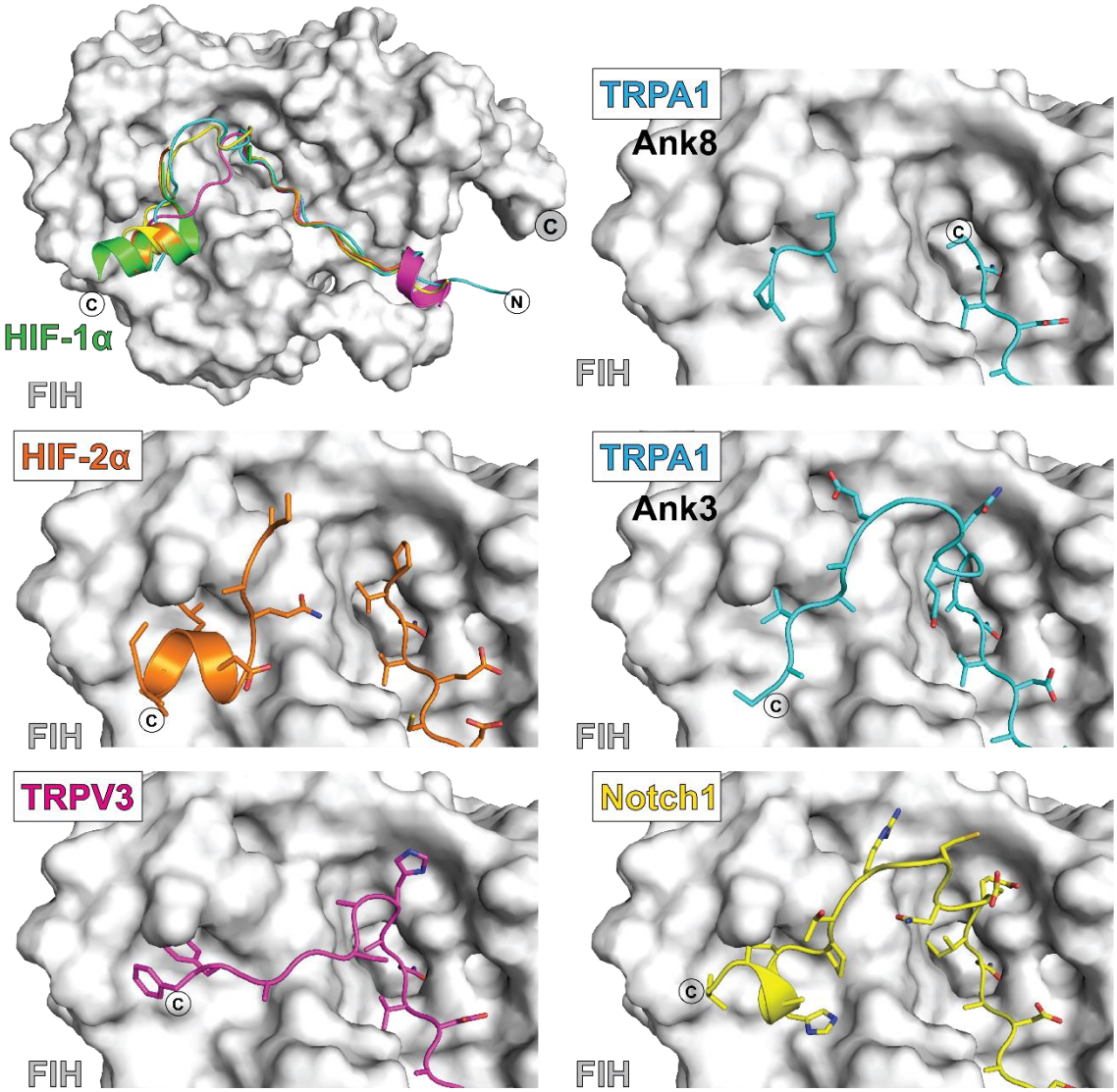


Figure 9 Differential FIH-Ankyrin binding. Views from crystal structures of FIH in complex with HIF-1/2α and several ankyrin substrates. Note that the site two of TRPA1 ankyrin-8 could not be determined unambiguously and were modeled as alanine to illustrate the peptide backbone. Density clearly indicates larger hydrophobic residues, but the identity of the residues is unclear. Notch1 is cross-linked to FIH (see chapter 5).

4.5 Conclusions and outlook

In summary, 5 crystal structures of FIH in complex with peptides derived from TRP channels and one structure of FIH in complex with a HIF-2α derived peptide were solved. To date, no crystal structure of FIH in complex with any ankyrin substrate derived peptides covering site two has been published.

Comparison of the FIH.HIF-2α complex and the FIH.TRP channel structures with a crystal structure of FIH in complex with HIF-1α and the crystal structure of the cross-linked FIH mouse Notch1 complex reveal differences in the binding of the different substrates. Binding of the substrate residues C-terminal of the hydroxylated residue (site two), varies between different substrates substantially. This could be of potential interest for the development of substrate specific FIH inhibitors. Variations in substrate binding close to the active site are of particular interest.

The TRP channel-derived peptides with elongated N-termini are accepted by FIH as substrates but could not be co-crystallised with FIH. Crystallisation trials of these peptides with FIH truncations, preferably crystallising in different crystal forms, might shed further light into additional FIH-substrate interactions.

Although the crystal structure of FIH in complex with $TRPA1_{144-171}$ is of comparably low resolution, the overall interaction site could be determined. In contrast to the other structures, the peptide interacts with the dimer partner over the -8 leucine binding pocket. Further studies with longer peptides could elucidate this distinct binding mode further.

4.6 References

1 Yu, F. H., Yarov-Yarovoy, V., Gutman, G. A. & Catterall, W. A. Overview of molecular relationships in the voltage-gated ion channel superfamily. *Pharmacol Rev* **57**, 387-395, doi:10.1124/pr.57.4.13 (2005).

2 Schüler, A. *et al.* The Rise and Fall of TRP-N, an Ancient Family of Mechanogated Ion Channels, in Metazoa. *Genome Biol Evol* **7**, 1713-1727, doi:10.1093/gbe/evv091 (2015).

3 Voolstra, O. & Huber, A. Post-Translational Modifications of TRP Channels. *Cells* **3**, 258-287, doi:10.3390/cells3020258 (2014).

4 Andersson, D. A., Chase, H. W. & Bevan, S. TRPM8 activation by menthol, icilin, and cold is differentially modulated by intracellular pH. *J Neurosci* **24**, 5364-5369, doi:10.1523/JNEUROSCI.0890-04.2004 (2004).

5 McKemy, D. D., Neuhausser, W. M. & Julius, D. Identification of a cold receptor reveals a general role for TRP channels in thermosensation. *Nature* **416**, 52-58, doi:10.1038/nature719 (2002).

6 Peier, A. M. *et al.* A TRP channel that senses cold stimuli and menthol. *Cell* **108**, 705-715 (2002).

7 Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389**, 816-824, doi:10.1038/39807 (1997).

8 Caterina, M. J., Rosen, T. A., Tominaga, M., Brake, A. J. & Julius, D. A capsaicin-receptor homologue with a high threshold for noxious heat. *Nature* **398**, 436-441, doi:10.1038/18906 (1999).

9 Venkatachalam, K. & Montell, C. TRP channels. *Annu Rev Biochem* **76**, 387-417, doi:10.1146/annurev.biochem.75.103004.142819 (2007).

10 Gees, M., Owsianik, G., Nilius, B. & Voets, T. TRP channels. *Compr Physiol* **2**, 563-608, doi:10.1002/cphy.c110026 (2012).

11 Yang, M. *et al.* Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidinyl residues within ankyrin repeat domains. *FEBS J* **278**, 1086-1097, doi:10.1111/j.1742-4658.2011.08022.x (2011).

12 Karttunen, S. *et al.* Oxygen-dependent hydroxylation by Factor Inhibiting HIF (FIH) regulates the TRPV3 ion channel. *J Cell Sci*, doi:10.1242/jcs.158451 (2014).

13 Miyake, T. *et al.* Cold sensitivity of TRPA1 is unveiled by the prolyl hydroxylation blockade-induced sensitization to ROS. *Nat Commun* **7**, 12840, doi:10.1038/ncomms12840 (2016).

14 So, K. *et al.* Hypoxia-induced sensitisation of TRPA1 in painful dysesthesia evoked by transient hindlimb ischemia/reperfusion in mice. *Sci Rep* **6**, 23261, doi:10.1038/srep23261 (2016).

15 Takahashi, N. *et al.* TRPA1 underlies a sensing mechanism for O2. *Nat Chem Biol* **7**, 701-711, doi:10.1038/nchembio.640 (2011).

16 Bautista, D. M. *et al.* TRPA1 mediates the inflammatory actions of environmental irritants and proalgesic agents. *Cell* **124**, 1269-1282, doi:10.1016/j.cell.2006.02.023 (2006).

17 Jordt, S. E. *et al.* Mustard oils and cannabinoids excite sensory nerve fibres through the TRP channel ANKTM1. *Nature* **427**, 260-265, doi:10.1038/nature02282 (2004).

18 Bandell, M. *et al.* Noxious cold ion channel TRPA1 is activated by pungent compounds and bradykinin. *Neuron* **41**, 849-857 (2004).

19 Moparthi, L. *et al.* Human TRPA1 is intrinsically cold- and chemosensitive with and without its N-terminal ankyrin repeat domain. *Proc Natl Acad Sci U S A* **111**, 16901-16906, doi:10.1073/pnas.1412689111 (2014).

20 Kremeyer, B. *et al.* A gain-of-function mutation in TRPA1 causes familial episodic pain syndrome. *Neuron* **66**, 671-680, doi:10.1016/j.neuron.2010.04.030 (2010).

21 Paulsen, C. E., Armache, J. P., Gao, Y., Cheng, Y. & Julius, D. Structure of the TRPA1 ion channel suggests regulatory mechanisms. *Nature* **520**, 511-517, doi:10.1038/nature14367 (2015).
22 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-2948, doi:10.1093/bioinformatics/btm404 (2007).
23 The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).
24 Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806, doi:10.1074/jbc.C200644200 (2003).
25 Elkins, J. M. *et al.* Oligomeric structure of proclavaminic acid amidino hydrolase: evolution of a hydrolytic enzyme in clavulanic acid biosynthesis. *Biochem J* **366**, 423-434, doi:10.1042/BJ20020125 (2002).
26 Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-Inducible Factor (FIH) as an Oxygen Sensor. *J Biol Chem*, doi:10.1074/jbc.M115.653014 (2015).
27 Wilkins, S. E. *et al.* Differences in hydroxylation and binding of Notch and HIF-1alpha demonstrate substrate selectivity for factor inhibiting HIF-1 (FIH-1). *Int J Biochem Cell Biol* **41**, 1563-1571, doi:10.1016/j.biocel.2009.01.005 (2009).
28 Coleman, M. L. *et al.* Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038, doi:10.1074/jbc.M704102200 (2007).
29 Yang, M. *et al.* Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. *J Biol Chem* **286**, 7648-7660, doi:10.1074/jbc.M110.193540 (2011).
30 Yang, M. *et al.* Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl* **52**, 1700-1704, doi:10.1002/anie.201208046 (2013).

5 Towards an FIH-ankyrin protein co-crystal structure

This chapter describes the work conducted towards obtaining a crystal structure of an FIH-ARD protein-protein complex. Crystallisation trials of FIH WT with intact ARD protein constructs, ASPP2, iASPP and 3CA, led to a crystal structure of the synthetic 3CA alone, but not in complex with FIH. N-terminal truncations of FIH based on crystal structures of FIH and sequence alignments of FIH from different organism were designed, cloned, expressed and crystallised. The FIH truncations crystallised in new crystal forms showed promising characteristics in terms of resolution and electron density map precision. Finally, new methodology for covalently and non-covalent cross-linking of FIH substrate complexes is presented.

5.1 Introduction

The binding mode of intact ARD containing protein constructs to FIH is a long-standing research challenge in Oxford research. The significant conformational changes that the ankyrin repeats have to undergo to bind to FIH are not fully understood[1,2]. Many intrinsically disordered proteins fold upon binding to protein partners[3-6]. The case of substrate unfolding upon binding, however, is rare with only a few examples reported[7]. Whether FIH can induce the unfolding of ankyrin substrates or only binds to an unfolded sub-population of ankyrins in solution is unknown. Stabilising the ankyrin fold however, makes ankyrin repeats less accessible to FIH mediated hydroxylation[8,9].

Recent work on the binding mode of peptides derived from TRP channels illustrates the distinct binding modes between different substrates (chapter 4). How this relates to an intact ankyrin repeat remains unclear. Further, studies on the ASPP protein family have revealed an unprecedented double hydroxylation of a single ankyrin by FIH in the case of intact ankyrin protein constructs but not in the case of peptide substrates. Most likely additional enzyme-substrate interactions play a key role (chapter 6). Additionally, hydrophobic motifs in the HIF-1α primary structure upstream of the structurally characterised FIH binding site have been reported to be of importance for FIH activity[10]. All of the above point towards the importance of further FIH-substrate interactions beyond the observed binding sites.

To characterise conformational changes induced by FIH-ARD interactions, crystallisation trials towards an intact FIH-ARD complex were pursued.

5.2 Non cross-linked FIH ankyrin crystallisation trials

5.2.1 Initial crystallisation trails of FIH-ARD complexes

For the crystallisation of FIH-ARD complexes, a total of 4 ARD proteins (3CA[8,11], mNotch1[1], iASPP and ASPP2, see chapter 6) were chosen. Initial trials to co-purify FIH with a mouse Notch1 construct were unsuccessful, likely because the complex apparently dissociated during size exclusion chromatography (SEC, Dr Michael McDonough, University of Oxford). Similar observations were made for FIH and 3CA which form, at least at 4 °C, no complex that is stable during SEC. FIH did, however, apparently co-purify during SEC in complex with NOG, Zn(II) and iASPP or ASPP2 (data not shown). Fractions containing the protein complex were collected, pooled and concentrated and crystallisation trials performed using commercial available broad screens, including PEG ION, Crystal SCREEN, PEG Rx (all Hampton Research) and PACT Premier, JCSG+, Proplex, Wizard (all Molecular Dimensions) at 4 °C and 20 °C. No crystals were observed in any of the conditions for both FIH complexes, with iASPP and ASPP2. Similar observations were made when the purified proteins of FIH and iASPP or ASPP2 were added together in the presence of NOG and zinc and used for crystallisation trials without any further purification.

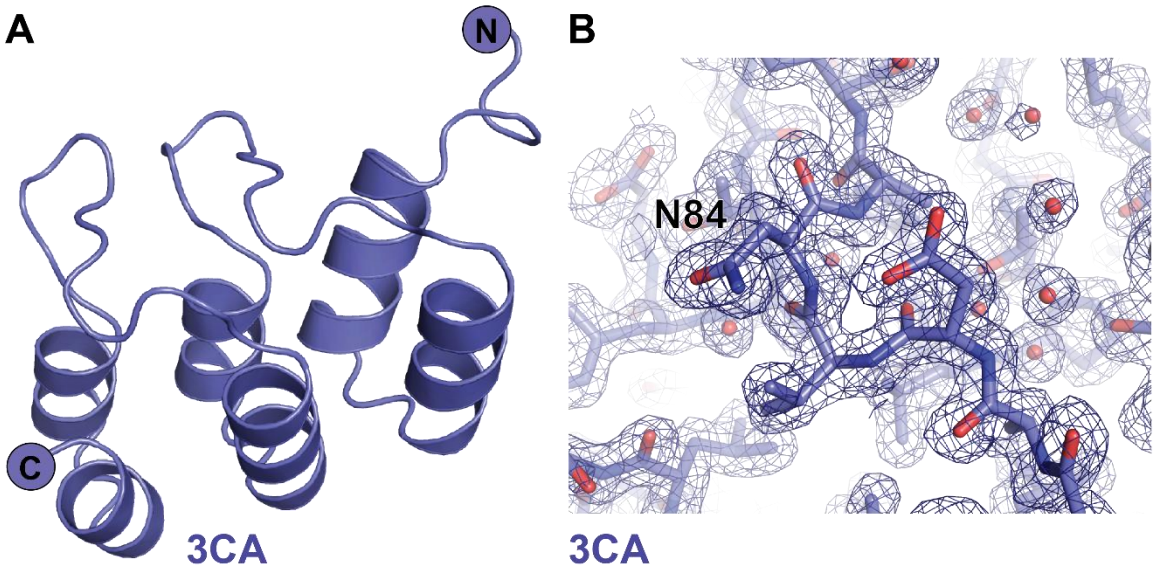


Figure 1 Views from a crystal structure of the synthetic 3 consensus ankyrin protein. A View from the overall fold of the 3CA protein illustrates the ankyrin fold. **B** View from a crystal structure of 3CA showing the FIH target residue asparagine-84 on the third ankyrin repeat; the 2mFo-Fc maps are contoured to 1σ around the protein. Water molecules are displayed as red spheres.

3CA was added to FIH in the presence of NOG and zinc, crystallisation trials were then performed. Crystals were observed in conditions containing sodium acetate and PEG6000. Further optimisation was performed, and optimisation plates designed (best conditions were 0.1 M NaOAc, 0.1-0.4 M $MgCl_2$, 18-23% PEG6000 and 2 mM $ZnCl_2$). In collaboration with Dr Rasheduzzamann Chowdhury, crystals were fished, and data collected.

Attempts to solve the crystal structure of the presumed FIH.3CA complex, revealed that 3CA had crystallised on its own in the presence of FIH (Figure 1). As the crystal structure of 3CA is known (2.0 Å, space group $P3_121$, PDB code 2ZGG[11]) no further optimisation was pursued. Data collection statistics for 3CA are shown in Table 1, refinement is still in progress.

Table 1 Data collection and refinement statistics of 3CA.

Data Collection	
Beamline (Wavelength, Å)	DLS I03 (0.9763)
Detector	Pilatus 6M-F
Data Processing	Xia2 3dii
Space group	$P2_12_12$
Cell dimensions	
a,*b*,*c* (Å)	26.1, 49.7, 54.1
α,β,γ(°)	90,90,90
No. of molecules/ASU	1
No. reflections	10488 (758)*
Resolution (Å)	36.6-1.63 (1.67-1.63)*
R_{merge}	0.176 (0.932)*
I/σI	15.1 (3.5)*
CC ½	
Completeness (%)	100 (100)*
Multiplicity	17.0 (17.3)*
Wilson B value ($Å^2$)	15.0

* Highest resolution shell in parentheses.

5.2.2 Variations of FIH constructs

5.2.2.1 Design of different FIH constructs

As described above, initial attempts to crystallise complexes of FIH WT with ARD proteins were unsuccessful, yielding only a structure of one ARD protein, 3CA, alone. Another strategy for structural determination of 2OG substrate complexes that has been successfully applied to PHDs and to ribosomal oxygenases, is the use of homologues from other organisms for crystallisation trials[12,13]. To date, all studied animals have a functional HIF pathway with at least one HIF isoform and one PHD isoform[14]. FIH is evolutionarily conserved in higher animals but only sporadically present in lower animals. The appearance of FIH would seems to correlate with the presence of a C-terminal transactivation domain of HIF[14]. Sequence alignments of FIH from selected metazoans (selection of organisms modified from Iyer et al[15]) are shown in Figure 2. FIH like enzymes from lower eukaryotes that do not possess a HIF pathway, such as *Monosiga brevicollis*, which could potentially act as an ankyrin hydroxylases, are also of interest.

Distinguishing FIH from other JmjC domain containing proteins, however, becomes more difficult with increasing evolutionary distance. Using characteristic domains like the FIH dimerisation domain for distinguishing FIH from *e.g*. JmjD6 or JmjD7 is used, but heavily relies on accurate bioinformatic prediction of the expressed proteins sequence and protein fold. Note that the C-terminal dimerisation domain is also closely related to the dimerisation domain of TYW5, a JmjC nucleic acid hydroxylase[16]. Attempts to sub-clone FIH from zebrafish and FIH like enzymes from *Monosiga* and *Drosophila melanogaster* are ongoing.

Comparison of the sequence of FIH from several organisms (Figure 2) shows highly conserved JmjC and C-terminal dimerisation domains. Some variations are observed in the N-terminus. Higher organisms appear to have elongated N-termini compared to lower organisms (Figure 2). At least in the crystal state, the N-terminus of FIH is not involved in substrate binding (Figure 3, see also chapter 4). In some structures the N-terminus of other molecules in the crystal lattice can occupy the site two substrate binding site of FIH (Figure 3, see also chapter 8) and thereby potentially interfere with substrate binding. Close analysis of the crystal packing reveals that the N-terminus is important for crystal packing in the tetragonal crystal form of FIH. Truncations of the N-terminus could potentially force FIH to crystallise in a different space group. In most reported crystal structures of FIH and crystal structures solved in the scope of this thesis, the N-terminus of FIH is disordered, thereby enhancing background scattering and lowering the overall resolution of the dataset.

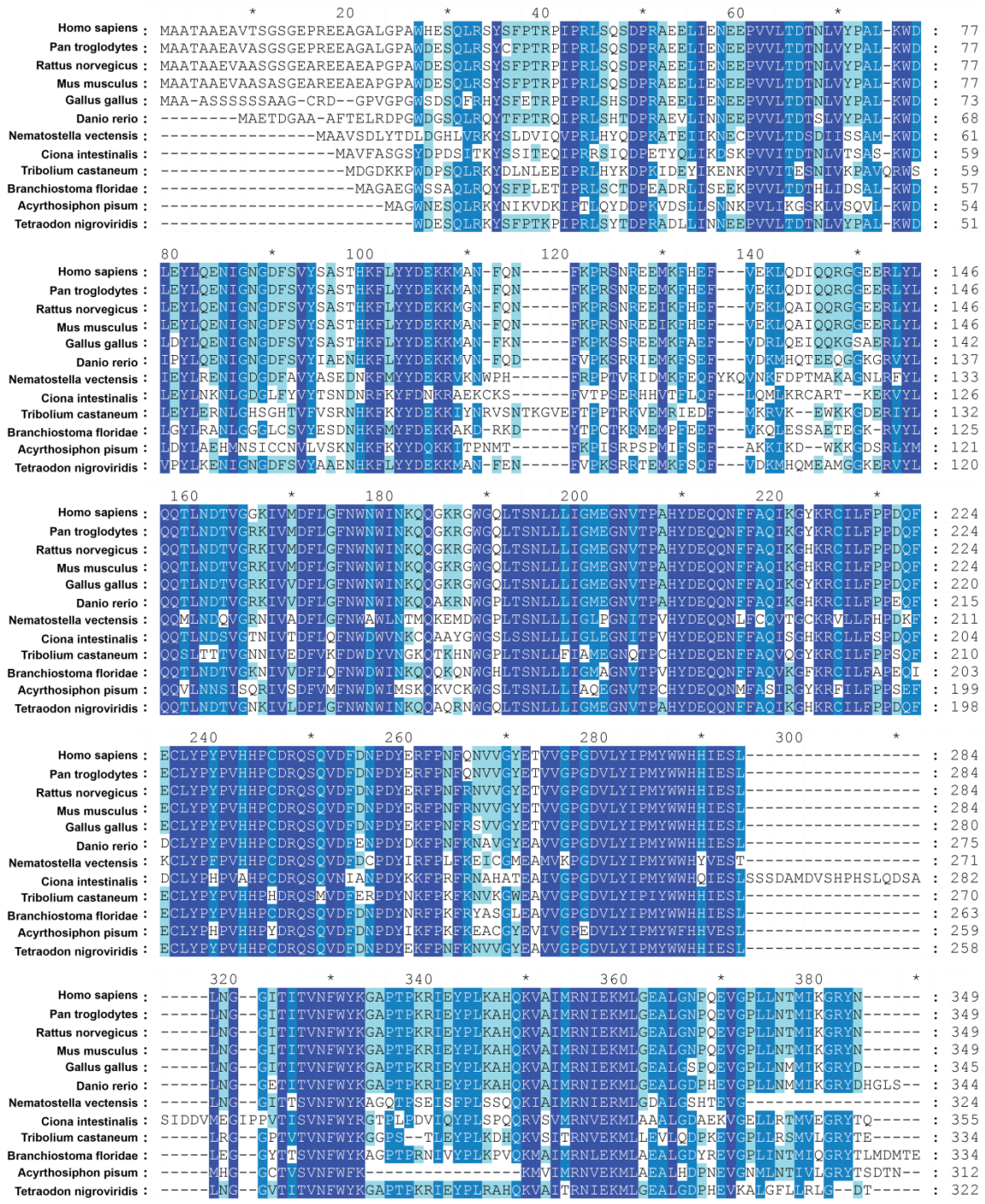


Figure 2 Sequence alignment of FIH from selected metazoans. Sequence alignment of FIH from humans, chimpanzee, rat, mouse, red junglefowl, zebrafish, starlet sea anemone, red flour beetle, Florida lancelet, pea aphid, sea squirt and pufferfish. Sequences as reported in UniProt[17]. Alignment modified from Iyer et al[15].

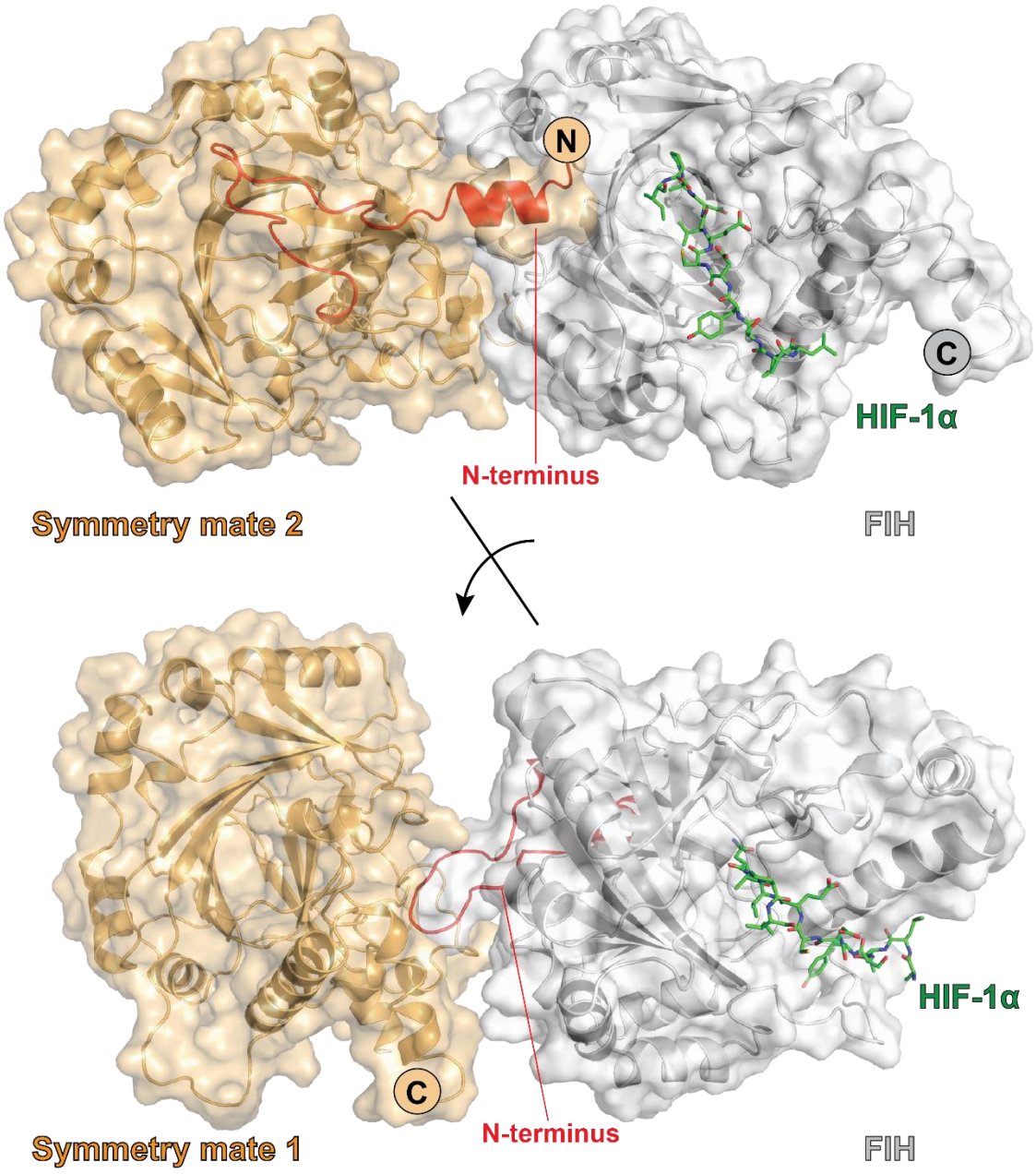


Figure 3 Views from a crystal structure of FIH in complex with HIF-1α illustrate crystal contacts. Views from a crystal structure of FIH in complex with a HIF-1α 19mer peptide (1.90 Å, chapter 8) illustrating interaction with the N-terminus of FIH within the crystal lattice.

The truncation of the disordered N-terminal residues of FIH may improve resolution by reducing background scattering. Further, breaking the tetragonal crystal form may open site two to binding site by reordering symmetry related proteins in the packing, which may presently block additional interactions of the enzyme-substrate complex. Truncating the N-terminus of FIH will also remove residues binding in the substrate binding groove of other units in the crystal, which potentially interfere with substrate binding. Based on FIH crystal structures, two N-terminally truncated FIH constructs were designed: $FIH_{27\text{-}349}$ and $FIH_{37\text{-}349}$.

5.2.2.2 Cloning and expression of FIH_{26-349} and FIH_{37-349}

FIH_{26-349} and FIH_{37-349} were cloned into a modified pET28a vector generated for this project where the thrombin cleavage site for the N-terminal His_6-tag was replaced by a TEV cleavages site (for experimental details see chapter 10).

Both truncations were found to be produced under similar conditions to those of FIH WT. Proteins were purified using a His-trap column (GE Healthcare) followed by an overnight tag cleavage and size exclusion chromatography (Figure 4). FIH_{26-349} shows a single peak on the SDS PAGE and denaturing MS, indicating pure protein. FIH_{37-349} shows a double band on the SDS PAGE and two species in the denaturing MS corresponding to the protein with cleaved and with non-cleaved His_6-tag, indicating incomplete tag-cleavage. The activity of the proteins was confirmed using MALDI-MS prior to crystallisation trials (data not shown).

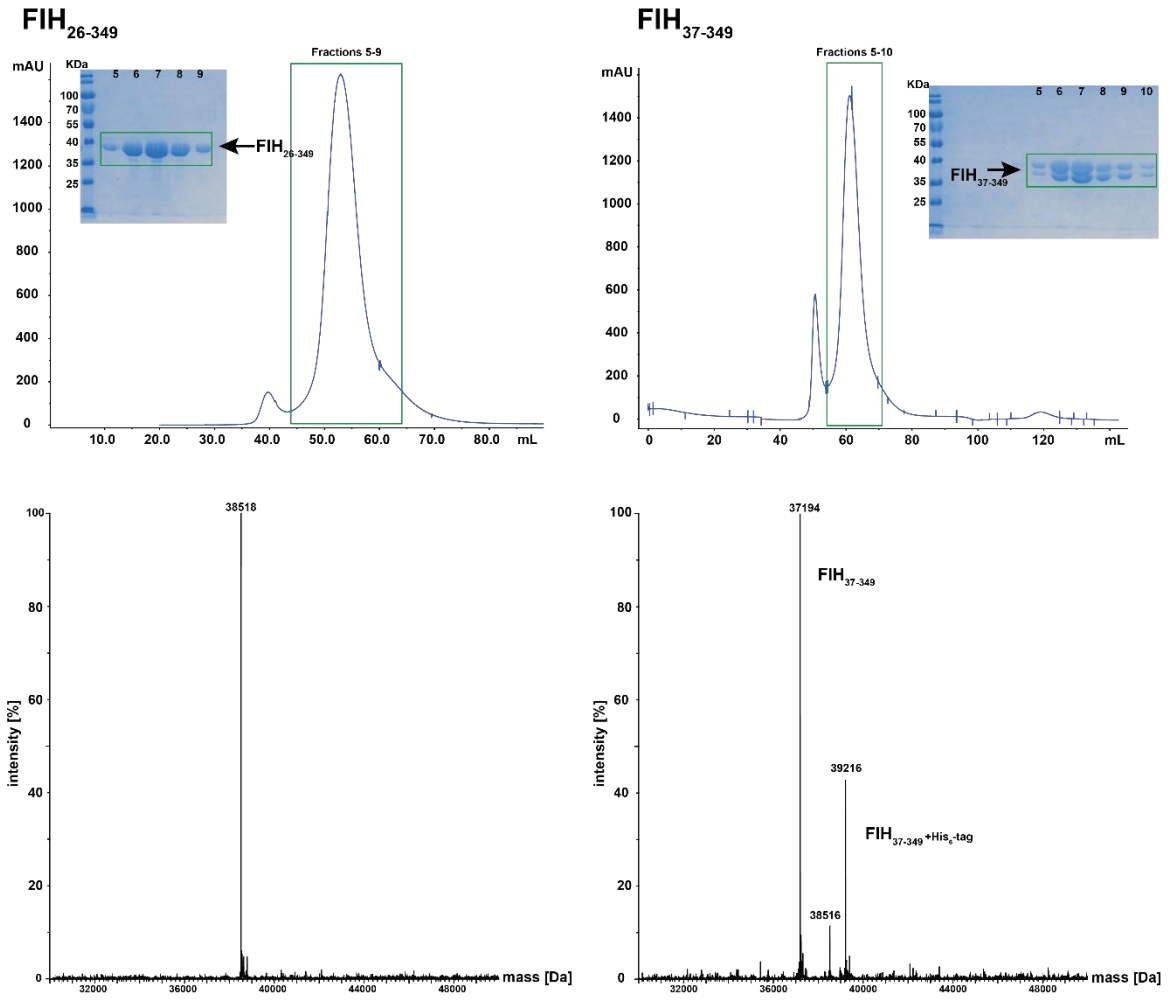


Figure 4 Purification of FIH N-terminal truncations. Purification of FIH_{26-349} and FIH_{37-349}. Top: Size exclusion chromatogram and SDS PAGE for the purified proteins. Bottom: Denaturing MS of the purified protein.

5.2.2.3 Crystallisation trials of FIH_{26-349} and FIH_{37-349}

Following the purification of the N-terminal FIH truncations and activity tests, crystallisation trials were pursued. For the crystallisation trails of FIH WT, the enzyme is incubated with NOG and zinc prior to crystallisation trials. FIH_{26-349} and FIH_{37-349} were similarly pre-incubated with zinc and NOG. FIH_{37-349} precipitated when zinc was added, and no crystallisation trials could be performed. Switching the metal to manganese overcame precipitation issues FIH_{37-349} and crystallisation trials were performed with manganese(II) substituting for the iron(II).

Both FIH truncations crystallised in several new crystal forms (Table 2 and Table 3). Only FIH_{26-349} crystallised in the $P4_12_12$ space group, whereas no crystals in this form could be observed for crystallisation trials with FIH_{37-349}. Notably, FIH_{37-349} crystallises, amongst other crystal forms, in the tetragonal space group *I*422. In the *I*422 space group FIH diffracts to higher resolution than the tetragonal $P4_12_12$ space group (accumulated data from several crystals, data not shown).

Comparison of the crystal structure of FIH WT and FIH_{37-349} shows that the active site is solvent exposed and that metal and co-factor binding modes are conserved between the two crystal forms (Figure 5). Notably, FIH crystals in the $P4_12_12$ space group have, compared to other crystals at similar resolution, noisy electron density maps. This can challenge the modelling of FIH binding partners, especially small molecule ligands. FIH_{37-349} crystallised in the *I*422 space group shows less noisy maps and lower overall B-factors for the structure, indicating less variation within the crystal lattice. This would make the FIH_{37-349} a suitable candidate for structural studies on FIH-ligand interactions, such as in crystallographic fragment screening (see also chapter 3).

The active site of FIH is in both crystal forms, $P4_12_12$ and *I*422, is solvent exposed, but the accessibility of the substrate binding sites varies. The site two binding site is blocked in the *I*422 space group of FIH_{37-349} which does not allow the binding of longer ankyrin substrates (Figure 5). The N-terminal binding site is, however, more accessible compared to the $P4_12_12$ crystal. This would be of interest for, *e.g.* the TRPA1 and TRPV3 peptides substrate with extended N-termini described in Chapter 4.

Crystallisation trials of FIH_{26-349} and FIH_{37-349} with ASPP2 however, did not yield any crystals for the protein-protein complex.

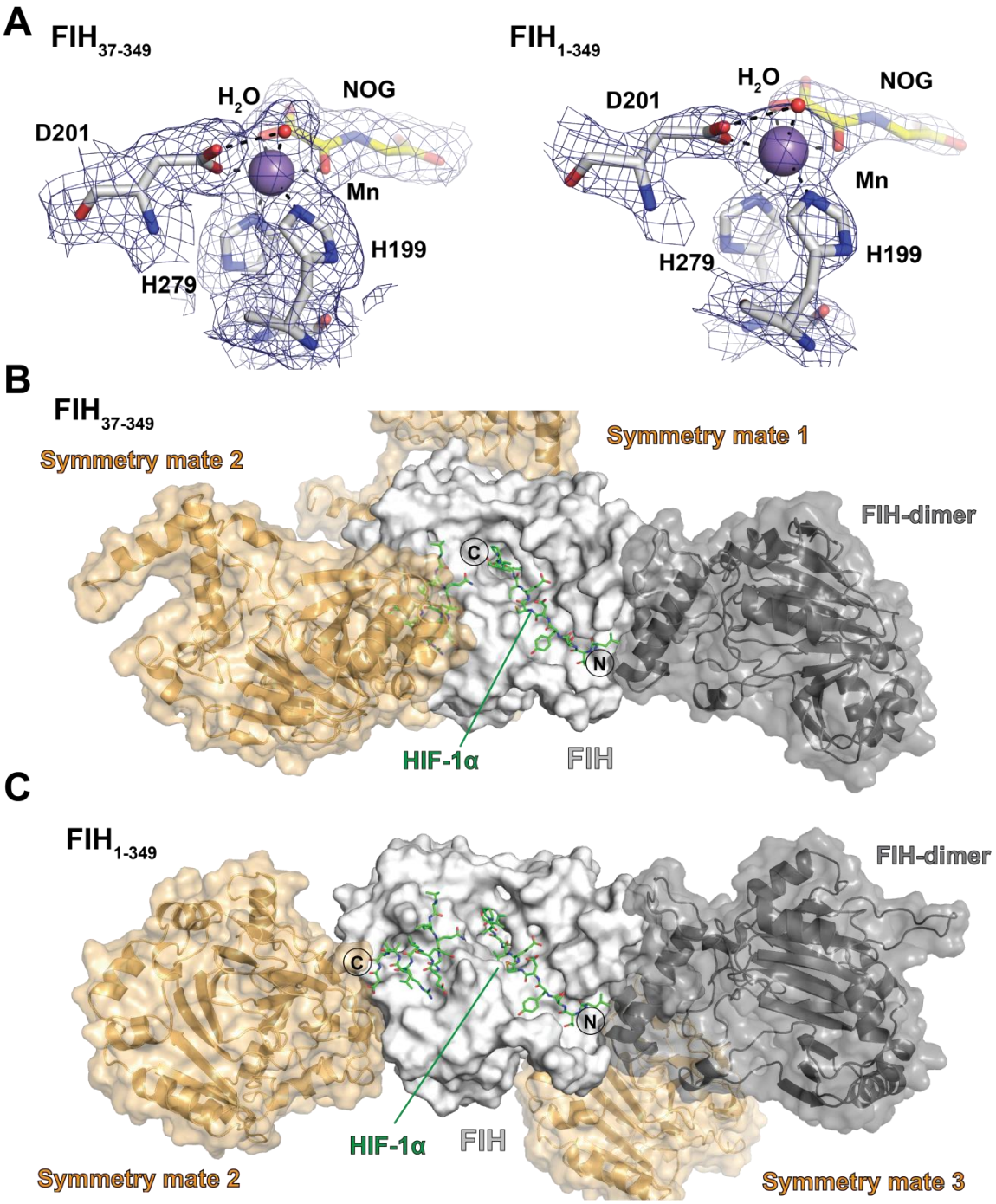


Figure 5 Views from crystal structures of FIH WT and FIH_{37-349}. A Views from a crystal structures of FIH_{37-349} and FIH WT in complex with manganese. 2mFo-Fc maps are contoured to 1σ. **B,C** Views from crystal structures of FIH with neighbouring FIH proteins in the crystal lattice. The dimer partner is coloured in dark grey, further crystal contacts in light orange. The substrate (green) binding site is highlighted by HIF-1α peptide binding modelled form PDB 1H2K[18].

Table 2 Data collection and refinement statistics of FIH_{37-349} and FIH WT in complex with Mn (II) and NOG.

Datasets	**FIH_{37-349}**.Mn(II).NOG crystal form 1	**FIH.**Mn(II).NOG
Data Collection	DLS I03 (0.9795)	DLS I02 (0.9795)
Beamline (Wavelength, Å)		
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	MOSFLM/SCALA	Xia 3dii (XDS)
Space group	*I*422	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	128.7, 128.7, 151.9	86.3, 86.3, 147.5
α,β,γ(°)	90, 90, 90	90, 90, 90
No. reflections	46562 (6691)*	37637 (2717)*
Resolution (Å)	98.16-1.95 (2.06-1.95)*	56.38-2.01 (2.06-2.01)*
R_{merge}	0.151 (3.697)*	0.061 (2.039)*
I/σI	16.2 (1.9)*	20.9 (1.4)*
CC1/2	0.999 (0.637)*	1.0 (0.6)*
Completeness (%)	100 (100)*	99.7 (99.2)*
Multiplicity	24.5 (24.8)*	13.0 (13.3)*
Wilson B value ($Å^2$)	39.4	50.9
Refinement	PHENIX	PHENIX
R_{work}/R_{free}†	0.1811/0.2013	0.198/0.2172
No. atoms		
-Enzyme (A/B)	2613	2727
-Ligand	NOG (10)	NOG (10)
-Water	157	81
Avg. B-factors		
Enzyme	55.6	74.7
Ligand	39.3	54.3
Water	58.6	66.7
R.m.s deviations		
-Bond lengths (Å)	0.010	0.003
-Bond angles (°)	1.079	0.591

* Highest resolution shell in parentheses.
†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

Table 3 Data collection statistics of FIH truncations.

Datasets	**FIH$_{26-349}$**.Zn(II).NOG crystal form 1	**FIH$_{26-349}$**.Zn(II).NOG crystal form 2
Beamline (Wavelength, Å)	In-house (RIGAKU FR-E Superbright) (1.5419)	DLS I03 (0.9795)
Detector	Saturn 944+	PILATUS 6M-F
Data Processing	HKL2000	Xia2 DIALS
Space group	$P4_12_12$	$I222$
Cell dimensions		
a,*b*,*c* (Å)	86.2, 86.2, 147.1	79.4, 142.3, 151.6
α,β,γ(°)	90, 90, 90	90, 90, 90
No. reflections	11613 (1121)*	21323 (1049)*
Resolution (Å)	50.0-3.0 (3.11-3.00)*	70.32-2.81 (2.86-2.81)*
R_{merge}	0.361 (0.992)*	0.200 (1.935)*
I/σI	6.5 (1.3)*	7.0 (1.2)*
CC 1/2	0.998 (0.804)*	1.0 (0.8)*
Completeness (%)	99.6 (99.7)*	100 (100)*
Multiplicity	12.9 (13.9)*	12.8 (13.0)*
Datasets	**FIH$_{26-349}$**.Zn(II).NOG crystal form 3	**FIH$_{37-349}$**.Mn(II).NOG crystal form 2
Beamline (Wavelength, Å)	In-house (RIGAKU FR-E Superbright) (1.5419)	DLS I03 (0.9795)
Detector	Saturn 944+	PILATUS 6M-F
Data Processing	HKL2000	Xia2 DIALS
Space group	$C121$	$I4_132$
Cell dimensions		
a,*b*,*c* (Å)	126.4, 134.5, 79.3	279.5, 579.5, 279.5
α,β,γ(°)	90.0, 96.1, 90.0	90, 90, 90
No. reflections	30802 (3094)*	56922 (2800)*
Resolution (Å)	50.0-2.85 (2.95-2.85)*	139.8-2.60 (2.64-2.60)*
R_{merge}	0.203 (0.890)*	0.256 (3.226)*
I/σI	6.1 (1.5)*	16.2 (2.0)*
CC 1/2	0.995 (0.673)*	1.0 (0.8)*
Completeness (%)	99.3 (99.8)*	100 (100)*
Multiplicity	3.2 (3.1)*	59.2 (59.9)*

* Highest resolution shell in parentheses.

5.2.3 Crystallisation trials using limited proteolysis

Following the crystallisation trials of FIH-ARD proteins with FIH WT and N-terminal truncations of FIH with ARD proteins, which did not yield crystals for the targeted protein-protein complex, further crystallisation strategies were pursued. With large conformational changes within the ankyrin protein likely upon FIH binding[1], it can be assumed that substantial substantial part of the ARD is flexible in the complex; thus may impair crystallisation.

To reduce the flexible loops or termini and to increase homogeneity of a sample, flexible regions can be cleaved[19]. If the flexible regions are unknown, *in situ* proteolysis approaches can be used to trim flexible loops under crystallisation conditions while folded parts of the protein are protected from protease degradation[20,21]. Using a protocol analogous to the "floppy-choppy" *in situ* proteolysis kit (Jena Bioscience) according to the manufacturers' protocol, protease resistance of FIH-ARD complexes were tested (Figure 7) and, following from initial protease resistance tests, crystallisation trials were performed using trypsin as the protease (Figure 6). Trypsin appeared to be the best tolerated protease by FIH, therefore crystallisation trials were performed as previously described with the supplementation of trypsin immediately prior to crystallisation trials. Small needle crystals (Figure 6) appeared within 4 weeks at 4 °C but did not diffract X-rays at a synchrotron (data not shown). Follow-up crystal optimisation attempts did not yield any further crystals.

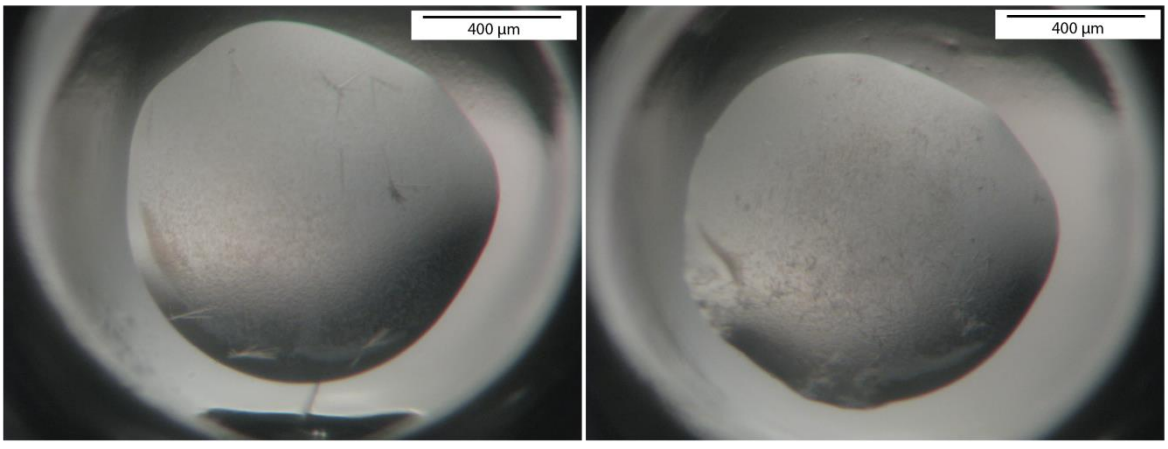


Figure 6 Crystals grown from limited hydrolysis. Crystals grown from FIH co-crystallisation experiments with ASPP2 at 4 °C, crystallisation solutions were supplemented with 1% trypsin immediately before crystallisation trials.

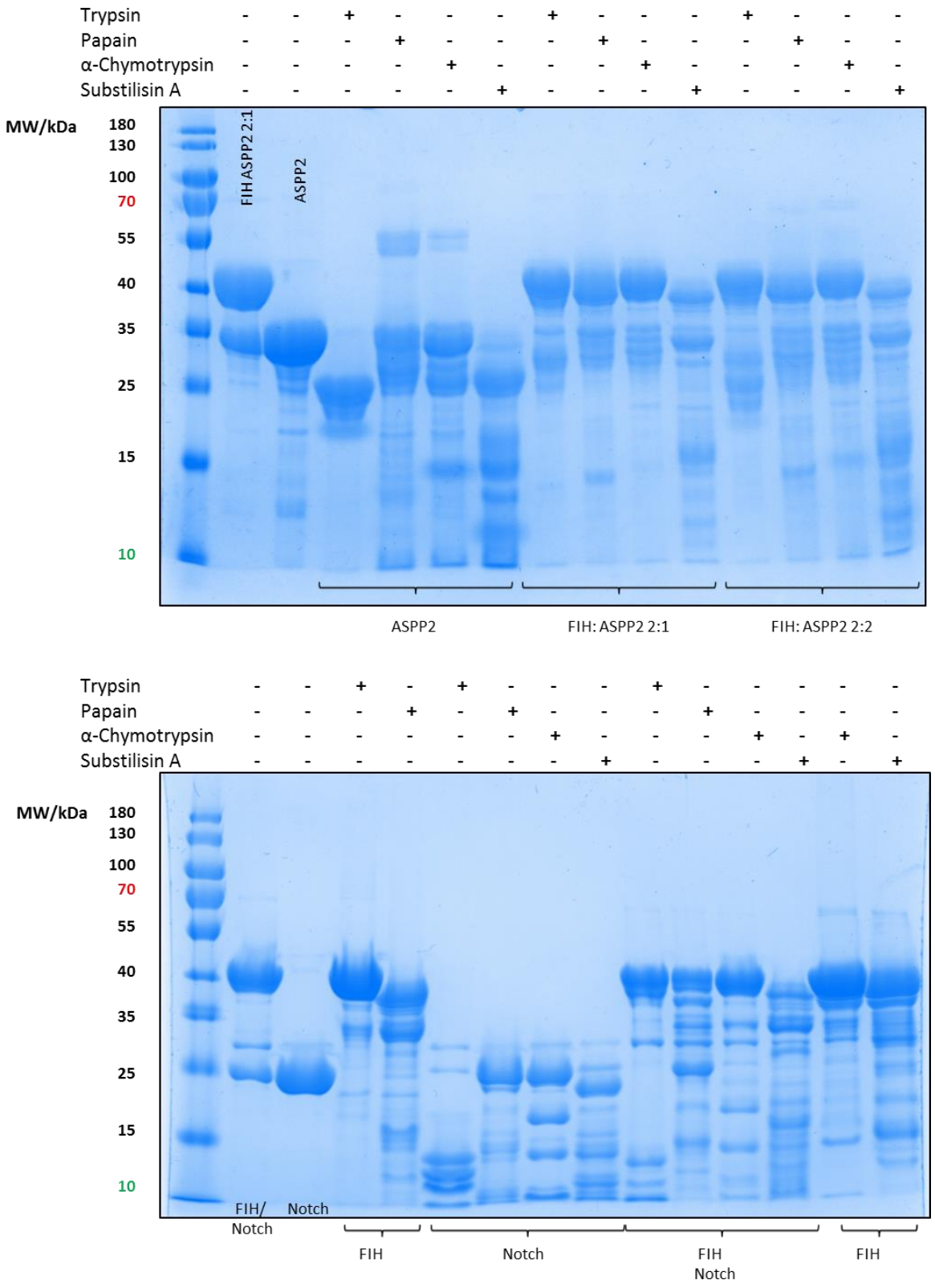


Figure 7 *In situ* protease resistance test of FIH and FIH-ARD complexes. According to "Floppy-Choppy" *in situ* proteolysis kit (Jena Bioscience).

5.3 Enzyme-substrate cross-linking approaches for FIH

5.3.1 Introduction to protein cross-linking

Obtaining well-diffracting crystals for protein-protein complexes is a long standing challenge in X-ray crystallography. The affinity of the substrates, solubility, transient nature of the interaction and heterogeneity of the sample can impair the formation of a homogenous enzyme-substrate complex suitable for crystallisation[22,23]. Many of these challenges can be overcome by stabilising the protein-protein or enzyme-substrate complexes with covalent cross-links.

Covalent cross-linking of low affinity or transient protein complexes is well established for many applications including, ChIP sequencing, proteomic mass spectrometry or structural biology[24-26].

For stabilisation of transient protein-protein interactions and protein-DNA complexes on a proteomic scale, no specific cross-linker is needed to capture a large range of protein complexes. Protein-nucleic acid complexes are routinely cross-linked using UV light or chemically cross linking with formaldehyde, of which the latter is reversible and frequently used in ChIP sequencing[27-30]. Aldehyde based cross-linkers are also frequently used to stabilise protein-protein complexes in structural biology with a preference for glutaraldehyde[23,31,32]. Chemical cross-link reactions using specific amino acids, *e.g*. primary amines or cysteine residues (Figure 8) have the advantage of spatial restrictions of the reactions and are often used in structural mass-spectrometry as they can only occur in complexes where the corresponding residues are within defined proximities[23]. Photo-inducible cross-linkers allow for time-controlled cross-linking (Figure 8).

Many of the protein cross-links described above are non-specific. In some cases, protein-protein complexes can have multiple interaction sites and several conformations (*i.e.* multiple oligomerisation states). This is also the case for the FIH-mNotch1 complex. The mNotch1 protein has two hydroxylation sites targeted by FIH within its ARDs[1]. To stabilise a single complex a site-specific approach is needed.

A well-established approach to stabilise proteins and protein-protein complexes is the introduction of engineered disulfide bonds. Pioneering work on protein stabilisation *via* disulfide bonds was conducted in the 1980s[33-36]. Using site specific disulfide bonds to restrict movement, protein folding, and unfolding dynamics were studied[37-40]. Following the pioneering work on disulfide bond engineering, stabilisation of proteins and protein-protein complexes *via* disulfide bonds was increasingly exploited for crystallography in the 1990s[41-44].

Cross-linking approaches for 2OG oxygenases were first reported for the human nucleotide demethylase Alkylated DNA repair protein AlkB homologue 2 (ABH2)[45]. Notably, a substitution reaction forming the disulfide cross-link between enzyme and substrate was employed (Figure 8) and not cysteine oxidation.

Following from established protein-protein disulfide cross-linking (see above) and enzyme DNA cross-linking procedures, substrate complexes of ribosomal protein oxygenases were reported[13] with, and subsequently without, cross-linking[46], followed by cross-linked PHD2.HIF-α complex structures[47] employing cysteine oxidation under crystallisation conditions (Figure 8).

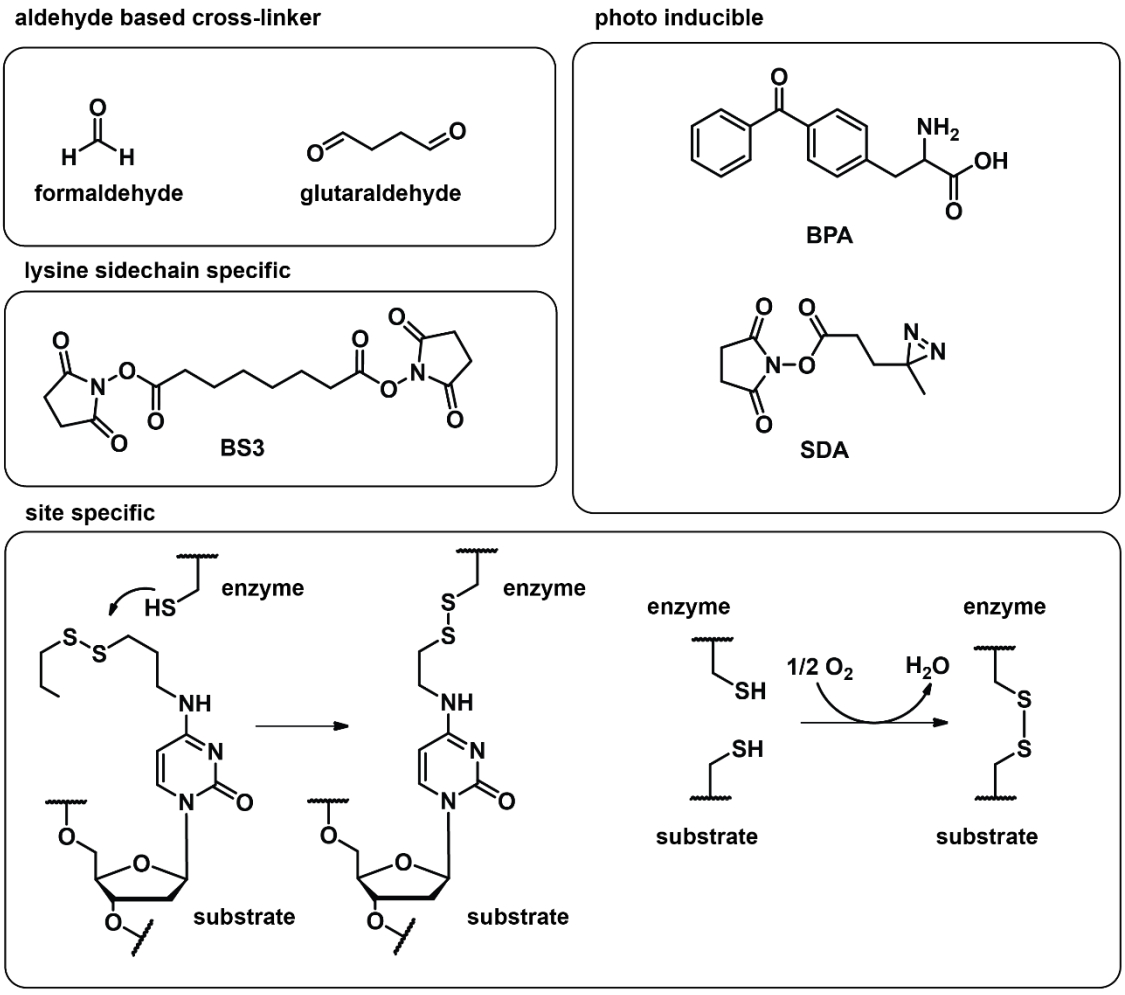


Figure 8 Examples of chemical cross-linking of protein-protein and enzyme-substrate complexes. Examples of aldehyde based cross-linkers are formaldehyde and glutaraldehyde[23,31,32]. Bis-sulfosuccinimidyl (BS3), reacts with primary amino groups[48]. Succinimidyl 4,4'-azipentanoate (SDA) and p-benzoyl-L-phenylalanine (BPA) are photo-inducible[49,50]. **Bottom**: Examples of site specific cross-linking[13,33-35,45,47].

5.3.2 Design of FIH-substrate cross-linking and first peptide structure

Analysis of crystal structures of FIH in complex with its substrates (chapter 1 and chapter 4) reveal a conserved substrate binding mode in site one of FIH and non-conserved substrate binding in site two of FIH (Figure 10). For disulfide cross-linking, the proximity and orientations of the cysteine residues is crucial[36,51,52]. With significant differences in substrate binding in site two of FIH (see chapter 4) potential cross-linking sites of the substrates with FIH in site one of FIH were further investigated.

FIH shows a substrate preference of -L^{-8}LxxGADV**N**- with the asparagine being the hydroxylation site (chapter 1). Comparing sequence alignments of known FIH substrates reveals that positions -5 and -6 to the hydroxylation sites are non-conserved (chapter 1) which allows for some substrate flexibility (Figure 10). This is further supported by alternative side chain conformations observed for phenylalanine-798 and phenylalanine-842 of HIF-1α and HIF-2α respectively at the -5 position of the hydroxylated asparagine. Close to this site is a loop (residues 304-311) which is often observed to be flexible and often not visible in the crystalline state (cumulative data from structural analysis of FIH crystals). The best positioned residue on FIH for potential cross-linking with the -5 residue of FIH is glycine-299 (Figure 10). It was noted that mutating glycine residues introduces more restrains in backbone conformations which was considered to be acceptable due to a flexible loop adjacent to the glycine. Glycine-299 of FIH was mutated to cysteine and the protein expressed and purified as described in chapter 10. Time-course assays (in collaboration with Dr Anthony Tumber, University of Oxford) with 1CA peptides show activity for both, FIH WT and the FIH G299C variant (Figure 9).

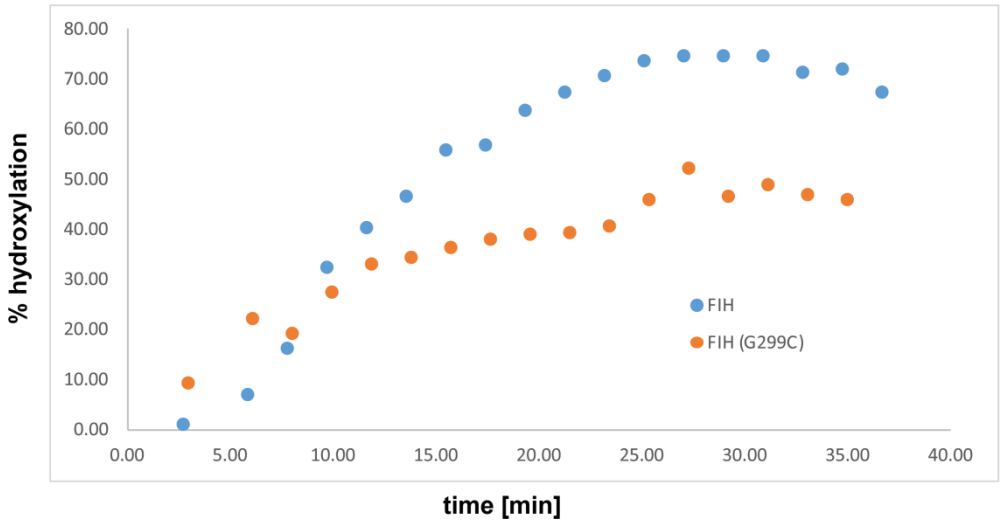


Figure 9 LC-MS based activity assay of FIH WT and FIH G299C. Reaction conditions: 100 nM FIH, 10 µM 1CA 20mer peptide, 10 µM 2OG, 100 µM L-ascorbic acid, 10 µM iron(II) ammonium sulfate, 50 mM Tris-HCl pH=7.5, 50 mM NaCl at ambient temperature.

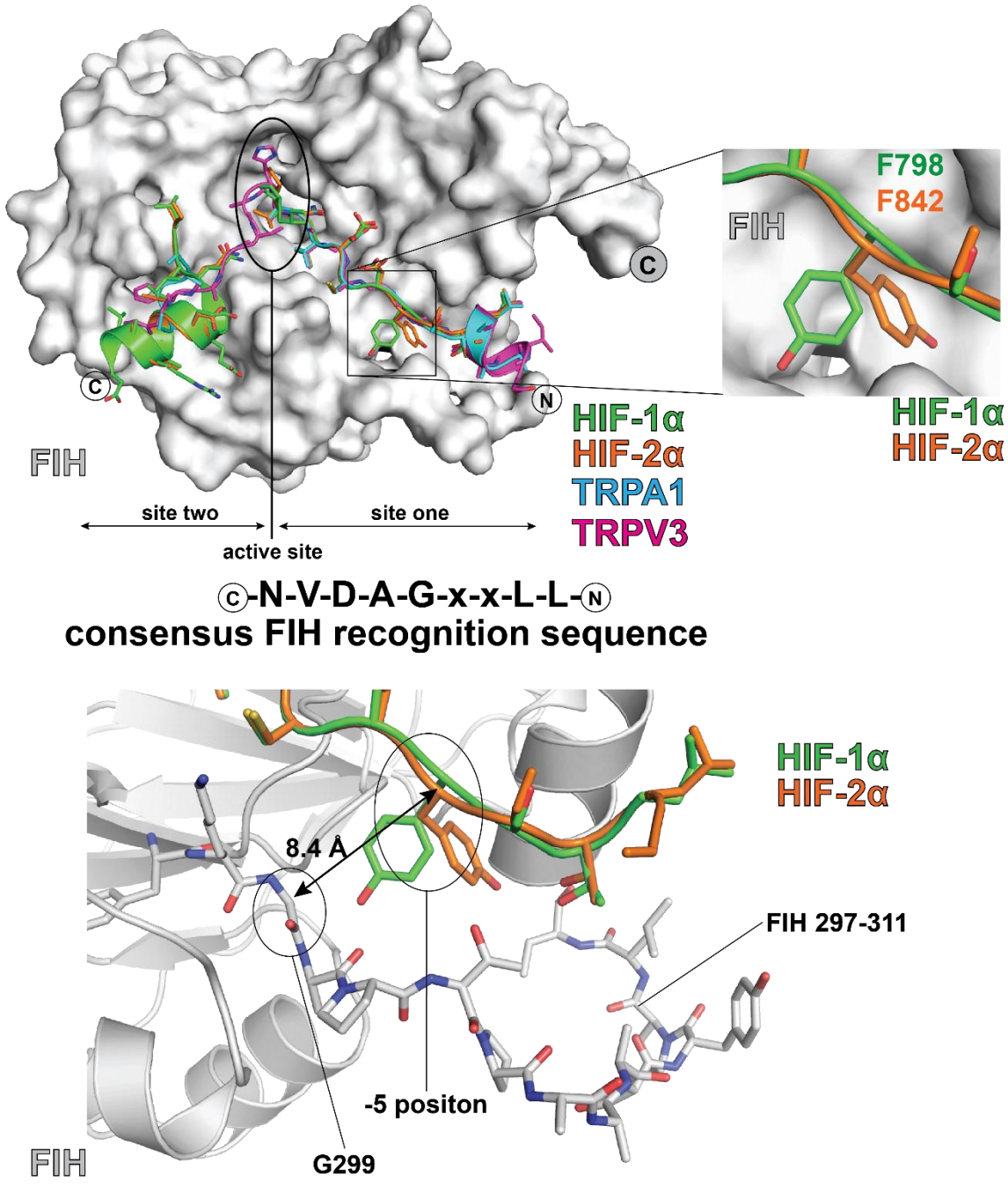


Figure 10 Design of an FIH-substrate disulfide cross-linking site. Top Views from crystal structures of FIH in complex with HIF-1α (2.15 Å, PDB code 1H2K), HIF-2α, TRPA1 and TRPV3 (see chapter 4). Views from crystal structures of FIH in complex with HIF-1/2α illustrate local variations in binding and potential cross-linking sites.

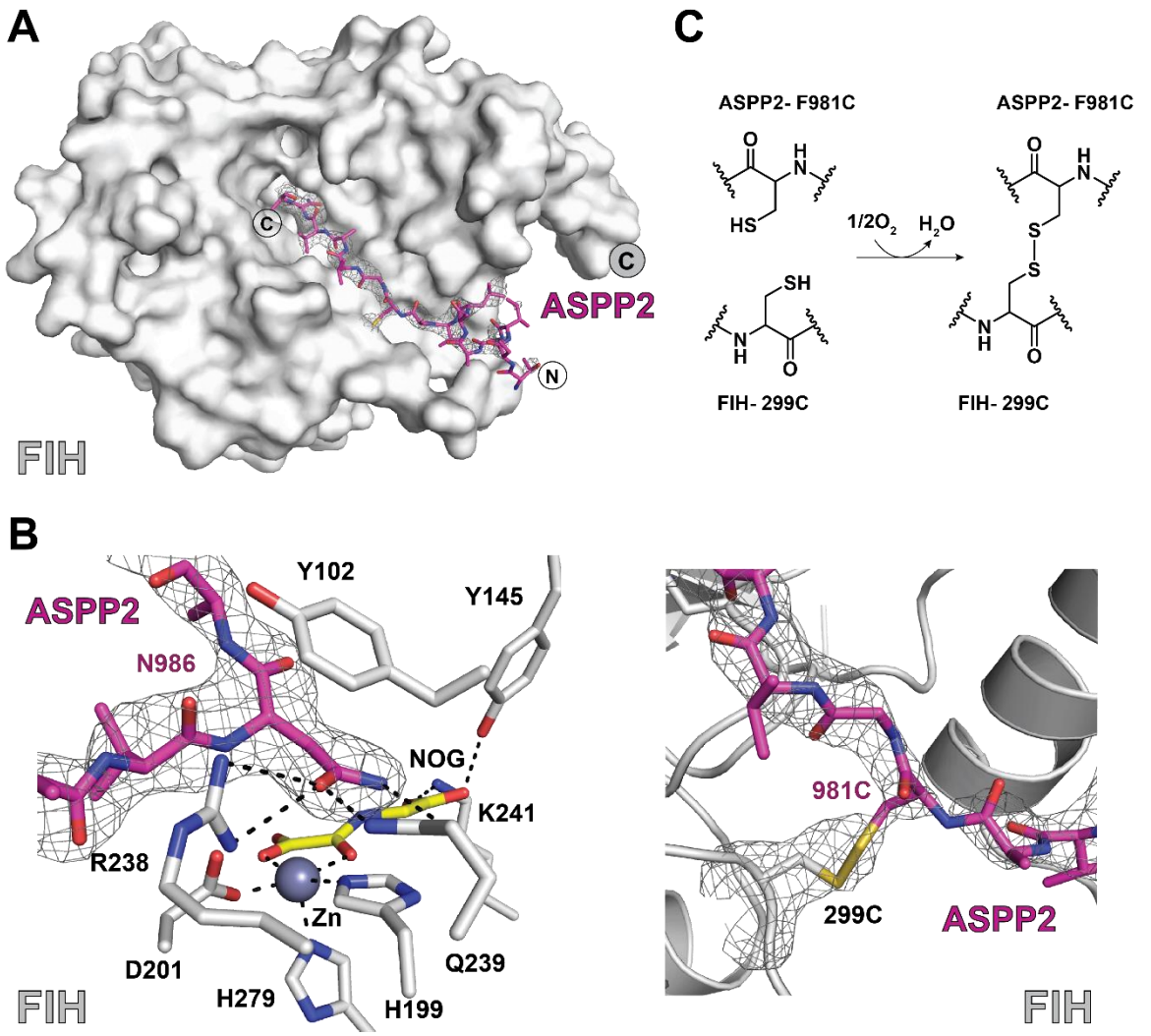


Figure 11 Views from a crystal structure of FIH in complex with an ASPP2 derived peptide. **A** View from a crystal structure of FIH in complex with ASPP2 F981C/N984A showing the overall binding mode of the substrate. **B** View from the active site and the cross-linking site of the FIH-ASPP2 complex. The mFo-Fc OMIT map is contoured to 3σ around the peptide. **C** Schematic mechanism of the disulfide cross-linking.

Initial cross-linking trials were performed using an ASPP2 F981C/N984A derived peptide with a cysteine at the -5 position to the hydroxylation site. FIH was incubated with a fivefold excess of peptide, in the presence of Zn(II) and NOG for 24 h at 4 °C prior to crystallisation trials. Crystallization trials were performed at 20 °C in sitting drop plates using optimised FIH crystallisation conditions (1.1-2.2 M ammonium sulphate, 2.5-6% v/v PEG 400). Crystals appeared within 16 h, were harvested and data collected (in collaboration with Dr Rasheduzzaman Chowdhury, University of Oxford). A structure was solved using molecular replacement (starting model PDB code 1H2K[18]) and the substrate modelled (Figure 11). Clear density for the cross-linking site could be observed for the disulfide bond confirming the success of the cross-link.

Table 4 Data collection and refinement statistics of an FIH G299C.Zn(II).NOG- ASPP2 N984A F981C complex.

Data Collection		**Refinement**	PHENIX
Beamline (Wavelength, Å)	In-house (RIGAKU FR-E Superbright) (1.5419)	R_{work}/R_{free}	0.2318/0.2750
Detector	Saturn 944+	No. atoms	
Data Processing	HKL2000	-Enzyme	2763
Space group	$P4_12_12$	-Ligand	NOG (10)
Cell dimensions		-Substrate	111
a,*b*,*c* (Å)	86.0, 86.0, 147.5	-Water	43
α,β,γ(°)	90, 90, 90	Avg. B-factors	
No. reflections	13507 (1335)*	Enzyme	48.4
Resolution (Å)	50.0-2.85 (2.95-2.85)*	Ligand	56.8
R_{merge}	0.200 (0.905)*	Substrate	61.8
I/σI	7.4 (1.7)*	Water	30.0
Completeness (%)	99.6 (99.8)*	R.m.s deviations	
Multiplicity	6.2 (6.0)*	-Bond lengths (Å)	0.007
Wilson B value (Å^2)	52.9	-Bond angles (°)	0.988

* Highest resolution shell in parentheses.

†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

5.3.3 Cross-linking of FIH with ARD substrates

Following from the initial FIH-substrate cross-linking trials with an ASPP2 derived peptide showing successful cross-linked substrate bound to the active site of FIH, further protein cross-linking sites were designed. As in the previously described crystallisation trials, mNotch1, 3CA, iASPP and ASPP2 were chosen for structural investigations. Cysteine mutations at the -5 position to the hydroxylation site were designed, namely: 3CA A79C, mNotch1 A1940C, iASPP A682C and ASPP2 F981C. Mutations were introduced using site directed mutagenesis and the proteins expressed and purified as described in chapter 10.

Crystallisation trials were performed as previously described (see section 5.3.2) with 24-48 h pre-incubation of the protein complex in the presence of NOG and zinc(II) prior to crystallisation trials. Crystallisation trials were performed at 20 °C and 4 °C.

All FIH crystal structures deposited in the Protein Data Bank (PDB, as of December 2017) are in the tetragonal space group $P4_12_12$. These crystals have characteristic bipyramidal morphology (for detailed description of FIH crystal morphology see chapter 8) although different crystal forms of FIH have been observed (see chapter 2 and chapter 4) but not yet published. Notably, crystals in the $P4_12_12$ space group appear within 12 h and continue to grow for another ~12 h when using peptide substrates. In case of the crystallisation trials of FIH G299C with mNotch1 A1940A, first crystals were observed after 24-48 h at 20 °C with new crystals appearing for 7 days in the above stated pyramidal crystal morphology. Crystal growth at 4 °C was significantly slower and tetragonal crystals appeared over the course of several weeks. Data collection at Diamond Light source showed that FIH G299C crystallised with mNotch1 in the $P4_12_12$ space group (Table 5). Solving a crystal structure using molecular replacement (search model PDB code 1H2K[18]) showed clear density for the substrate in site one and site two corresponding to residues 1934-1959 of mNotch1 (Figure 12) but surprisingly no further density for the other residues of the Notch construct. The overall binding within the substrate binding groove appears to be similar to previously observed ankyrin substrate binding (chapter 3). Notably, no cross-linking between mNotch1 and FIH was observed, modelling the cross-linking site lead to strong negative density in the mFo-Fc maps after refinement (Figure 12C). To investigate if the mNotch1 protein was subject to hydrolysis, ~40-50 crystals were fished, washed in well solution, and analysed by SDS PAGE next to FIH G299C, mNotch1 A1940C and the complex sample after incubation but prior to crystallization trials. No band for the mNotch1 construct was be observed in the crystal sample by SDS PAGE (Figure 12D).

The absence of the mNotch1 protein band in the crystal sample on the SDS PAGE gel suggests hydrolysis of the protein construct under the crystallisation conditions. Whether this is due to a destabilization of the Notch1 construct upon FIH binding, metal induced peptide backbone cleavage, or from potential protease contamination, is unclear. Kinetics of the crystallisation would imply slow hydrolysis of the mNotch1 construct followed by fast crystallisation of the cleaved FIH.mNotch1 complex. Degradation of the FIH protein was not observed.

Attempts to characterise the FIH.mNotch1 cross-linked crystal complex via non-denaturing or denaturing mass-spectrometry were not successful, likely due to the low amount of sample available from the crystals.

Notably, no cross-linking between FIH and the mNotch1 construct was observed. FIH crystals are, compared to other crystals, weakly diffracting for which the beam transmission and exposure time needed to be increased for data collections (see also chapter 3). Disulfide bonds are known to be vulnerable to radiation damage and frequently observed to be damaged by synchrotron radiation[53-57]. To reduce radiation damage, data for the FIH.mNotch1 complex crystals were collected using an in-house X-ray source (Table 5). In collaboration with Dr Rasheduzzaman Chowdhury (University of Oxford), FIH.mNotch1 crystals were fished, data collected, and a structure solved using molecular replacement.

A structure of the FIH.mNotch1 complex solved using data from an in-house X-ray source (FR-E Superbright, rotating copper anode, Rigaku) showed, as previously observed for the synchrotron data, clear density for the mNotch1 substrate in site one and site two of the FIH substrate binding groove but, again, no density for further substrate residues (Figure 12). Notably, at the designed cross-linking site, density for a metal mediated cross-linking, *via* additional zinc ions, instead of covalently cross-linking *via* disulfide bonds were observed.

In addition to the FIH.mNotch1 complexes, crystals were observed in the above described bipyramidal morphology for the FIH-3CA complex at 20 °C only. Collecting data using the in-house X-ray source and solving a structure, as previously described (see above), revealed electron density for the 3CA construct bound in the active site and the substrate binding groove of FIH (Figure 13). Notably, only in the substrate binding site one of FIH, electron density corresponding to bound 3CA was observed, corresponding to residues 71-87 of the synthetic ankyrin construct. Clear density could be observed for the cross-linking site between FIH and 3CA, showing the formation of the designed disulfide bridge between the enzyme and the substrate.

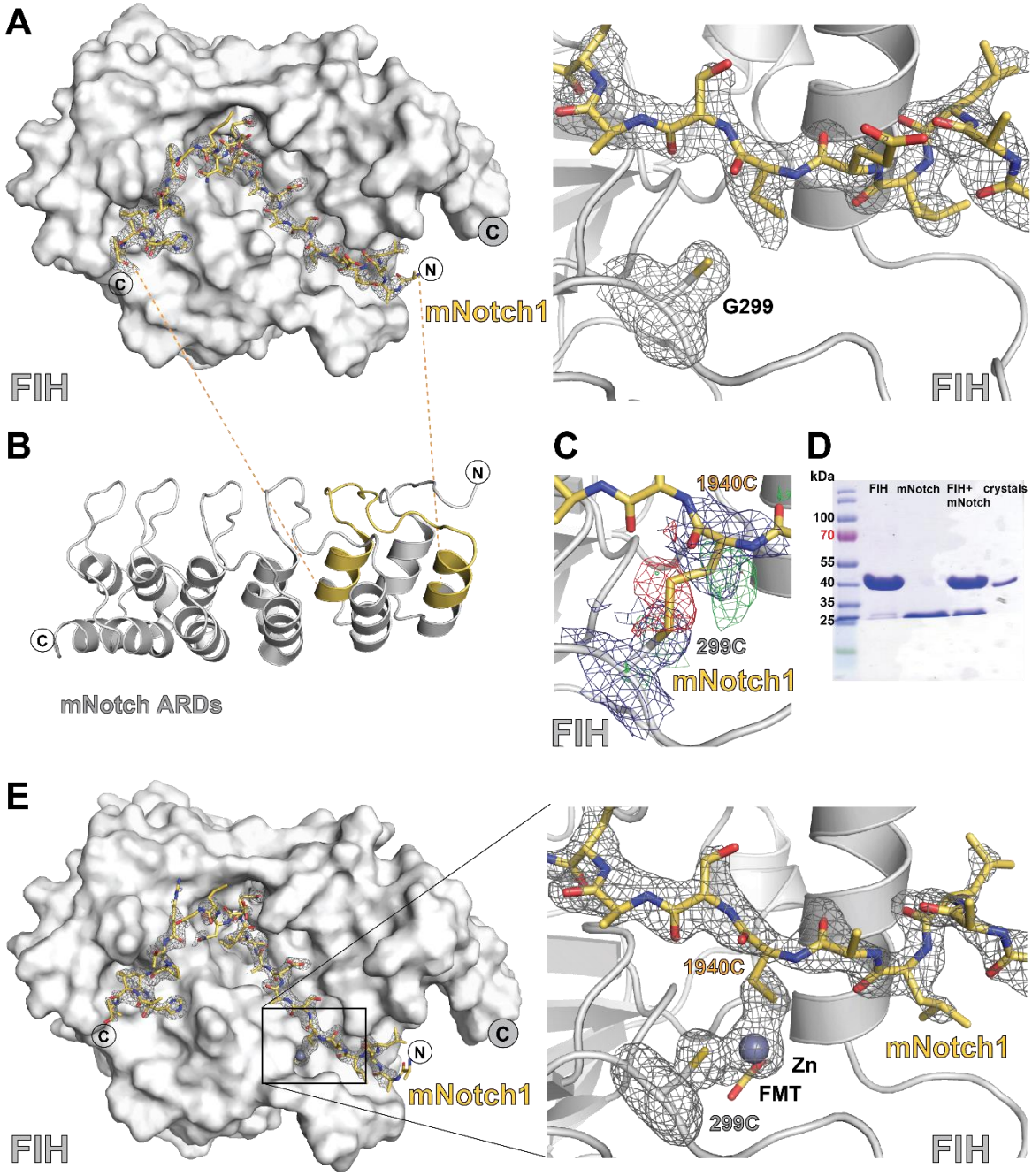


Figure 12 Views from crystal structures of FIH in complex with an mNotch1 ARD protein construct. A Views from a crystal structure of FIH in complex with mNotch1, collected at Diamond Light Source, illustrating the overall binding and enzyme substrate interactions at the designed cross-linking site; mFo-Fc OMIT maps are contoured to 3σ around mNotch1 and coloured in grey. **B** Views from a crystal structure of mNotch1 protein (2.25 Å, PDB code 2QC9[1]) illustrates the residues visible in complex with FIH. **C** Refinement of an FIH.mNotch1 complex with modelled disulfide cross-link. 2mFo-Fc maps are contoured to 1σ around the cross-linking site and coloured in blue, mFo-Fc maps are contoured to ±3σ around the cross-linking site and coloured in green for positive and red for negative values. **C** SDS PAGE of proteins used for cross-linking, crystallisation sample and of the FIH.mNotch1 complex crystals. **D** Views from a crystal structure of FIH in complex with mNotch1, collected using an in-house X-ray source, illustrating the overall binding mode of the substrate to FIH and the enzyme substrate interactions at the designed cross-linking site; mFo-Fc OMIT maps are contoured to 3σ around mNotch1 and coloured in grey.

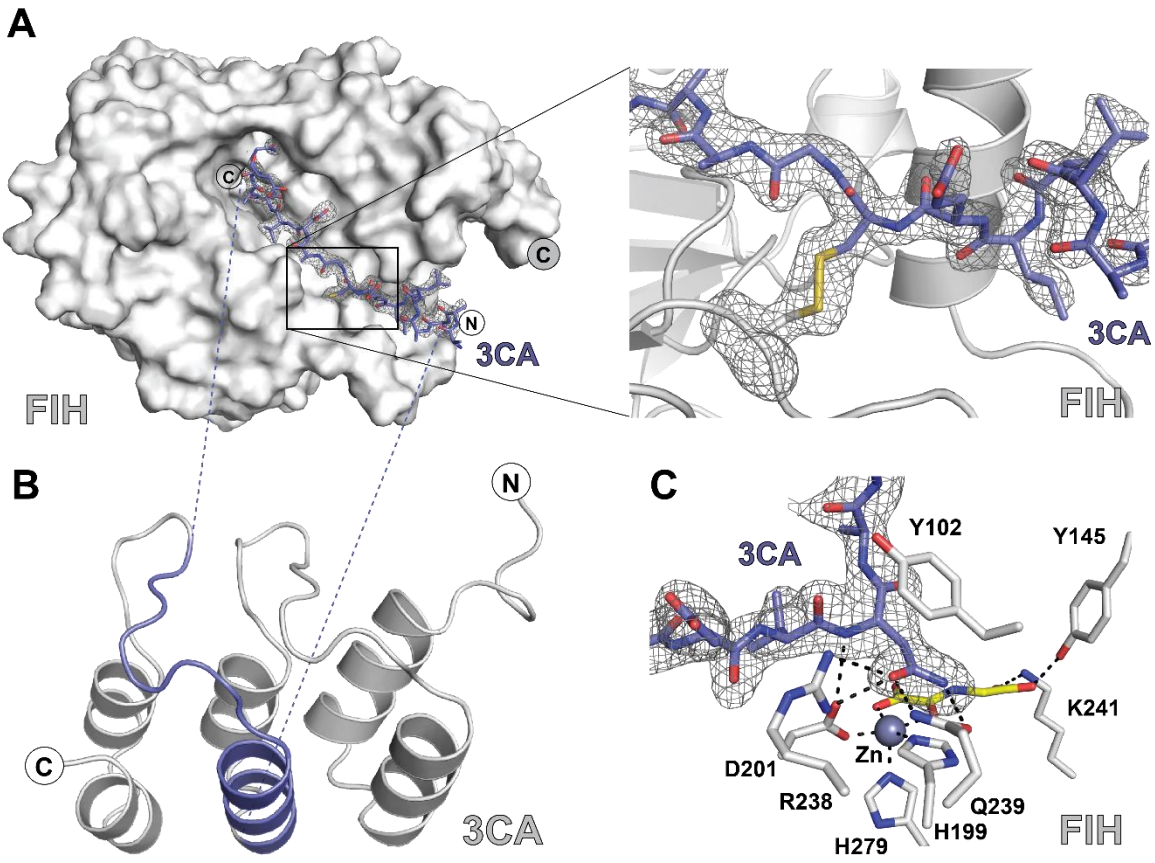


Figure 13 Views from a crystal structure of a cross-linked FIH.3CA complex. A Views from a crystal structure of FIH in complex with 3CA illustrates overall binding mode and disulfide cross-linking site. **B** View from a crystal structure of 3CA (1.63 Å, see abouve) illustrates the residues visible in the FIH complex crystal structure. **C** View from the active site of a cross-linked FIH.3CA complex. mFoFc OMIT maps are contoured to 3σ around 3CA and the cross-linking site.

In addition to the tetragonal crystals of the FIH.3CA complex, a second crystal form was observed in a hexagonal crystal morphology (Figure 14). This crystal form was not observed for any of the other FIH-substrate complexes. Several datasets were collected using Diamond Light Source and an in-house X-ray source. Diffraction could be observed between 5-15 Å by optical inspection. The spots could however not be indexed using MOSLFM, XDS, DIALS and HKL2000 and the structure could not be solved. Optimisation attempts did not improve the diffraction of the crystals.



Figure 14 Crystals from FIH G299C – ARD proteins co-crystallisation. Top Crystals grown from FIH G299C and 3CA 79C co-crystallisation at 20 °C. **Bottom** Crystals grown from FIH G299C co-crystallised with ASPP2 F981C and mNotch1 A1940C at 4 °C.

Unlike for the FIH.mNocht1 and FIH.3CA complexes, no crystals in the $P4_12_12$ space group were observed for co-crystallisation of ASPP2 F981C and iASPP A682C with FIH G299C. Two different crystal morphologies were observed when ASPP2 F981C was co-crystallised with FIH G299C. The first crystals grew within 3 months of crystallisation trials in a Proplex broad screen (Figure 14), diffraction was observed until ~20 Å (data not shown), the second crystal form grew as clusters of plates $P2_12_12_1$ space group within 6-12 month and showed X-ray diffraction >3 Å (Table 6). The overall data quality was poor for all crystals screened, with multiple crystal lattices present. Solving a structure using molecular replacement (search model PDB code 1H2K[18]) revealed that only FIH G299C was present in the crystals with no extra electron density corresponding to bound substrate. No further optimisation attempts were pursued.

Table 5 Data collection and refinement statistics of FIH G299C complexes with mNotch1 and 3CA.

Datasets	**FIH G299C**.Zn(II).NOG mNotch1 A1940A	**FIH G299C**.Zn(II).NOG mNotch1 A1940A	**FIH G299C**.Zn(II).NOG 3CA 80C
Data Collection			
Beamline (Wavelength, Å)	DLS I03 (0.9795)	In-house (RIGAKU FR-E Superbright) (1.5419)	In-house (RIGAKU FR-E Superbright) (1.5419)
Detector	PILATUS 6M-F	Saturn 944+	Saturn 944+
Data Processing	Xia2 3dii (XDS)	HKL2000	HKL2000
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.4, 86.4, 150.5	86.2, 86.2, 145.4	86.1, 86.1, 147.4
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. reflections	42916 (3121)*	36017 (3575)*	35321 (3405)*
Resolution (Å)	86.35-1.94 (1.99-1.94)*	21.12-2.02 (2.09-2.02)*	50.0-2.05 (2.10-2.05)*
R_{merge}	0.092 (2.529)*	0.079 (0.764)*	0.122 (0.944)*
I/σI	20.6 (1.4)*	16.6 (2.1)*	9.4 (1.4)*
CC 1/2	1.0 (0.6)*		0.515 (0.998)*
Completeness (%)	99.8 (99.8)*	97.9 (98.0)*	99.2 (98.4)*
Multiplicity	19.2 (17.6)*	4.0 (3.9)*	4.6 (4.0)*
Wilson B value (Å^2)	40.4	36.7	34.5
Refinement	PHENIX	PHENIX	PHENIX
R_{work}/R_{free}	0.1723/0.2011	0.1823/0.2134	0.1924/0.2245
No. atoms			
-Enzyme (A/B)	2666	2669	2618
-Substrate	179	178	110
-Ligand	NOG (10)	NOG (10)	NOG (10)
-Water	204	210	222
Avg. B-factors			
Enzyme	56.7	46.6	43.1
Substrate	100.6	81.7	57.6
Ligand	45.0	50.0	33.0
Water	60.3	54.0	47.7
R.m.s deviations			
-Bond lengths (Å)	0.003	0.003	0.006
-Bond angles (°)	0.562	0.776	0.586

* Highest resolution shell in parentheses.

† $R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

Table 6 Data collection statistics of FIH G299C crystallised in the presence of ASPP2 F981C an Zn(II) and NOG.

Data Collection	
Beamline (Wavelength, Å)	DLS I03 (0.9795)
Detector	Pilatus 6M-F
Data Processing	Xia2 3d (XDS)
Space group	$P2_12_12$
Cell dimensions	
a,*b*,*c* (Å)	49.3, 172.1, 210.3
α,β,γ(°)	90, 90, 90
No. of molecules/ASU	4
No. reflections	51644 (3801)*
Resolution (Å)	64.92-2.68 (2.75-2.68)*
R_{merge}	0.500 (2.778)*
I/σI	4.9 (0.9)*
CC ½	1.0 (0.4)*
Completeness (%)	900 (99.9)*
Multiplicity	9.2 (10.1)*
Wilson B value ($Å^2$)	48.5

* Highest resolution shell in parentheses.

5.3.4 Impact of cross-linking on substrate binding modes

Introducing covalent and non-covalent cross-links in the substrate binding site of proteins bears the risk of significantly changing the substrate binding mode, or at least stabilising one of potentially many possible conformations as it was previously observed[13,46]. To investigate the impact of the cross-links on the binding mode of FIH substrates, cross-linked structures were compared to non-cross-linked FIH substrate complex structures (Figure 15).

The overall binding mode of all FIH.ARD cross-linked complex structures (*i.e.* for FIH.mNotch1, FIH.ASPP2, FIH.3CA complexes) is conserved compared to their non cross-linked reference structure (Figure 15). Local variations can only be observed for ASPP2 F981C and 3CA 79C complexes at the cross-linking site. Noted are the differences in resolution for the FIH.ASPP2 complexes and FIH.CA complexes of ~1 Å and ~0.35 Å respectively between the different structures.

The overall conserved binding mode between cross-linked and non cross-linked peptide binding modes with FIH verifies the cross-linking approach as a suitable crystallisation method for otherwise inaccessible FIH-substrate complexes.

The reason why no cross-linking was observed between FIH and ASPP2 protein, in contrast to the corresponding peptide, remains unclear. Two potential explanations seem possible. Firstly, the complex forms reversibly and only fraction of FIH which did not bind to ASPP2 crystallised and is thereby removed from the equilibrium. In the case of a covalently bound substrate *via* disulfide, this seems unlikely but, at least in the case of mNotch1 binding, a non covalent cross-link was formed. Alternatively, the complex may not have formed under the crystallisation conditions. ASPP2 has been characterised as FIH substrate but interactions of ASPP2 with FIH have only been observed at 37 °C and not ambient temperature (chapter 6). Attempts follow the cross-linking *via* mass-spectrometry have not been successful due to detection issues.

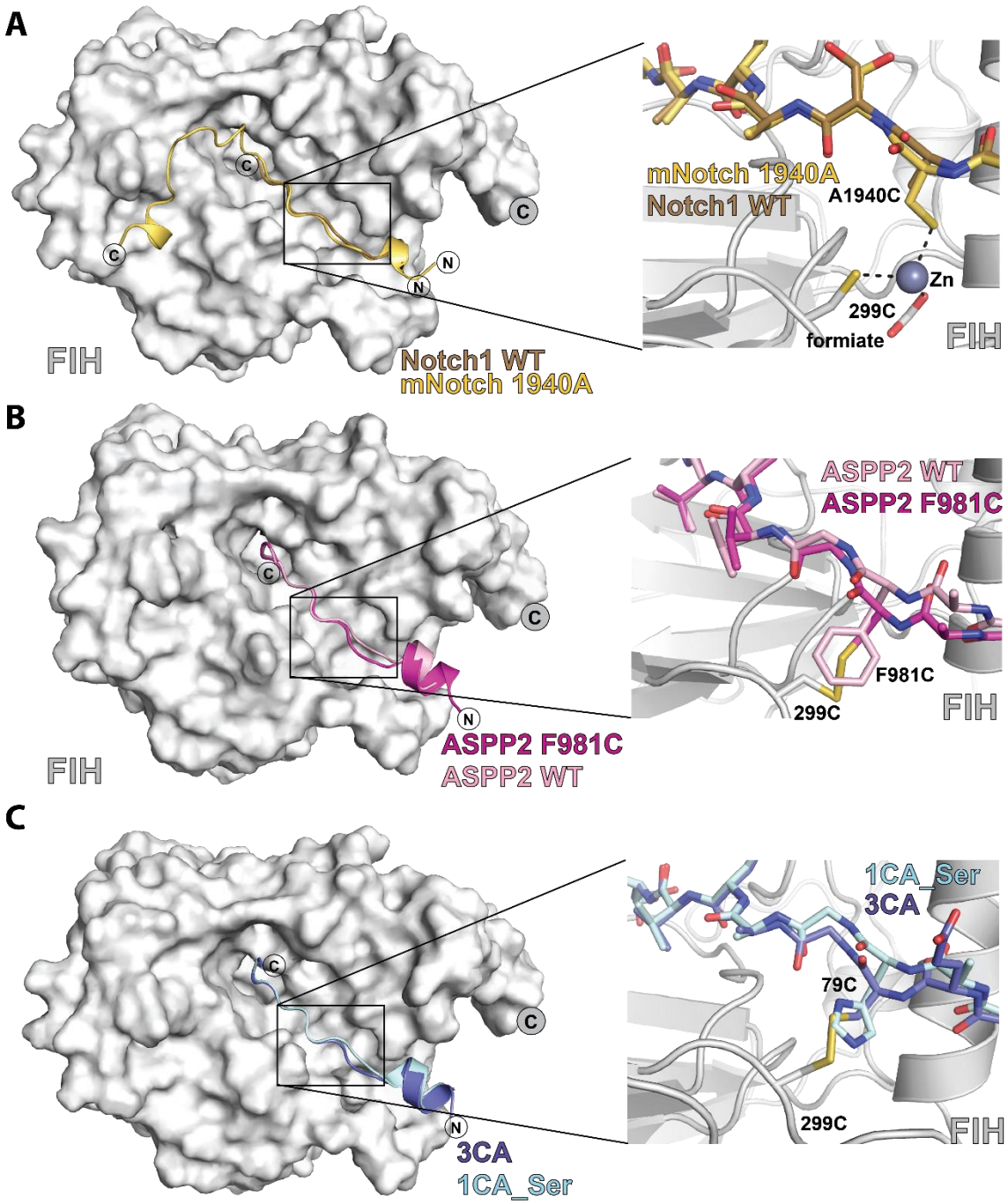


Figure 15 Comparison of cross-linked FIH-substrate complex structures with none cross-linked FIH substrate complex structures. A Views from a crystal structure of mNotch1 and a Notch1 fragment in complex with FIH (2.4 Å, PDB code 3P3N[1]) illustrate the impact of the non-covalent cross-link on FIH substrate binding. **B** Views from crystal structures of FIH in complex with ASPP2 derived peptide ASPP2 F981C/N984A and ASPP2 WT (1.89 Å, chapter 6) illustrate the impact of the covalent disulfide cross-link on FIH substrate binding. **C** Views from crystal structures of FIH in complex with 3CA 79C and a 1 consensus Ankyrin serine peptide (2.39 Å, PDB code 4B7K[58]) illustrate the impact of the covalent disulphide cross-link on FIH substrate binding for a consensus ankyrin sequence.

5.4 Conclusions and outlook

Several methodologies for the crystallisation of protein-protein complexes were applied to crystallise an FIH.ARD complex. 3CA crystallised in the presence of FIH WT separately. Application of *in situ* proteolysis approached did not result in diffracting FIH-substrate crystals.

Using N-terminally truncated constructs of FIH yielded several new FIH crystal forms. $FIH_{37\text{-}349}$ crystallises in the *I*422 space group to high resolution (>2 Å) and gives better defined electron density maps compared to data from FIH crystals in the $P4_12_12$. Better defined electron density maps facilitate the modelling of small molecules binding to FIH *e.g.* for crystallographic fragment screening.

A cross-linking strategy for FIH-substrate complexes was developed based on oxidation of cysteines to form a disulfide bonds. In the case of short peptides, cross-linking successfully immobilised the peptide in the substrate binding site. In the case of a FIH.3CA complex, a disulfide cross-link was successfully established but only a fraction of the 3CA protein was visible in a crystal structure of the FIH.3CA complex. In the case of an FIH.mNotch1 complex, a metal mediated cross-link between FIH and mNotch1 was found instead of a disulphide bond formed. The cross-link is only observable for in-house datasets and not for synchrotron datasets, possibly due to radiation damage. For iASPP and ASPP2 constructs, no evidence for cross-linking could be observed.

To take into account variations in affinity and possible temperature dependence of ankyrin binding to FIH, new strategies are likely needed for the stabilisation of the FIH ankyrin complex. Complementary techniques to X-ray crystallography, such as small angle scattering (SAXS) or NMR of labelled ankyrin repeats, may provide further insights into the domain re-arrangements of ARDs upon FIH binding. Mass spectrometry based approaches for the study of FIH ankyrin complex, such as hydrogen-deuterium exchange[59-62], could provide structural information on further binding sites between the ARD proteins and FIH and could even elucidate potential cross-talk between different ARD or HIF substrates bound to the same FIH dimer.

5.5 References

1 Coleman, M. L. *et al.* Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038, doi:10.1074/jbc.M704102200 (2007).

2 Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-Inducible Factor (FIH) as an Oxygen Sensor. *J Biol Chem*, doi:10.1074/jbc.M115.653014 (2015).

3 Spolar, R. S. & Record, M. T. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263**, 777-784 (1994).

4 Dyson, H. J. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**, 54-60 (2002).

5 Shammas, S. L., Crabtree, M. D., Dahal, L., Wicky, B. I. & Clarke, J. Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. *J Biol Chem* **291**, 6689-6695, doi:10.1074/jbc.R115.692715 (2016).

6 Jensen, M. R., Ruigrok, R. W. & Blackledge, M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr Opin Struct Biol* **23**, 426-435, doi:10.1016/j.sbi.2013.02.007 (2013).

7 Yu, T. K. *et al.* An unusual protein-protein interaction through coupled unfolding and binding. *Angew Chem Int Ed Engl* **53**, 9784-9787, doi:10.1002/anie.201404750 (2014).

8 Hardy, A. P., Prokes, I., Kelly, L., Campbell, I. D. & Schofield, C. J. Asparaginyl beta-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J Mol Biol* **392**, 994-1006, doi:10.1016/j.jmb.2009.07.070 (2009).

9 Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-inducible Factor (FIH) as an Oxygen Sensor. *J Biol Chem* **290**, 19726-19742, doi:10.1074/jbc.M115.653014 (2015).

10 Wilkins, S. E. *et al.* Differences in hydroxylation and binding of Notch and HIF-1alpha demonstrate substrate selectivity for factor inhibiting HIF-1 (FIH-1). *Int J Biochem Cell Biol* **41**, 1563-1571, doi:10.1016/j.biocel.2009.01.005 (2009).

11 Kelly, L., McDonough, M. A., Coleman, M. L., Ratcliffe, P. J. & Schofield, C. J. Asparagine beta-hydroxylation stabilizes the ankyrin repeat domain fold. *Mol Biosyst* **5**, 52-58, doi:10.1039/b815271c (2009).

12 Scotti, J. S. *et al.* Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Natl Acad Sci U S A* **111**, 13331-13336, doi:10.1073/pnas.1409916111 (2014).

13 Chowdhury, R. *et al.* Ribosomal oxygenases are structurally conserved from prokaryotes to humans. *Nature*, doi:10.1038/nature13263 (2014).

14 Loenarz, C. *et al.* The hypoxia-inducible transcription factor pathway regulates oxygen sensing in the simplest animal, Trichoplax adhaerens. *EMBO Rep* **12**, 63-70, doi:10.1038/embor.2010.170 (2011).

15 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, doi:10.1093/nar/gkq265 (2010).

16 Kato, M. *et al.* Crystal structure of a novel JmjC-domain-containing protein, TYW5, involved in tRNA modification. *Nucleic Acids Res* **39**, 1576-1585, doi:10.1093/nar/gkq919 (2011).

17 The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).

18 Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806, doi:10.1074/jbc.C200644200 (2003).

19 Dale, G. E., Oefner, C. & D'Arcy, A. The protein as a variable in protein crystallization. *J Struct Biol* **142**, 88-97 (2003).
20 Wernimont, A. & Edwards, A. In situ proteolysis to generate crystals for structure determination: an update. *PLoS One* **4**, e5094, doi:10.1371/journal.pone.0005094 (2009).
21 Dong, A. *et al.* In situ proteolysis for protein crystallization and structure determination. *Nat Methods* **4**, 1019-1021, doi:10.1038/nmeth1118 (2007).
22 Acuner Ozbabacan, S. E., Engin, H. B., Gursoy, A. & Keskin, O. Transient protein-protein interactions. *Protein Eng Des Sel* **24**, 635-648, doi:10.1093/protein/gzr025 (2011).
23 Chichili, V. P. R., Kumar, V. & Sivaraman, J. A method to trap transient and weak interacting protein complexes for structural studies. *Intrinsically Disord Proteins* **1**, e25464, doi:10.4161/idp.25464 (2013).
24 Back, J. W., de Jong, L., Muijsers, A. O. & de Koster, C. G. Chemical cross-linking and mass spectrometry for protein structural modeling. *J Mol Biol* **331**, 303-313 (2003).
25 Sinz, A. Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions. *Mass Spectrom Rev* **25**, 663-682, doi:10.1002/mas.20082 (2006).
26 Trakselis, M. A., Alley, S. C. & Ishmael, F. T. Identification and mapping of protein-protein interactions by a combination of cross-linking, cleavage, and proteomics. *Bioconjug Chem* **16**, 741-750, doi:10.1021/bc050043a (2005).
27 Chodosh, L. A. UV crosslinking of proteins to nucleic acids. *Curr Protoc Mol Biol* **Chapter 12**, Unit 12.15, doi:10.1002/0471142727.mb1205s36 (2001).
28 Moss, T., Dimitrov, S. I. & Houde, D. UV-laser crosslinking of proteins to DNA. *Methods* **11**, 225-234, doi:10.1006/meth.1996.0409 (1997).
29 Tian, B., Yang, J. & Brasier, A. R. Two-step cross-linking for analysis of protein-chromatin interactions. *Methods Mol Biol* **809**, 105-120, doi:10.1007/978-1-61779-376-9_7 (2012).
30 Toth, J. & Biggin, M. D. The specificity of protein-DNA crosslinking by formaldehyde: in vitro and in drosophila embryos. *Nucleic Acids Res* **28**, e4 (2000).
31 QUIOCHO, F. A. & RICHARDS, F. M. INTERMOLECULAR CROSS LINKING OF A PROTEIN IN THE CRYSTALLINE STATE: CARBOXYPEPTIDASE-A. *Proc Natl Acad Sci U S A* **52**, 833-839 (1964).
32 Lusty, C. J. A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *J Appl Crystallogr.* **32**, 106-112, doi:10.1107/S002188989801053X (1999).
33 Sauer, R. T. *et al.* An engineered intersubunit disulfide enhances the stability and DNA binding of the N-terminal domain of lambda repressor. *Biochemistry* **25**, 5992-5998 (1986).
34 Wells, J. A. & Powers, D. B. In vivo formation and stability of engineered disulfide bonds in subtilisin. *J Biol Chem* **261**, 6564-6570 (1986).
35 Scrutton, N. S., Berry, A. & Perham, R. N. Engineering of an intersubunit disulphide bridge in glutathione reductase from Escherichia coli. *FEBS Lett* **241**, 46-50 (1988).
36 Perry, L. J. & Wetzel, R. Disulfide bond engineered into T4 lysozyme: stabilization of the protein toward thermal inactivation. *Science* **226**, 555-557 (1984).
37 Ramakrishnan, V. *et al.* Geofold: topology-based protein unfolding pathways capture the effects of engineered disulfides on kinetic stability. *Proteins* **80**, 920-934, doi:10.1002/prot.23249 (2012).
38 Mansfeld, J., Vriend, G., Van den Burg, B., Eijsink, V. G. & Ulbrich-Hofmann, R. Probing the unfolding region in a thermolysin-like protease by site-specific immobilization. *Biochemistry* **38**, 8240-8245, doi:10.1021/bi990008p (1999).
39 Abkevich, V. I. & Shakhnovich, E. I. What can disulfide bonds tell us about protein energetics, function and folding: simulations and bioninformatics analysis. *J Mol Biol* **300**, 975-985, doi:10.1006/jmbi.2000.3893 (2000).

40 Clarke, J. & Fersht, A. R. Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32**, 4322-4329 (1993).

41 Heinz, D. W. & Matthews, B. W. Rapid crystallization of T4 lysozyme by intermolecular disulfide cross-linking. *Protein Eng* **7**, 301-307 (1994).

42 Clarke, J., Henrick, K. & Fersht, A. R. Disulfide mutants of barnase. I: Changes in stability and structure assessed by biophysical methods and X-ray crystallography. *J Mol Biol* **253**, 493-504, doi:10.1006/jmbi.1995.0568 (1995).

43 van den Akker, F. *et al.* Crystal structure of heat-labile enterotoxin from Escherichia coli with increased thermostability introduced by an engineered disulfide bond in the A subunit. *Protein Sci* **6**, 2644-2649, doi:10.1002/pro.5560061219 (1997).

44 Velanker, S. S. *et al.* Disulfide engineering at the dimer interface of Lactobacillus casei thymidylate synthase: crystal structure of the T155C/E188C/C244T mutant. *Protein Sci* **8**, 930-933, doi:10.1110/ps.8.4.930 (1999).

45 Yang, C. G. *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961-965, doi:10.1038/nature06889 (2008).

46 Wang, C. *et al.* Structure of the JmjC domain-containing protein NO66 complexed with ribosomal protein Rpl8. *Acta Crystallogr D Biol Crystallogr* **71**, 1955-1964, doi:10.1107/S1399004715012948 (2015).

47 Chowdhury, R. *et al.* Structural basis for oxygen degradation domain selectivity of the HIF prolyl hydroxylases. *Nat Commun* **7**, 12673, doi:10.1038/ncomms12673 (2016).

48 Clerc, A., Catty, P. & Bennett, N. Interaction between cGMP-phosphodiesterase and transducin alpha-subunit in retinal rods. A cross-linking study. *J Biol Chem* **267**, 19948-19953 (1992).

49 Schmidt, C. & Urlaub, H. Combining cryo-electron microscopy (cryo-EM) and cross-linking mass spectrometry (CX-MS) for structural elucidation of large protein assemblies. *Curr Opin Struct Biol* **46**, 157-168, doi:10.1016/j.sbi.2017.10.005 (2017).

50 Xie, J. & Schultz, P. G. A chemical toolkit for proteins--an expanded genetic code. *Nat Rev Mol Cell Biol* **7**, 775-782, doi:10.1038/nrm2005 (2006).

51 Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Lett* **588**, 206-212, doi:10.1016/j.febslet.2013.11.024 (2014).

52 Craig, D. B. & Dombkowski, A. A. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics* **14**, 346, doi:10.1186/1471-2105-14-346 (2013).

53 Sutton, K. A. *et al.* Insights into the mechanism of X-ray-induced disulfide-bond cleavage in lysozyme crystals based on EPR, optical absorption and X-ray diffraction studies. *Acta Crystallogr D Biol Crystallogr* **69**, 2381-2394, doi:10.1107/S0907444913022117 (2013).

54 Bury, C. *et al.* Radiation damage to nucleoprotein complexes in macromolecular crystallography. *J Synchrotron Radiat* **22**, 213-224, doi:10.1107/S1600577514026289 (2015).

55 Burmeister, W. P. Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr D Biol Crystallogr* **56**, 328-341 (2000).

56 Garman, E. F. & Owen, R. L. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **62**, 32-47, doi:10.1107/S0907444905034207 (2006).

57 Ravelli, R. B. & Garman, E. F. Radiation damage in macromolecular cryocrystallography. *Curr Opin Struct Biol* **16**, 624-629, doi:10.1016/j.sbi.2006.08.001 (2006).

58 Yang, M. *et al.* Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. *J Biol Chem* **286**, 7648-7660, doi:10.1074/jbc.M110.193540 (2011).

59 Pantazatos, D. *et al.* Rapid refinement of crystallographic protein construct definition employing enhanced hydrogen/deuterium exchange MS. *Proc Natl Acad Sci U S A* **101**, 751-756, doi:10.1073/pnas.0307204101 (2004).

60 Spraggon, G. *et al.* On the use of DXMS to produce more crystallizable proteins: structures of the T. maritima proteins TM0160 and TM1171. *Protein Sci* **13**, 3187-3199, doi:10.1110/ps.04939904 (2004).

61 Sharma, S. *et al.* Construct optimization for protein NMR structure analysis using amide hydrogen/deuterium exchange mass spectrometry. *Proteins* **76**, 882-894, doi:10.1002/prot.22394 (2009).

62 Khan, A. G. *et al.* Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. *Nature* **509**, 381-384, doi:10.1038/nature13117 (2014).

6 Scoping the catalytic landscape of FIH mediated reactions

This chapter describes work performed to scope the catalytic landscape of FIH mediated hydroxylation. ASPP proteins are characterised as FIH substrates using peptide, recombinant protein and cell based hydroxylation assays. In addition to the predicted hydroxylation sites within the ankyrin repeats of all ASPP proteins, an unprecedented additional FIH dependent second hydroxylation was observed when recombinant protein, or transfected full-length ASPP protein in human cell lines, was incubated with FIH. ASPP fragments are not doubly hydroxylated. Further investigations with recombinant forms of p18-INK4c and ASB11 proteins also manifested FIH mediated double hydroxylation suggesting a potentially broader scope of double hydroxylations on ARD proteins.

6.1 Introduction: The ASPP protein family

Tight control of the cell cycle and cell growth is essential in higher organisms as uncontrolled proliferation can lead to cancer. The tumour suppressor protein p53 plays a crucial role in cellular homeostasis, response to cellular stress and DNA damage[1-4]. In the case of moderate DNA damage and cellular stress, p53 induces cell-cycle arrest, DNA repair and metabolic homeostasis[5-10]. In the case of severe stress, however, p53 causes senescence and apoptosis[11-14]. Due to its central role in DNA damage response, p53 is often referred to as the "guardian of the genome"[15] and is, to the extent of current knowledge, the most important tumour suppressor protein in humans[16,17]. This is underlined by the observation that the p53 gene is mutated in ~50% of all cancers[18].

The activity of p53 is tightly controlled by many cellular interactions. Amongst others, the Apoptosis-Stimulating Proteins of p53 (also known as Ankyrin-repeat-,SH3-domain and proline rich-region proteins, ASPP) are key regulators of p53 activity[19]. The ASPP protein family consists of 3 members, the pro-apoptotic ASPP1 and ASPP2[20-26], and the anti-apoptotic protein iASPP[27,28]. Despite their opposing effects on cell-fate decisions, all ASPP proteins share a similar domain architecture and a highly conserved C-terminal region which consists of four Ankyrin repeats and adjacent SH3 domains (Figure 1). The C-terminal region are also reported to be binding sites for p53, and its two family members p63 and p73[25,29].

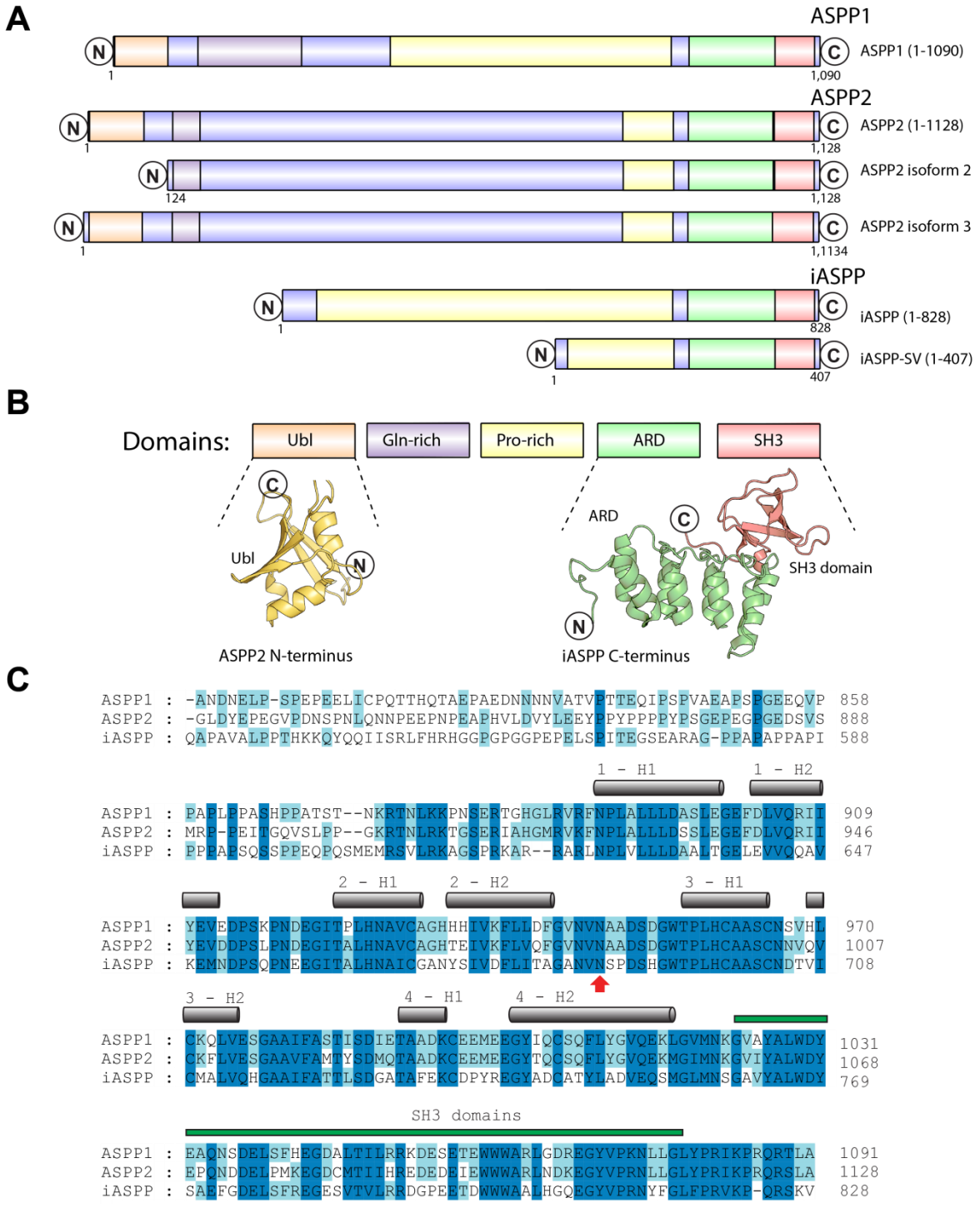


Figure 1 Domain structures of ASPP proteins. A Overview of the domain structures of the three ASPP proteins and reported isoforms. Isoforms as reported in UniProt[30], domain annotation according to InterPro[31] and structural analysis[29,32,33]. **B** Views from an NMR structure of the N-terminus of ASPP2 generated from PDB code 2UWQ[34], view from a crystal structure of iASPP C-terminus generated from PDB code 2VGE, 2.10 Å[33]. **C** Sequence alignment of the C-termini of ASPP1, ASPP2, and iASPP illustrates sequence homology between all three ASPP isoforms. The canonical FIH hydroxylation site within the ankyrin repeats is marked by a red arrow.

The function of ASPP proteins in the regulation of p53 target gene expression is relatively well understood. ASPP1 and ASPP2 induce the expression of pro-apoptotic genes, whereas iASPP regulates the expression of pro-survival genes (Figure 2). ASPP1 and ASPP2 are frequently found to be mutated or down-regulated in cancer whereas iASPP is found to be up-regulated in some cancers[18]. Recent studies suggest, however, that the physiological roles of iASPP and ASPP2 might be more complex and more context dependent than originally proposed. iASPP has been shown to have tumour suppression activity under specific circumstances (Dr Paul Miller, unpublished data). In addition, ASPP1 and ASPP2 were shown to activate Yes associate protein 1 (YAP1) and transcriptional coactivator with PDZ-binding motif (TAZ), both of which are oncogenic effector proteins of the HIPPO pathway which regulates organ growth (Figure 2)[35-39].

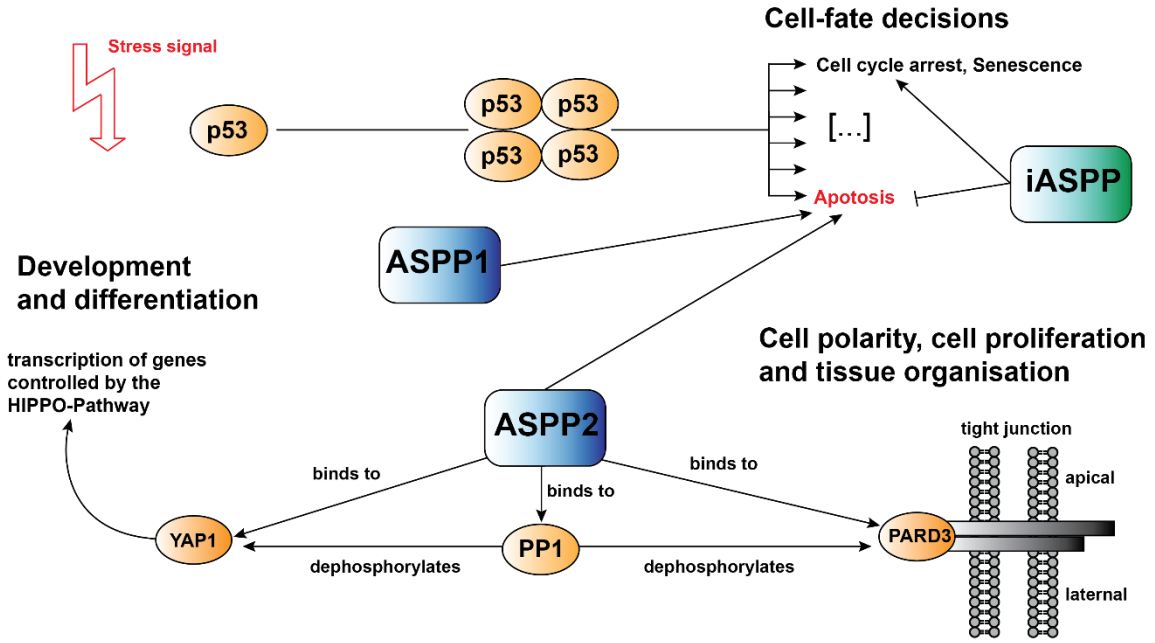


Figure 2 Physiological role of ASPP2 in cells. ASPP2 binds to p53 and induces the p53 dependent expression of pro-apoptotic genes. Additionally, ASPP2 interacts with PARD3 and YAP1 and can induce PP1 mediated dephosphorylation of both, thereby playing a crucial role in cell polarity, proliferation, tissue organization and development and differentiation.

In vivo studies show that ASPP2 KO mice have a dysregulated central nervous system and show increased spontaneous tumourigenesis, linking ASPP2s role as a tumour suppressor and keeper of the cell polarity[23,40].

Subsequently, ASPP2 has been shown to interact with a large number of proteins, including: partitioning defective 3 homolog (PARD3, an integral factor for the development and maintenance of cell polarity and part of the tight junction complex[40,41]); protein phosphatase 1 (PP1)[36,37,39,42]; and adenomatous polyposis coli protein-like (APCL)[43]. The cellular localisation of ASPP2 is a crucial factor for the formation of its protein complexes. The substantial number of cellular interaction partners of ASPP2 within various cellular signalling pathways could suggest a function as a scaffold protein for pathways beyond p53 regulation and the maintenance of cell polarity. In the Biogrid database[44,45], more than 100 ASPP2 protein-protein interactions are reported from high-throughput interaction studies. Although the number of proteins directly binding to ASPP2 is likely lower, it demonstrates that ASPP2 interacts with large numbers of proteins. ASPP2 interactions are further discussed in chapter 7.

Amongst other signalling molecules, ASPP2 has been reported to interact with FIH[46]. Further, it has been shown that ASPP2 is hydroxylated at asparagine-986 in an FIH dependent manner. Effects of FIH mediated ASPP2 hydroxylation on ASPP2-p53 interactions have not been found but an effect of FIH KO on ASPP2 association with PARD3 was reported[46]. With the ASPP2-PARD3 interaction site located at the N-terminus of ASPP2 and FIH-ASPP2 interaction sites located at the C-terminus of the protein, a direct link between the two effects seems unlikely.

6.2 FIH mediated hydroxylation of ASPP proteins

6.2.1 FIH interacts with ASPP proteins in U2OS cells

The interaction of ASPP2 with FIH has recently been reported using peptide and protein constructs[46]. However, the FIH-ASPP2 interactions were not studied at endogenous levels. Thus, experiments were performed to determine whether the FIH interaction is unique to ASPP2 or shared with other ASPP family members (ASPP1 and iASPP). Endogenous ASPP2 was immunoprecipitated from U2OS cells (for optimisation of endogenous ASPP2-IP see chapter 7). Only a faint band was observed for FIH (Figure 3A), in accordance with previous data on transient FIH-substrate complexes (see chapter 1). To stabilise transient FIH-substrate complexes, the cells were treated with dimethyloxalylglycine (DMOG). FIH tagged with the human influenza hemagglutinin tag (FIH-HA), the catalytically inactive iron binding site mutant FIH-HA D201A and the dimerisation mutant FIH-HA L340R (reported as inactive[47]) were transfected into U2OS FIH KO cells, DMOG was added and FIH was immunoprecipitated after 24 h (Figure 3B).

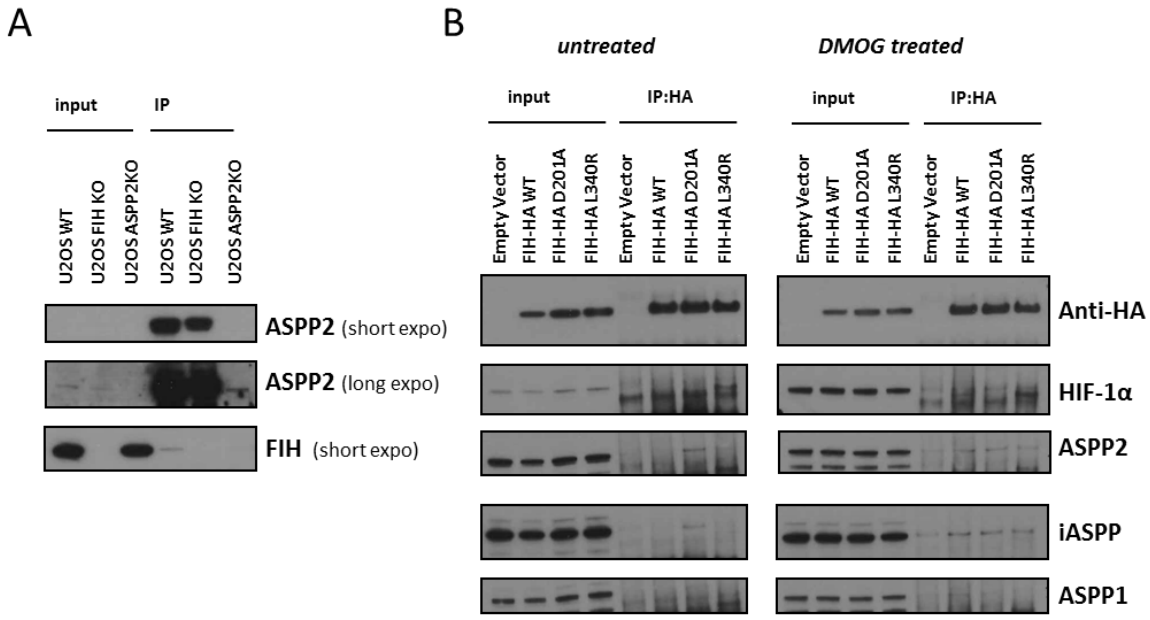


Figure 3 FIH interacts with ASPP proteins in U2OS cells. A Endogenous ASPP2 co-immunoprecipitated FIH from U2OS cells. **B** FIH-HA WT and the L340R variant of FIH can co-IP ASPP proteins from U2OS FIH KO cells after DMOG treatment. The FIH-HA D201A variant co-immunoprecipitated ASPP proteins unaffected by DMOG treatment.

As previously observed for the immunoprecipitation of endogenous ASPP2 WT from U2OS cells (Figure 3), nearly no ASPP2 was co-immunoprecipitated with FIH-HA from the U2OS FIH KO cells. A clear band for ASPP2 could, however, be observed in the IP of FIH-HA from DMOG treated USOS cells, supporting the idea that inhibiting FIH substrate turnover by NOG stabilises the transient FIH-substrate complex.

FIH D201A was shown to be catalytically inactive despite its ability to turn-over 2OG and its ability to bind substrates[47]. The enzyme-substrate complex is locked in the resting state (for mechanistic discussion on FIH catalysed protein hydroxylation see chapter 1) and is believed to be comparable to NOG treatment. This is further supported by the fact that DMOG treatment of the U2OS cells did not have any visible effects on the ability of FIH-HA D201A to co-immunoprecipitate the ASPP proteins (Figure 3).

FIH-HA L340R is reported to be unable to bind HIF-α substrates[48]. Surprisingly, FIH-HA L340R was able, like FIH-HA WT, to co-immunoprecipitate ASPP proteins after DMOG treatment. This could suggest that the FIH L340R variant is catalytically active in cells. Re-sequencing of the plasmid confirmed that the mutation was indeed present.

Out of the three ASPP proteins, only ASPP2 has been reported to be hydroxylated by FIH[46], whereas iASPP was proposed not to be accepted by FIH as substrate and ASPP1 has not been characterised. These results suggest, in contrast to the literature, that all ASPP proteins interact with FIH.

6.2.2 ASPP proteins interact with FIH and can compete with HIF-α as a substrate

A key feature of FIH-ankyrin interactions is the ability of ankyrins to bind to FIH and thereby compete with HIF-α as FIH substrate[49-52]. Therefore, it was of interest to investigate whether the ASPP proteins were competitors of FIH substrates in regard to HIF-1α. A peptide derived from HIF-1α CAD (HIF-1α 35mer, residues 789-822, Tarhonskaya et al[53]) was incubated in the presence of equimolar ARD proteins (3CA, iASPP and ASPP2) and HIF-1α hydroxylation levels were monitored over 20 min (Figure 4). Addition of the three consensus ankyrin construct (3CA) produced only a limited effect on the hydroxylation levels of HIF-1α whereas iASPP, and ASPP2 showed substantial reduction of HIF-1α hydroxylation. The experiments were repeated with a peptide derived from HIF-2α CAD (HIF-2α 35mer, residues 832-866, Tarhonskaya et al[53]) with comparable results (Figure 4B).

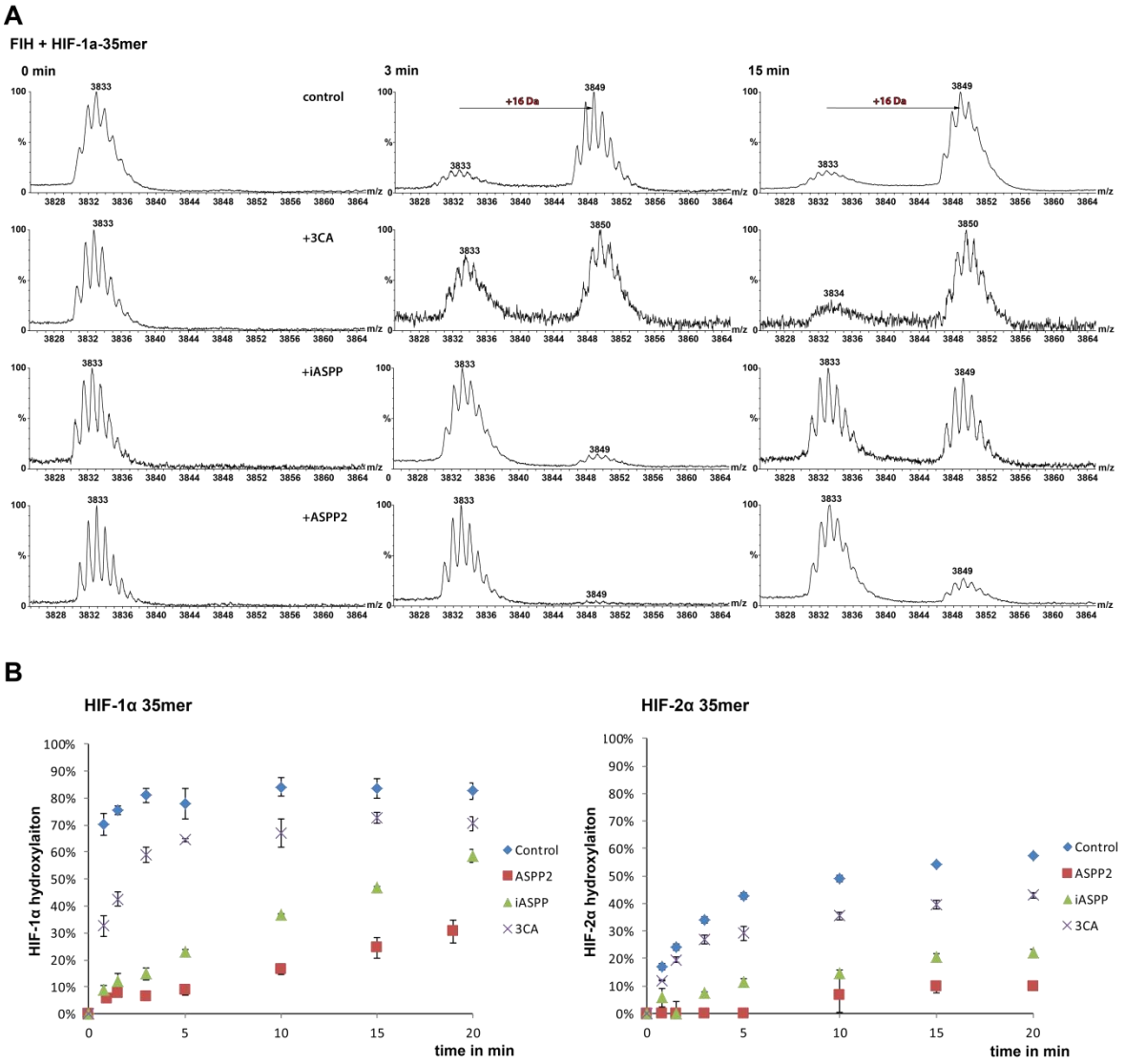


Figure 4 Competition of ASPP proteins with HIF-α for FIH activity *in vitro*. A Representative MALDI-MS spectra of HIF-1α 35mer peptide hydroxylation in the presence of no ankyrin, 3CA, iASPP and ASPP2. **B** Time courses of HIF-1α 35mer peptide hydroxylation and HIF-2α 35mer peptide hydroxylation in the presence of ankyrin competitors. Reaction conditions: 4 μM FIH, 100 μM HIF-α 35mer, 100 μM ARD protein, 1 mM ascorbate, 1 mM 2OG, 200 μM Fe(II), 50 mM Tris-HCl pH = 7.5 incubated for indicated times at 37 °C.

To investigate if the observed differences between ASPP2, iASPP and 3CA in their ability to compete with HIF-α for FIH binding may result from potentially different binding affinities to FIH, surface plasmon resonance (SPR) were pursued. ASPP2 and iASPP were immobilised on a sensor chip and K_ds measured by SPR. The SPR measurements gave a K_d of 2.2 μM for ASPP2 and 1.1 μM for iASPP (Figure 5). Surprisingly, this is not in accordance with the competition data; stronger binding would have been expected from ASPP2. A possible reason for the discrepancy could be that immobilising the substrate on the sensor chip might hinder FIH binding. Furthermore, only one substrate is bound to the surface, the dimeric nature of FIH and the possibility of multiple substrates binding in

cooperative manner would not be considered using SPR under these conditions. Attempts to immobilise FIH on the sensor chip were not pursued due to time restraints.

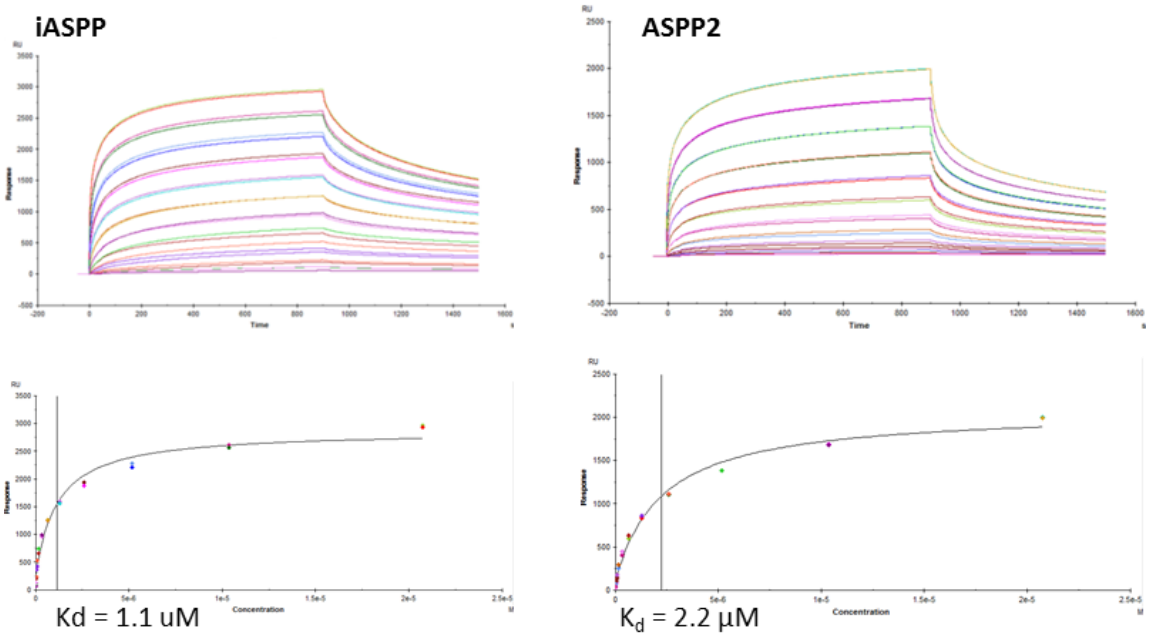


Figure 5 SPR measurements of iASPP and ASPP2 with FIH. All measurements were carried out at 25 °C. Biotinylated ASPP2 and iASPP were immobilized on the sensor chip via streptavidin (~2,000 response units, RU) and measured with an FIH flow.

In the crystalline state, the available evidence is that FIH can bind one peptide substrate per molecule[51,54-56], so it is possible two substrates can bind per dimer in solution. It was therefore tested if FIH could bind multiple intact ARD protein substrates simultaneously. In collaboration with Mr Joseph Donlan (University of Oxford), non-denaturing mass-spectrometric analysis were performed on FIH-ankyrin complexes. The results indicate that FIH can indeed bind one or two ankyrin proteins simultaneously (Figure 6).

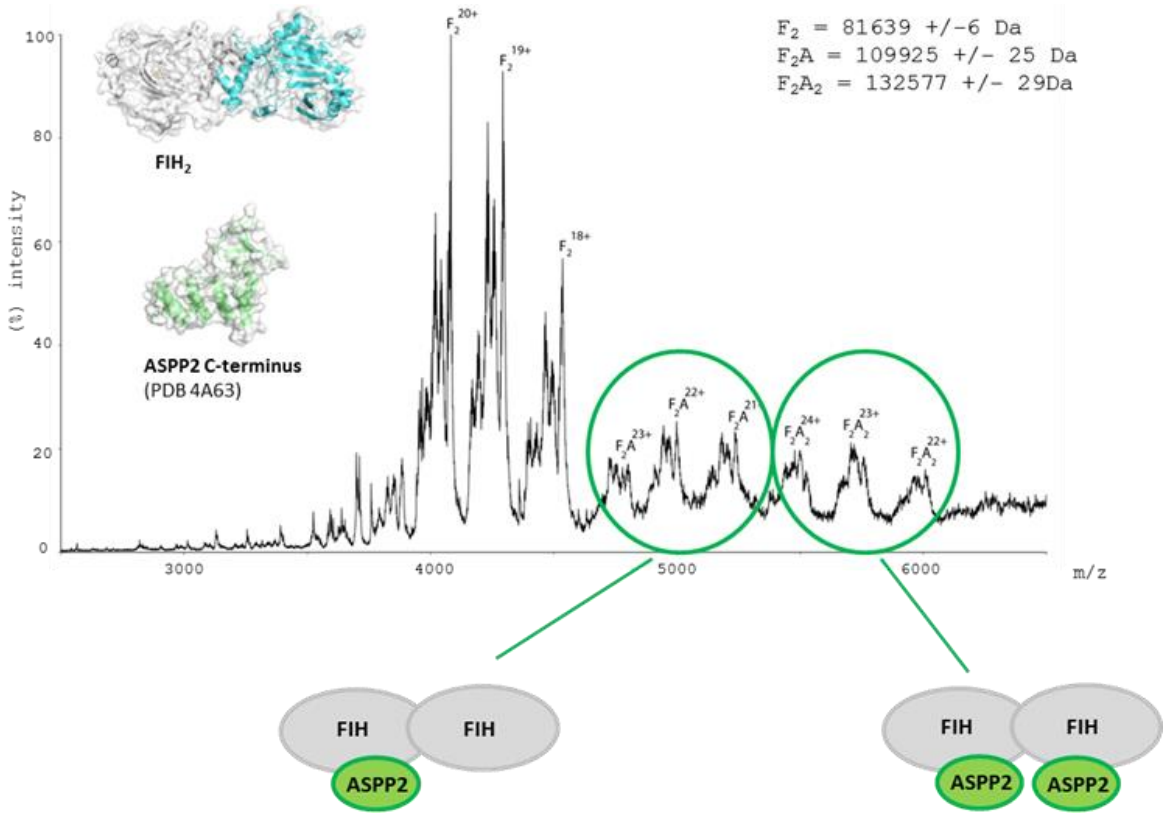


Figure 6 Non-denaturing MS of FIH-ASPP2 complexes. MS spectrum of the FIH-ASPP complex shows three distinct species: the FIH dimer, the FIH dimer in complex with one ASPP2 protein and the FIH dimer in complex with two ASPP2 proteins. The MS spectra were acquired by Mr Joseph Donlan, University of Oxford).

To remove any potential structural interference from the SPR chip and to enable the binding of multiple substrates, isothermal titration calorimetry (ITC) was pursued in collaboration with Dr Sarah Picaud (University of Oxford). To determine whether binding between an ankyrin and FIH could be determined by ITC, and to establish the experimental parameters, a consensus ankyrin (1CA) peptide (see also Tarhonskaya et al[53]) was analysed for binding to FIH. To avoid substrate hydroxylation and metal oxidation, 2OG was substituted with NOG and Fe(II) substituted with Zn(II). FIH was added to the cell and a solution of 1CA peptide was titrated into the cell under mixing. The overall energy released upon substrate addition was very low however, a K_d of ~100 nM was calculated, and the experiment determined that two 1CA peptides could simultaneously bind to 1 dimer of FIH (Figure 7).

The same parameters were then used to characterise the binding of 3CA to FIH. 3CA was found to precipitate under the experimental conditions. To rule out the possibility that the high protein concentration of 3CA was the cause of instability, a solution of FIH was titrated into 3CA, however, no evidence for stable binding was accrued (Figure 7).

To exclude a substrate specific effect, further ITC measurements with iASPP were performed. As with the 3CA, iASPP precipitated when concentrated for the titration under the experimental conditions. With iASPP placed in the sample cell and FIH titrated into it, some energetic changes could be observed. These, however, appear to be more related to the addition of FIH into the cell solution rather than the binding of substrate. Whether the observed effects originate from the possible dilution energy of the FIH stock or possible dissociation of the FIH dimer is unknown (Figure 7).

A possible explanation of the lack of binding observed between FIH and the ARD proteins could be the temperature of the experiments. The ITC measurements were performed at 15 °C (to increase protein stability for chapter 6, Appendix), whereas the hydroxylation assays were performed at 37 °C. FIH is active at room temperature (see supporting information for chapter 6, Appendix) but at least in the case of ASPP2 protein, hydroxylation was only observed at 37 °C (see supporting information). Potential effects of the ambient temperature on ankyrin unfolding might play a role in FIH-ankyrin binding.

Attempts to determine the binding affinities of ARD proteins to FIH and to explain the discrepancy between SPR and ITC measurements were not pursued further.

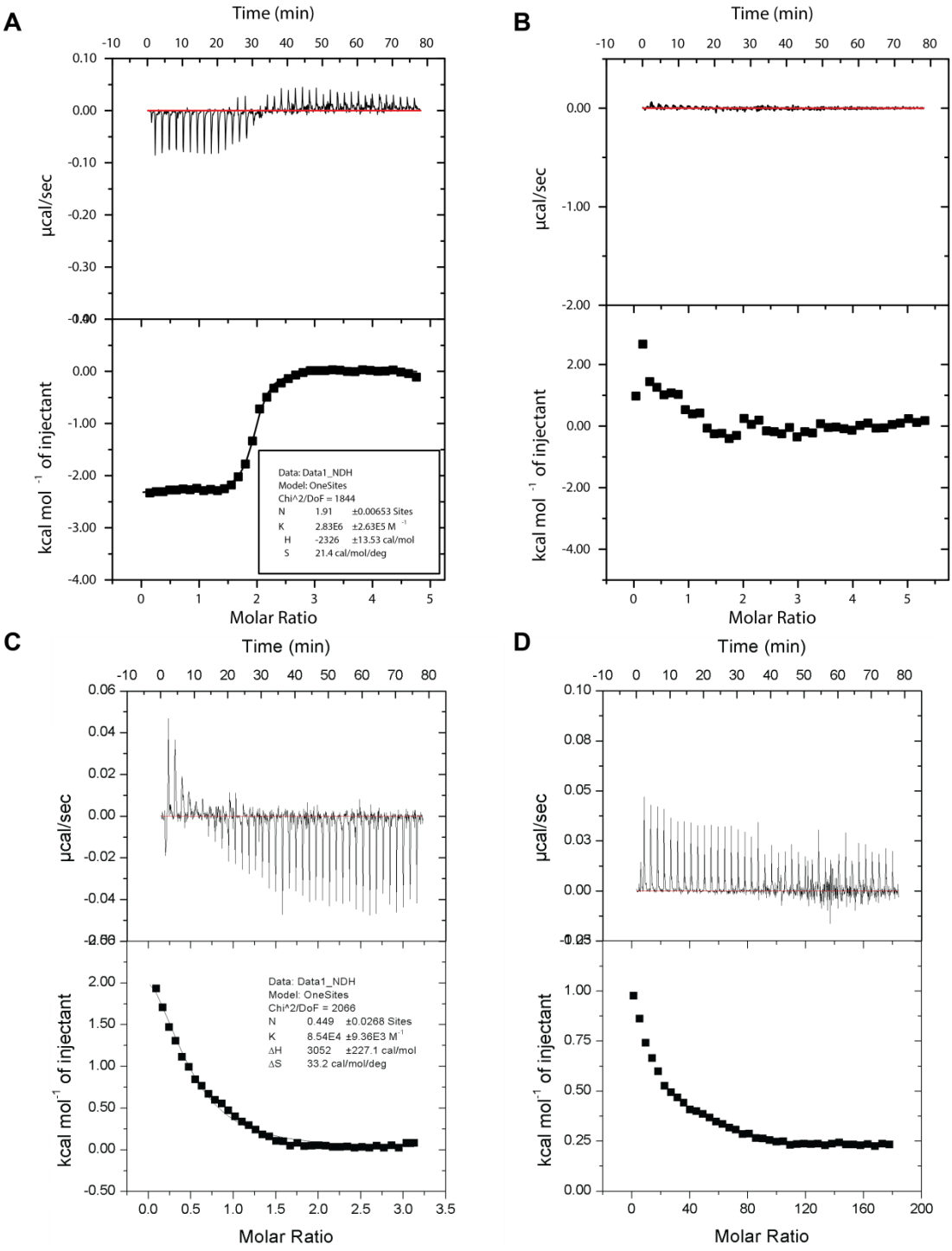


Figure 7 ITC measurements of FIH-ankyrin interactions. A A solution of 1CA peptide (1 mM) was added to FIH (50 µM) in the ITC cell. **B**,**C** A solution of FIH (800 µM) was added to **B** 3CA (50 µM), **C** iASPP (50 µM). **D** No ankyrin control, FIH (800 µM) was added to sample chamber in the absence of any ARD proteins. All measurements were carried out at 15 °C.

6.2.3 FIH mediated hydroxylation of recombinant iASPP and ASPP2

Following from the competition data, where iASPP and ASPP2 could compete with HIF-α peptides for FIH, it was investigated if recombinant forms of $ASPP2_{889-1128}$ and $iASPP_{625-828}$ were accepted by FIH as substrates.

Sequence analysis of ASPP2 and iASPP using the ExPASy web server[57], suggests that the FIH hydroxylation site for both proteins is not readily accessible for MS-MS analysis with trypsin digest. The ankyrin repeats do not contain any lysine or arginine residues (after which trypsin preferably cleaves the peptides) in direct proximity to the hydroxylation site. This would lead to long peptides after digestion (>30 amino acids in ASPP2 and >80 amino acids in iASPP), making the MS-MS analysis challenging. Attempts to substitute trypsin for chymotrypsin[58], which was predicted by the ExPASy web server (see above) to yield shorter peptides covering the site of interest, did not give the desired results. Optimisation attempts (*i.e*. changing buffer, reaction temperature and protease concentration) did not improve coverage of the hydroxylation site in the MS/MS (data not shown), thus an alternative approach was chosen. Instead of using a specific protease, yielding predictable peptides, elastase (a less specific protease when compared to trypsin or chymotrypsin) was chosen.

The use of elastase as a protease for MS/MS analysis has the advantage that it results in various, partially overlapping peptides which lead to near complete sequence coverage given enough starting material. Elastase displays, compared to trypsin, a broader and partially selective cleavage preference[59,60]. The broader cleavage preference increase coverage over inaccessible or other difficult protein sequences but results in more diverse peptides which make reliable quantifications of peptide modifications challenging.

Recombinant forms of $ASPP2_{889-1128}$ and $iASPP_{625-828}$ were incubated with FIH and co-factors for 3 h at 37 °C. The reactions were quenched by protein precipitation followed by mass-spectrometry analysis. MS-MS, analysis of the ASPP proteins provided further evidence that both iASPP and ASPP2 are accepted by FIH as substrates (see Figure 9 and Figure 10). Surprisingly, a second unprecedented hydroxylation was observed at the -2 position to the canonical hydroxylation site (Figure 9 and Figure 10).

Using elastase, the quantity of double hydroxylation could not be determined; however, using spectral and peptide counts, the levels of double hydroxylation were estimated to be ~20-30% for ASPP2 and for iASPP ~10-15%.

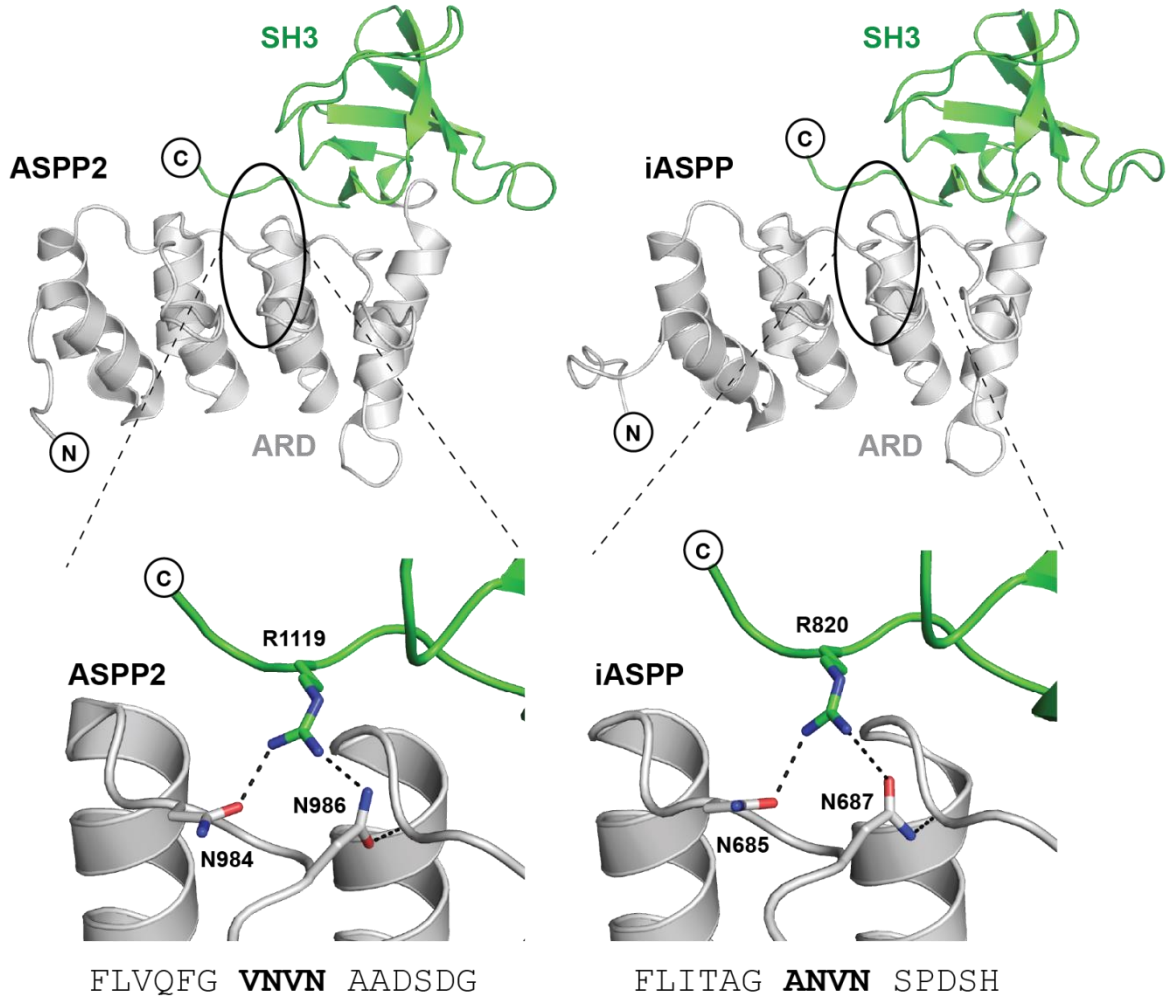
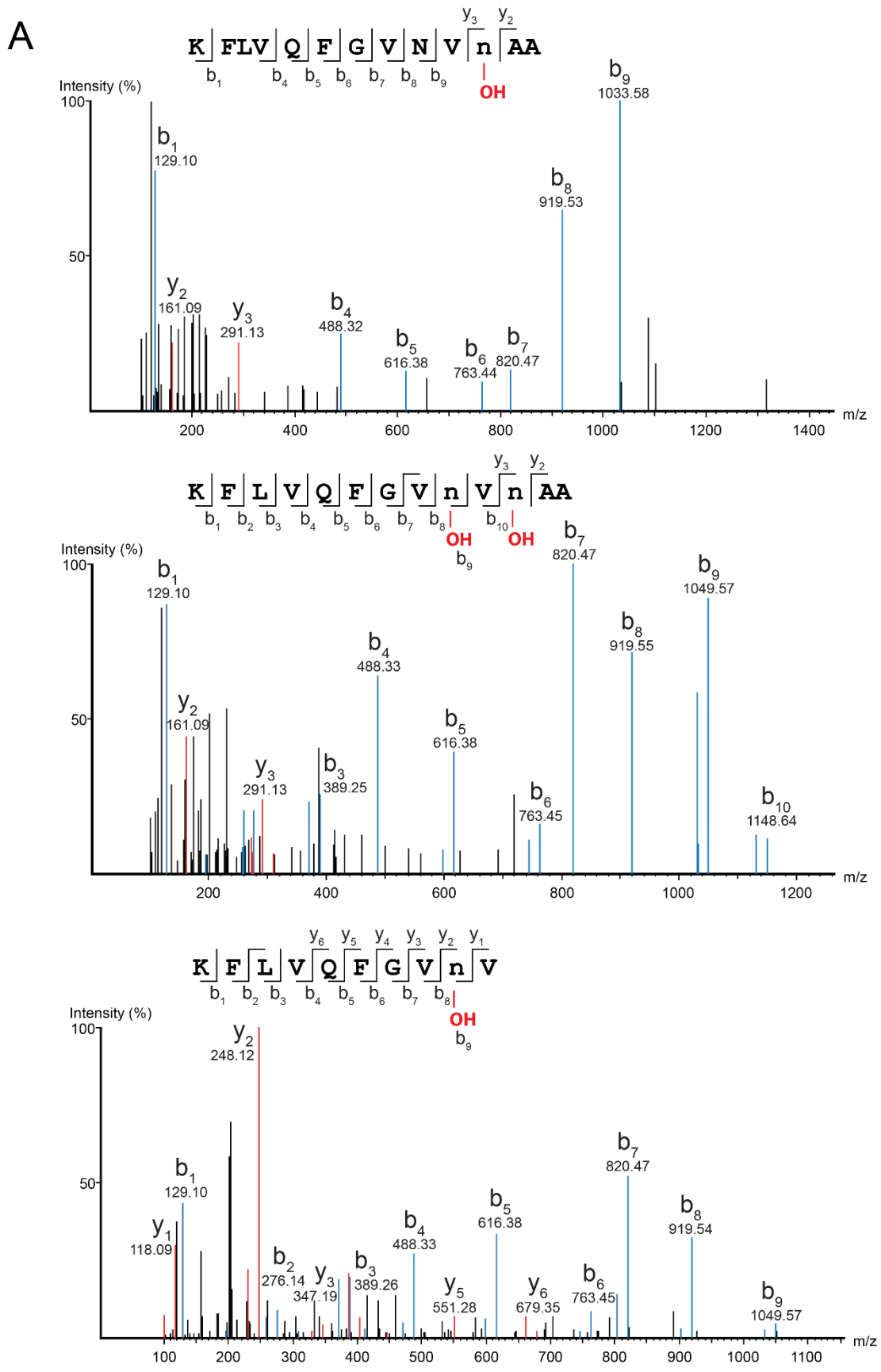


Figure 8 Positioning of the hydroxylation site of ASPP2 and iASPP. Views from crystal structures of ASPP2 (2.27 Å, PDB code 4A63[29]) and iASPP (2.10 Å, PDB code 2VGE) illustrate the positioning and intra molecular interactions of the hydroxylated asparagine residues within the C-terminus of ASPP2 and iASPP.

ASPP2 and iASPP have similar domain structure in their C-termini and manifest high sequence homology in their ARDs and adjacent SH3 domain (see also Figure 1). ASPP2 incorporates a “VNVN” motif with two valines preceding the two hydroxylated asparagines. iASPP has an “ANVN” motif with an alanine preceding the first asparagine on the non-canonical hydroxylation site (Figure 8). The -1 valine (with respect to the canonical hydroxylation sites) has previously been described as an important feature of FIH substrates (see also chapter 1) and the substitution of the -3 valine for alanine could potentially account for the differences observed between ASPP2 and iASPP with respect to double hydroxylation. Note that in both cases the hydroxylated residues appear to be of structural importance.



(Figure continues on next page)

B

K|FLV|Q|F|G|V|N|V|n|AA
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.09	112.08	65.05	K					13
2	276.17	258.16	259.14	138.59	F	1294.68	1276.67	1277.65	647.84	12
3	389.26	371.24	372.23	195.13	L	1147.61	1129.60	1130.58	574.31	11
4	488.32	470.31	471.30	244.66	V	1034.53	1016.52	1017.50	517.76	10
5	616.38	598.37	599.36	308.69	Q	935.46	917.45	918.43	468.23	9
6	763.44	745.44	746.42	382.23	F	807.40	789.39	790.37	404.20	8
7	820.47	802.46	803.45	410.74	G	660.33	642.32	643.30	330.67	7
8	919.53	901.53	902.51	460.27	V	603.31	585.30	586.28	302.15	6
9	1033.58	1015.57	1016.56	517.29	N	504.24	486.23	487.21	252.62	5
10	1132.65	1114.64	1115.62	566.83	V	390.20	372.19	373.17	195.60	4
11	1262.69	1244.68	1245.66	631.84	N(+15.99)	291.13	273.12	274.10	146.06	3
12	1333.73	1315.72	1316.70	667.36	A	161.09	143.08	144.07	81.05	2
13					A	90.05	72.04	73.03	45.53	1

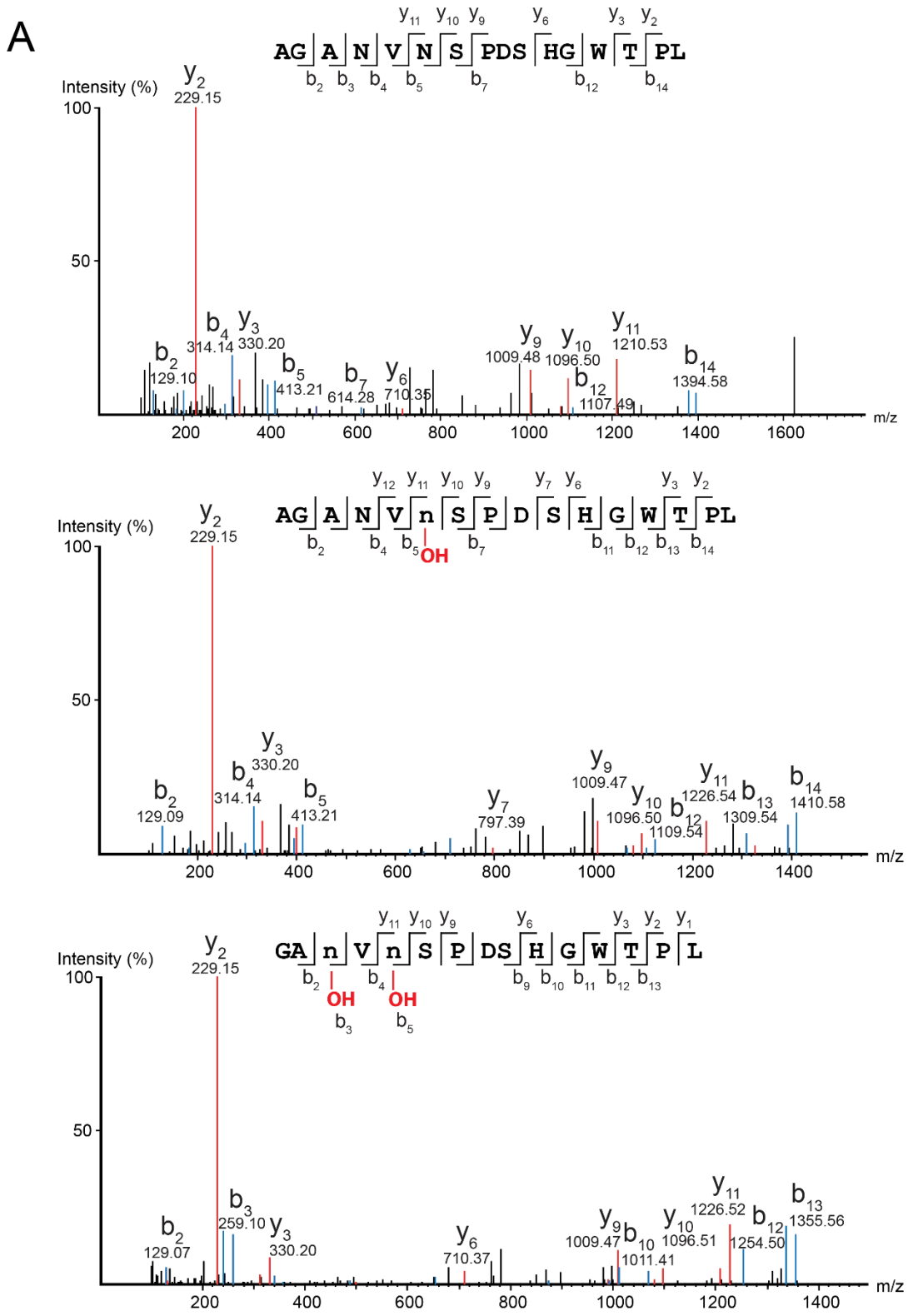
K|F|L|V|Q|F|G|V|n|V|n|AA
OH OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.09	112.08	65.05	K					13
2	276.17	258.16	259.11	138.59	F	1310.67	1292.66	1293.65	655.84	12
3	389.25	371.25	372.23	195.11	L	1163.61	1145.59	1146.58	582.30	11
4	488.33	470.31	471.30	244.66	V	1050.52	1032.51	1033.49	525.76	10
5	616.38	598.37	599.37	308.69	Q	951.45	933.44	934.43	476.23	9
6	763.45	745.43	746.42	382.23	F	823.39	805.38	806.37	412.20	8
7	820.47	802.46	803.45	410.74	G	676.33	658.32	659.30	338.66	7
8	919.55	901.53	902.51	460.27	V	619.30	601.29	602.28	310.14	6
9	1049.57	1031.56	1032.56	525.29	N(+15.99)	520.24	502.23	503.21	260.62	5
10	1148.64	1130.61	1131.62	574.82	V	390.20	372.19	373.17	195.60	4
11	1278.68	1260.67	1261.66	639.84	N(+15.99)	291.13	273.12	274.10	146.06	3
12	1349.72	1331.71	1332.69	675.36	A	161.09	143.08	144.07	81.05	2
13					A	90.05	72.04	73.03	45.53	1

K|F|L|V|Q|F|G|V|n|V
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.09	112.08	65.05	K					10
2	276.14	258.16	259.11	138.59	F	1038.56	1020.55	1021.53	519.78	9
3	389.26	371.25	372.23	195.11	L	891.49	873.48	874.47	446.28	8
4	488.33	470.31	471.30	244.66	V	778.41	760.40	761.38	389.70	7
5	616.38	598.37	599.36	308.70	Q	679.35	661.33	662.35	340.17	6
6	763.45	745.45	746.43	382.23	F	551.28	533.27	534.26	276.14	5
7	820.47	802.46	803.45	410.74	G	404.23	386.21	387.20	202.61	4
8	919.54	901.53	902.51	460.27	V	347.19	329.18	330.21	174.10	3
9	1049.57	1031.56	1032.55	525.29	N(+15.99)	248.12	230.11	231.10	124.56	2
10					V	118.09	100.08	101.07	59.54	1

Figure 9 FIH mediated hydroxylation of recombinant ASPP2 protein. The results provide evidence that the C-terminus of recombinant ASPP2 is hydroxylated at two asparagines within the same ankyrin repeat. Reaction conditions: 1 µM ASPP2, 0.5 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), 50 mM Tris-HCl pH = 7.5 incubated for 3 h at 37 °C.



(Figure continues on next page)

B

AG|A|N|V|N|S|PDS|HG|W|T|PL

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	72.04	54.03	55.02	36.52	A					16
2	129.10	111.06	112.04	65.03	G	1551.72	1533.71	1534.69	776.36	15
3	200.10	182.09	183.11	100.55	A	1494.70	1476.69	1477.67	747.85	14
4	314.14	296.14	297.12	157.57	N	1423.66	1405.65	1406.63	712.33	13
5	413.21	395.20	396.19	207.11	V	1309.62	1291.61	1292.59	655.31	12
6	527.26	509.25	510.22	264.13	N	1210.53	1192.54	1193.52	605.77	11
7	614.28	596.28	597.26	307.64	S	1096.50	1078.50	1079.46	548.75	10
8	711.34	693.33	694.32	356.17	P	1009.48	991.46	992.45	505.24	9
9	826.37	808.36	809.34	413.68	D	912.42	894.41	895.39	456.71	8
10	913.40	895.39	896.37	457.20	S	797.39	779.38	780.37	399.20	7
11	1050.46	1032.45	1033.43	525.73	H	710.35	692.35	693.34	355.68	6
12	1107.49	1089.47	1090.45	554.24	G	573.30	555.29	556.28	287.15	5
13	1293.56	1275.55	1276.53	647.28	W	516.28	498.27	499.25	258.64	4
14	1394.58	1376.58	1377.58	697.80	T	330.20	312.19	313.18	165.60	3
15	1491.66	1473.65	1474.63	746.33	P	229.15	211.14	212.13	115.08	2
16					L	132.10	114.09	115.07	66.55	1

AG|A|N|V|n(OH)|S|P|D|S|H|G|W|T|PL

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	72.04	54.03	55.02	36.52	A					16
2	129.09	111.06	112.04	65.03	G	1567.71	1549.70	1550.69	784.36	15
3	200.10	182.09	183.11	100.55	A	1510.69	1492.68	1493.67	755.85	14
4	314.14	296.14	297.12	157.57	N	1439.65	1421.64	1422.63	720.33	13
5	413.21	395.20	396.19	207.11	V	1325.62	1307.60	1308.59	663.31	12
6	543.25	525.24	526.23	272.13	N(+15.99)	1226.54	1208.53	1209.52	613.77	11
7	630.28	612.27	613.26	315.64	S	1096.49	1078.50	1079.44	548.75	10
8	727.34	709.33	710.35	364.17	P	1009.47	991.46	992.45	505.24	9
9	842.36	824.35	825.34	421.68	D	912.42	894.41	895.39	456.71	8
10	929.40	911.39	912.37	465.20	S	797.39	779.38	780.37	399.20	7
11	1066.45	1048.44	1049.43	533.73	H	710.35	692.35	693.34	355.68	6
12	1123.47	1105.47	1106.46	562.24	G	573.30	555.29	556.28	287.15	5
13	1309.54	1291.55	1292.53	655.24	W	516.28	498.27	499.25	258.64	4
14	1410.58	1392.59	1393.57	705.80	T	330.20	312.19	313.18	165.60	3
15	1507.66	1489.65	1490.63	754.33	P	229.15	211.14	212.13	115.08	2
16					L	132.10	114.09	115.07	66.55	1

GA|n(OH)|V|n(OH)|S|P|DS|H|G|W|T|P|L

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	58.03	40.02	41.00	29.51	G					15
2	129.07	111.06	112.04	65.03	A	1526.69	1508.68	1509.66	763.84	14
3	259.10	241.09	242.08	130.05	N(+15.99)	1455.65	1437.64	1438.62	728.32	13
4	358.17	340.16	341.15	179.59	V	1325.61	1307.60	1308.59	663.31	12
5	488.21	470.20	471.18	244.61	N(+15.99)	1226.52	1208.51	1209.49	613.77	11
6	575.24	557.23	558.22	288.12	S	1096.51	1078.50	1079.46	548.75	10
7	672.30	654.25	655.27	336.65	P	1009.47	991.47	992.45	505.24	9
8	787.32	769.31	770.30	394.16	D	912.42	894.41	895.39	456.71	8
9	874.34	856.34	857.33	437.68	S	797.39	779.38	780.37	399.20	7
10	1011.41	993.40	994.39	506.21	H	710.37	692.35	693.34	355.68	6
11	1068.43	1050.42	1051.41	534.72	G	573.30	555.29	556.28	287.15	5
12	1254.50	1236.50	1237.49	627.76	W	516.28	498.27	499.23	258.64	4
13	1355.56	1337.55	1338.53	678.28	T	330.20	312.17	313.18	165.60	3
14	1452.61	1434.60	1435.59	726.81	P	229.15	211.14	212.13	115.08	2
15					L	132.10	114.09	115.07	66.55	1

Figure 10 FIH mediated hydroxylation of recombinant iASPP protein. The results provide further evidence that the C-terminus of recombinant iASPP is hydroxylated at two asparagines within the same ankyrin repeat. Reaction conditions 1 μM iASPP, 0.5 μM FIH, 1 mM ascorbate, 1 mM 2OG, 200 μM, Fe(II) 50 mM Tris-HCl pH = 7.5 incubated for 3 h at 37 °C.

6.2.4 FIH mediated hydroxylation of ASPP proteins in U2OS cells:

Following from the *in vitro* hydroxylation assays of ASPP2 and iASPP with FIH (see above), investigations on whether ASPP family members could be hydroxylated in an FIH-dependent manner in human cells, were carried out. Vectors encoding for V5-tagged versions (see also chapter 7) of the members of the ASPP family (ASPP1, ASPP2 and iASPP) were transfected into U2OS WT or co-transfected into U2OS FIH KO cells with either empty pEF6 vector, FIH WT, FIH D201A or FIH L340R (Figure 11). ASPP proteins were immunoprecipitated and analysed by LC-MS/MS. Neither of the ASPP proteins showed significant hydroxylation in either the FIH KO cells co-transfected with empty vector or when transfected with the catalytically inactive FIH D201A mutant in the LC-MS/MS analysis (Figure 11). The endogenous levels of FIH in U2OS WT cells were sufficient to hydroxylate significant amounts of the ASPP proteins at one site. In the sample with overexpressed FIH WT, evidence for double hydroxylation could be detected for all three ASPPs. Surprisingly, FIH L340R, showed similar hydroxylation activity compared to the FIH WT protein. This result is in accordance with previously observed cellular interaction studies of FIH with ASPP proteins (Figure 3).

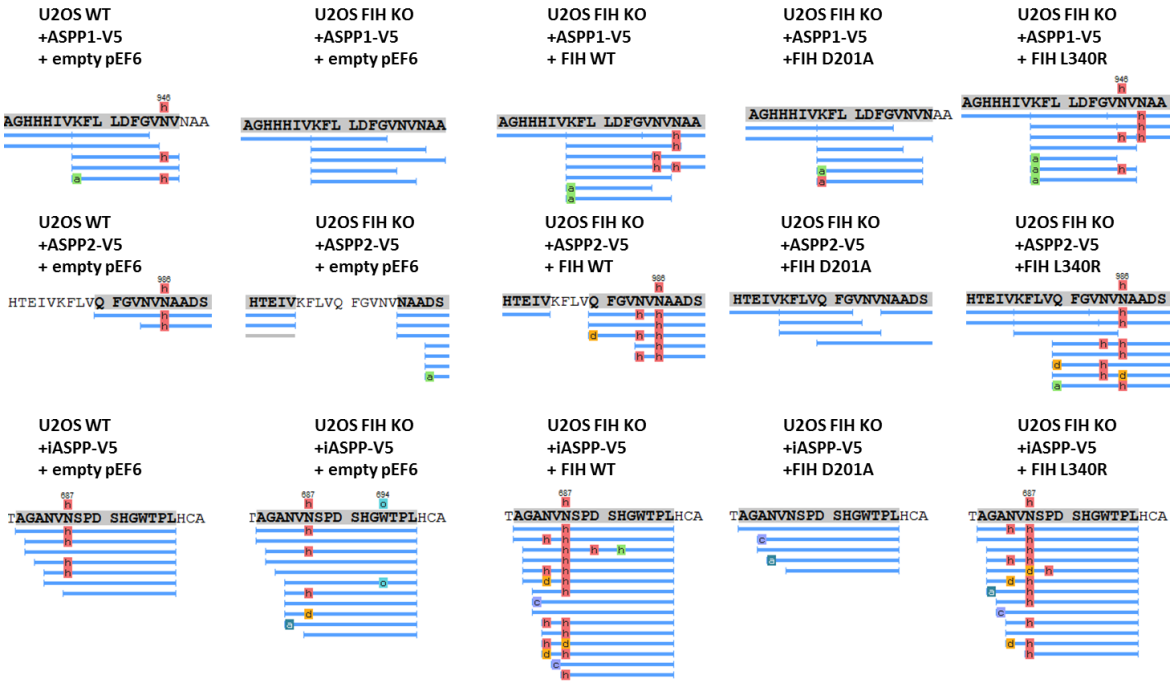
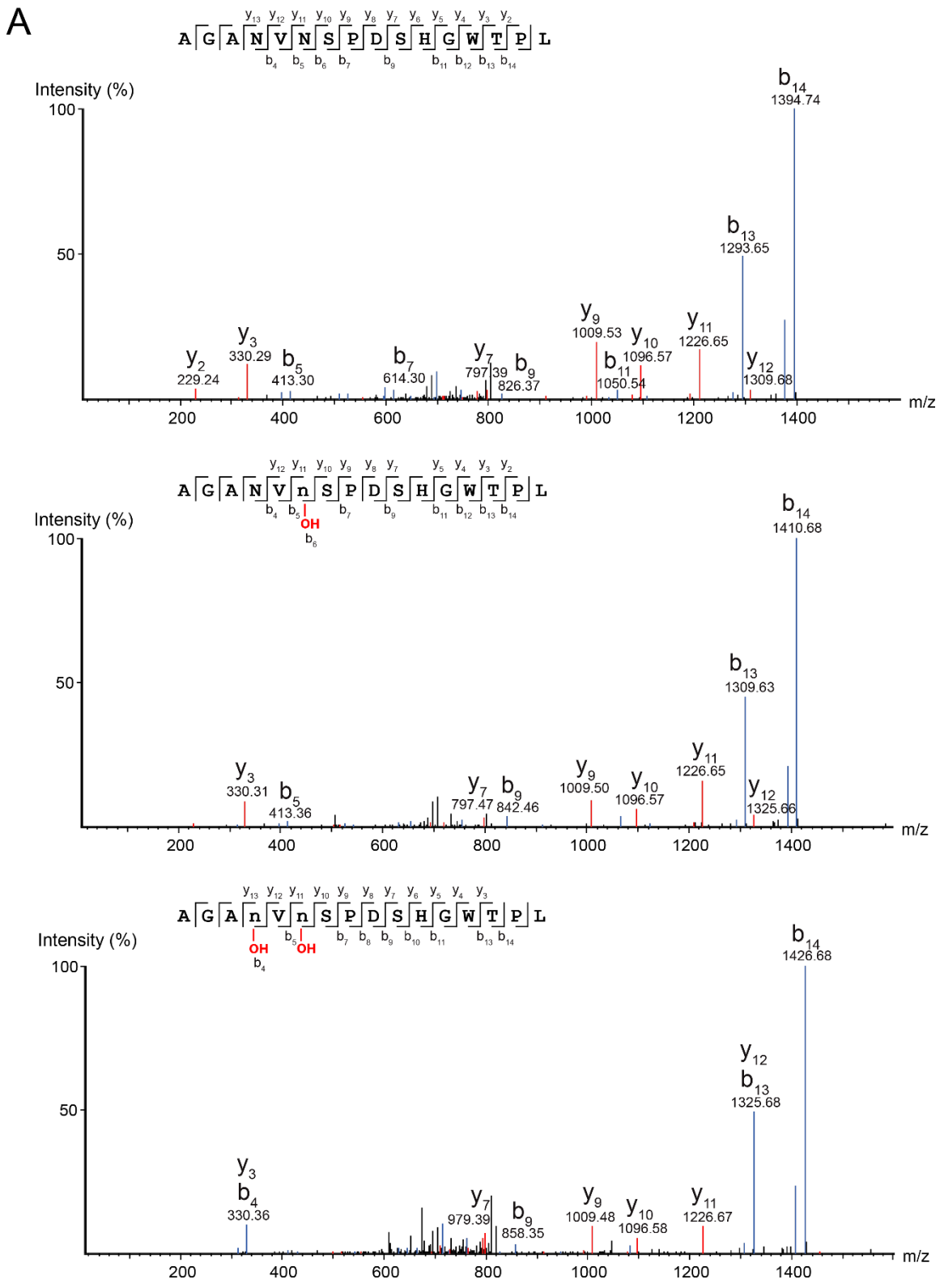


Figure 11 Overview of FIH mediated hydroxylation of V5-tagged ASPP proteins in U2OS cells. V5-tagged ASPP2 was overexpressed in U2OS cells and U2OS FIH KO cells which were co-overexpressed with empty pEF6 vector, FIH WT, FIH D201A and FIH L340R. ASPP2 was immunoprecipitated with anti-V5 antibody and the protein digested with Elastase. Blue lines indicate identified peptides of the protein. Red boxes mark annotated hydroxylation sites. Data was analysed using PEAKS7.5 proteomic software.

The overall peptide coverage results for overexpression of ASPP1-V5, ASPP2-V5 and iASPP5-V5 in U2OS cells were very high (>85%) but only a few peptides covered the hydroxylation site, probably due to the lack of cleavage specificity of elastase (see above).

It is notable that evidence for the second hydroxylation (N685 in iASPP, N946 in ASPP1 and N984 in ASPP2) was only detected in the samples with overexpressed FIH. This might suggest a dependency of the second hydroxylation on FIH levels. Since the ASPP proteins have been overexpressed in U2OS cells, the effect of endogenous FIH on hydroxylation of the ASPP proteins at endogenous levels was of interest. Several attempts to investigate endogenous ASPP2 hydroxylation levels were performed.

ASPP2 was immunoprecipitated, PAGE performed, and bands corresponding to ASPP2 were excised from the gel and subject to in-gel digestion with chymotrypsin or elastase, followed by LC-MS/MS analysis (see chapter 10.5 for a detailed protocol). The observed peptides from the digestion did not manifest coverage over the proposed hydroxylation site (for optimisation of endogenous ASPP2-IP see chapter 7). To overcome the issues of incomplete in-gel digestion and minimise material loss on the gel, in-solution digestion was attempted. The anti-ASPP2 antibodies available showed some level of unspecific binding and low affinity. Thus, a high amount of antibody was used (see also chapter 7). This can be an issue for the in-solution digest as peptides from the antibody can mask the proteins of interest. To avoid antibody contamination, the ASPP2 antibody was cross-linked to beads (for cross-linking optimization see chapter 7). ASPP2 was immunoprecipitated from U2OS cell lysate using anti-ASPP2 antibody cross-linked beads and the precipitated proteins were subject to in solution digest using elastase. Although antibody contamination was reduced, contaminants in the IP masked most of the ASPP2 peptides and no coverage over the hydroxylation site could be achieved in the LC-MS/MS (data not shown).



(Figure continues on next page)

B

A G A N V N S P D S H G W T P L

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	72.04	54.03	55.02	36.52	A					16
2	129.07	111.06	112.04	65.03	G	1551.72	1533.71	1534.69	776.42	15
3	200.10	182.09	183.08	100.55	A	1494.70	1476.69	1477.67	747.87	14
4	314.37	296.25	297.34	157.57	N	1423.81	1405.65	1406.63	712.50	13
5	413.30	395.29	396.27	207.11	V	1309.68	1291.61	1292.32	655.31	12
6	527.35	509.29	510.25	264.13	N	1210.65	1192.64	1193.52	605.77	11
7	614.30	596.35	597.31	307.64	S	1096.60	1078.57	1079.48	548.75	10
8	711.34	693.33	694.49	356.17	P	1009.53	991.57	992.45	505.24	9
9	826.37	808.36	809.34	413.68	D	912.51	894.41	895.32	456.71	8
10	913.40	895.32	896.63	457.20	S	797.39	779.38	780.37	399.20	7
11	1050.54	1032.68	1033.55	525.73	H	710.48	692.35	693.34	355.68	6
12	1107.54	1089.47	1090.45	554.24	G	573.27	555.26	556.42	287.15	5
13	1293.65	1275.61	1276.64	647.54	W	516.45	498.33	499.25	258.64	4
14	1394.74	1376.72	1377.58	698.05	T	330.29	312.32	313.18	165.60	3
15	1491.66	1473.65	1474.63	746.58	P	229.24	211.14	212.13	115.08	2
16					L	132.10	114.09	115.07	66.55	1

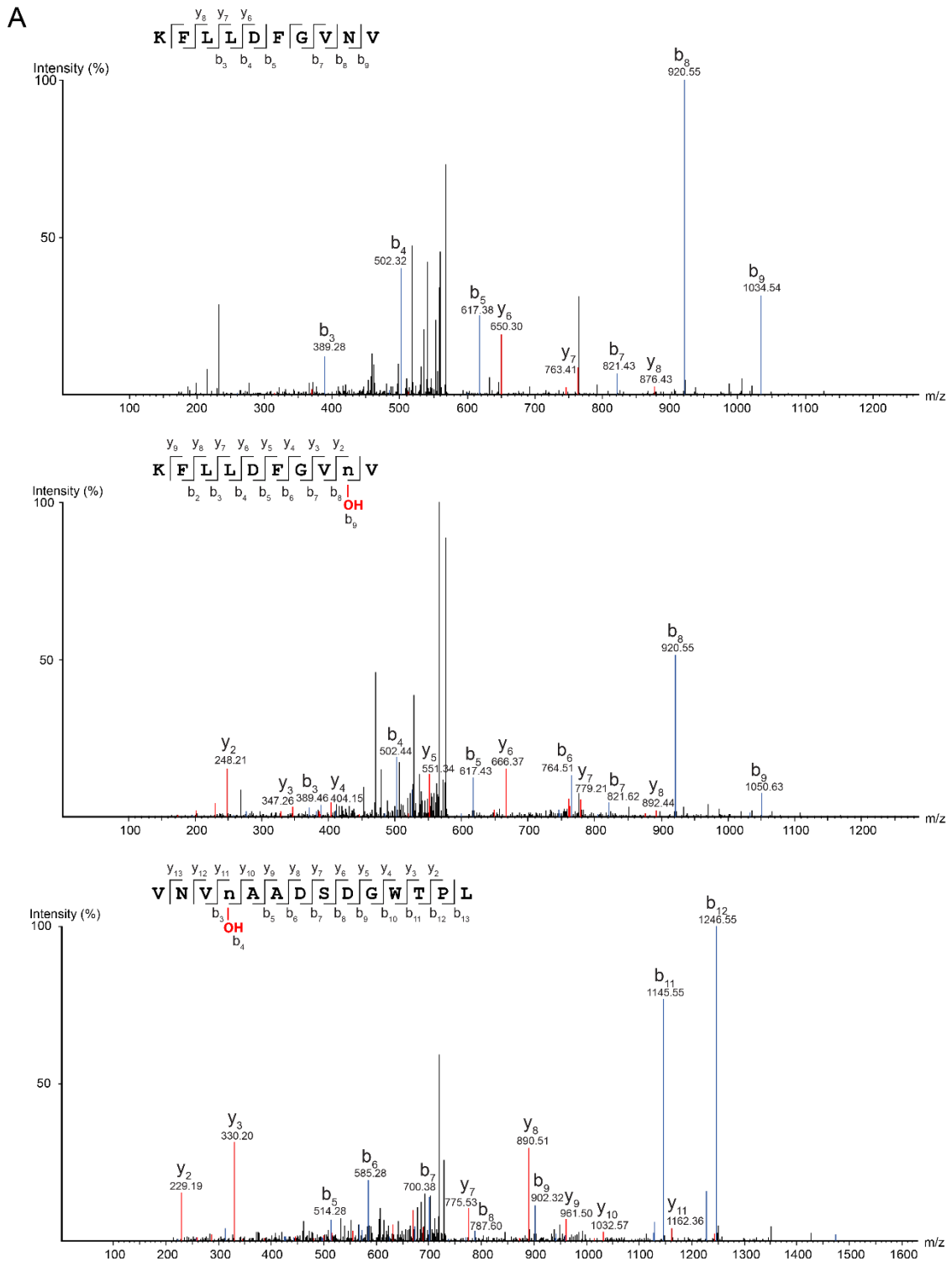
A G A N V n S P D S H G W T P L
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	72.04	54.03	55.02	36.52	A					16
2	129.07	111.06	112.04	65.03	G	1567.71	1549.70	1550.69	784.25	15
3	200.10	182.09	183.08	100.55	A	1510.69	1492.68	1493.67	755.83	14
4	314.24	296.14	297.12	157.57	N	1439.65	1421.73	1422.63	720.29	13
5	413.36	395.23	396.26	207.11	V	1325.66	1307.82	1308.59	663.31	12
6	543.15	525.24	526.17	272.13	N(+15.99)	1226.65	1208.64	1209.52	613.77	11
7	630.27	612.27	613.26	315.64	S	1096.57	1078.50	1079.48	548.75	10
8	727.34	709.33	710.31	364.17	P	1009.50	991.53	992.45	505.17	9
9	842.46	824.35	825.52	421.68	D	912.56	894.23	895.54	456.71	8
10	929.40	911.39	912.56	465.47	S	797.47	779.54	780.37	399.20	7
11	1066.48	1048.44	1049.43	533.73	H	710.36	692.35	693.36	355.68	6
12	1123.58	1105.47	1106.55	562.24	G	573.40	555.29	556.45	287.15	5
13	1309.63	1291.52	1292.71	655.37	W	516.49	498.42	499.25	258.64	4
14	1410.68	1392.69	1393.58	705.80	T	330.31	312.24	313.18	165.60	3
15	1507.66	1489.65	1490.63	754.52	P	229.21	211.14	212.13	115.08	2
16					L	132.10	114.09	115.07	66.55	1

A G A n V n S P D S H G W T P L
OH OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	72.04	54.03	55.02	36.52	A					16
2	129.07	111.06	112.04	65.03	G	1583.71	1565.70	1566.68	792.31	15
3	200.10	182.09	183.08	100.55	A	1526.69	1508.68	1509.82	763.86	14
4	330.36	312.33	313.11	165.57	N(+15.99)	1455.70	1437.64	1438.62	728.49	13
5	429.32	411.37	412.18	215.10	V	1325.68	1307.69	1308.59	663.41	12
6	559.25	541.36	542.38	280.12	N(+15.99)	1226.67	1208.75	1209.58	613.77	11
7	646.38	628.36	629.25	323.64	S	1096.58	1078.61	1079.48	548.75	10
8	743.31	725.32	726.41	372.17	P	1009.48	991.47	992.29	505.34	9
9	858.35	840.52	841.33	429.68	D	912.22	894.41	895.39	456.71	8
10	945.47	927.35	928.53	473.20	S	797.39	779.38	780.37	399.18	7
11	1082.52	1064.44	1065.57	541.73	H	710.31	692.38	693.57	355.68	6
12	1139.47	1121.38	1122.44	570.30	G	573.21	555.28	556.28	287.15	5
13	1325.68	1307.69	1308.52	663.41	W	516.53	498.32	499.30	258.64	4
14	1426.68	1408.66	1409.57	713.97	T	330.36	312.33	313.18	165.60	3
15	1523.65	1505.64	1506.62	762.57	P	229.15	211.14	212.13	115.08	2
16					L	132.10	114.09	115.07	66.55	1

Figure 12 FIH mediated hydroxylation of iASPP-V5 in U2OS cells. **A** MS-MS spectra indicating unmodified iASPP peptides, singly hydroxylated peptides at asparagine-687 and double hydroxylated peptides at asparagine-687 and asparagine-685. **B** Ion tables from the spectra above. Experimental Conditions: iASPP-V5 in pcDNA3. The vectors were transfected and overexpressed for 24 h in U2OS and U2OS FIH CRISPR KO cells and immunoprecipitated with anti-V5 antibody. Proteins were run on SDS-PAGE and the band correspondent to iASPP excised and the protein digested using elastase, 16 h at 37 °C.



(Figure continues on next page)

B

K F L L D F G V N V

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.09	112.08	65.05	K					10
2	276.17	258.16	259.19	138.59	F	1023.55	1005.54	1006.52	512.24	9
3	389.28	371.24	372.23	195.13	L	876.43	858.47	859.46	438.74	8
4	502.32	484.38	485.31	251.67	L	763.41	745.39	746.37	382.20	7
5	617.38	599.36	600.33	309.18	D	650.30	632.30	633.29	325.66	6
6	764.43	746.42	747.41	382.72	F	535.29	517.28	518.26	268.14	5
7	821.43	803.45	804.40	411.23	G	388.22	370.24	371.19	194.61	4
8	920.55	902.51	903.50	460.76	V	331.20	313.19	314.20	166.10	3
9	1034.54	1016.59	1017.54	517.78	N	232.13	214.12	215.10	116.56	2
10					V	118.09	100.08	101.06	59.54	1

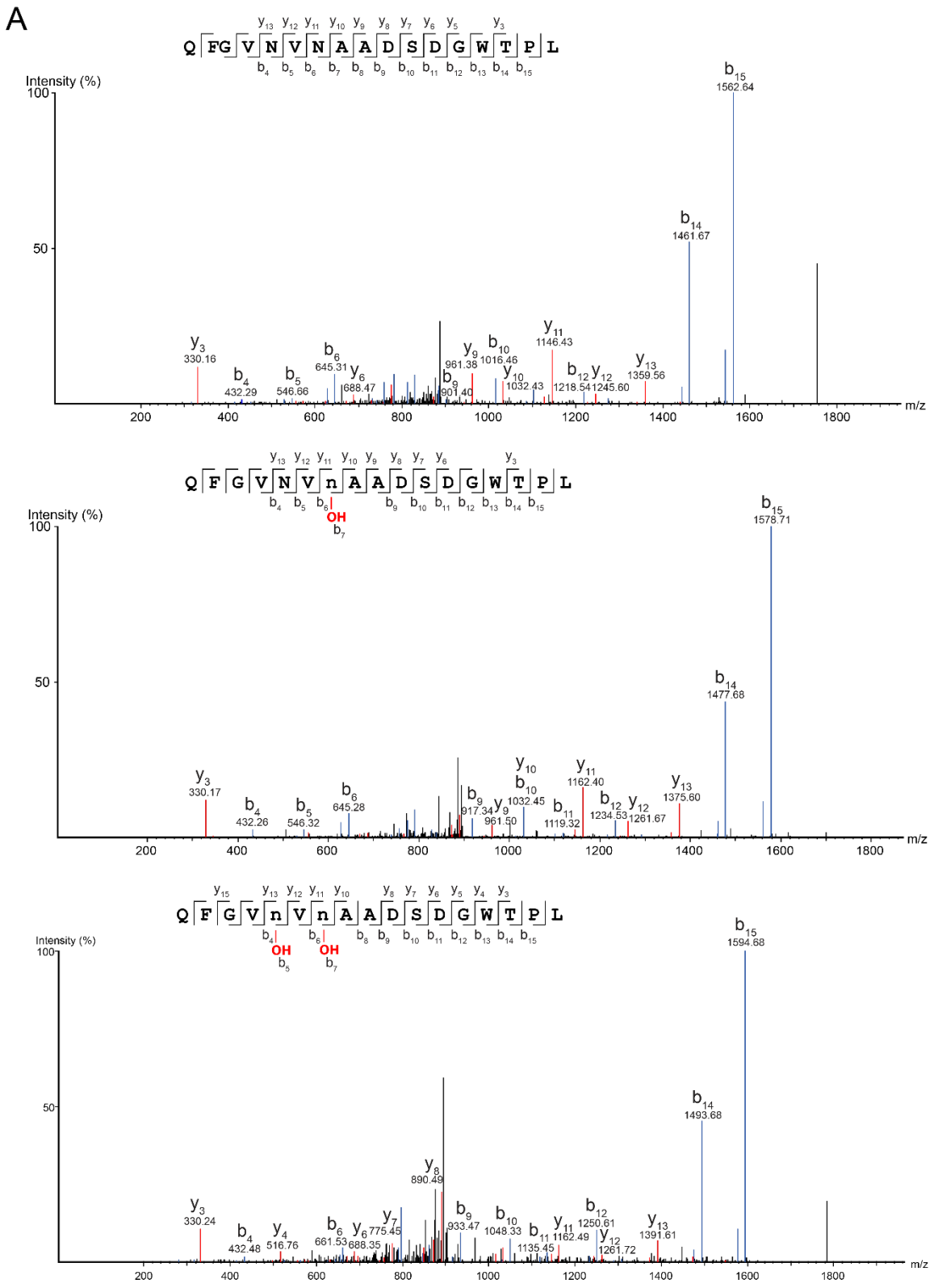
K F L L D F G V n V
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.10	111.09	112.08	65.05	K					10
2	276.34	258.29	259.14	138.59	F	1039.55	1021.54	1022.52	520.72	9
3	389.46	371.34	372.23	195.59	L	892.44	874.43	875.38	446.83	8
4	502.44	484.41	485.31	251.67	L	779.21	761.70	762.37	390.20	7
5	617.43	599.20	600.20	309.18	D	666.37	648.78	649.28	333.65	6
6	764.51	746.48	747.78	382.26	F	551.34	533.27	534.26	276.34	5
7	821.62	803.45	804.57	411.21	G	404.15	386.17	387.48	202.30	4
8	920.55	902.51	903.61	460.76	V	347.26	329.25	330.17	174.34	3
9	1050.63	1032.54	1033.32	525.71	N(+15.99)	248.21	230.19	231.10	124.56	2
10					V	118.09	100.08	101.06	59.54	1

V N V n A A D S D G W T P L
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	100.08	82.07	83.05	50.54	V					14
2	214.12	196.11	197.09	107.56	N	1375.61	1357.78	1358.59	688.13	13
3	313.27	295.18	296.16	157.09	V	1261.57	1243.56	1244.54	631.34	12
4	443.32	425.23	426.34	222.42	N(+15.99)	1162.36	1144.49	1145.55	581.75	11
5	514.28	496.28	497.56	257.36	A	1032.57	1013.99	1015.44	516.73	10
6	585.28	567.50	568.37	293.15	A	961.50	943.42	944.24	481.34	9
7	700.38	682.32	683.30	350.66	D	890.51	872.38	872.91	445.41	8
8	787.60	769.20	770.35	394.18	S	775.53	757.35	758.34	387.91	7
9	902.32	884.38	885.36	451.69	D	688.13	670.17	671.30	345.15	6
10	959.23	940.98	942.26	480.20	G	573.41	555.23	556.24	287.24	5
11	1145.55	1127.60	1128.49	573.41	W	516.21	498.32	499.27	258.27	4
12	1246.55	1228.57	1229.51	624.13	T	330.20	312.48	313.27	165.60	3
13	1343.59	1325.58	1326.56	672.69	P	229.19	211.14	212.13	115.08	2
14					L	132.10	114.09	115.07	66.55	1

Figure 13 FIH mediated hydroxylation of ASPP1-V5 in U2OS cells. **A** MS-MS spectra indicating un-modified ASPP1 peptides, singly hydroxylated peptides at asparagine-946 and at asparagine-948. **B** Ion tables from the spectra above. Experimental Conditions: ASPP1-V5 in pcDNA3. The vectors were transfected and overexpressed for 24 h in U2OS FIH CRISPR KO cells and immunoprecipitated with anti-V5 antibody. Proteins were run on SDS-PAGE and the band correspondent to ASPP1 excised and the protein digested using elastase, 16 h at 37 °C.



(Figure continues on next page)

B

Q|FG|V|N|V|N|A|A|D|S|D|G|W|T|P|L

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.07	111.06	112.04	65.03	Q					17
2	276.13	258.12	259.11	138.57	F	1662.78	1644.77	1645.75	832.19	16
3	333.16	315.29	316.13	167.08	G	1515.71	1497.70	1498.68	758.35	15
4	432.29	414.31	415.27	216.61	V	1458.69	1440.72	1441.66	729.73	14
5	546.66	528.26	529.15	273.63	N	1359.56	1341.72	1342.59	680.31	13
6	645.31	627.25	628.31	323.17	V	1245.60	1227.59	1228.80	623.33	12
7	759.48	741.37	742.41	380.19	N	1146.43	1128.45	1129.48	573.27	11
8	830.29	812.41	813.09	415.27	A	1032.43	1014.45	1015.44	516.73	10
9	901.40	883.01	884.41	451.41	A	961.38	943.30	944.40	481.21	9
10	1016.46	998.47	999.45	508.74	D	890.44	872.39	873.13	445.69	8
11	1103.51	1085.53	1086.49	551.76	S	775.29	757.17	758.34	388.18	7
12	1218.54	1200.53	1201.03	609.77	D	688.47	670.32	671.12	344.67	6
13	1275.50	1257.55	1258.53	638.28	G	573.27	555.19	556.28	287.15	5
14	1461.67	1443.71	1444.61	731.59	W	516.28	498.06	498.90	258.64	4
15	1562.64	1544.63	1545.66	781.98	T	330.16	312.19	313.18	165.60	3
16	1659.74	1641.73	1642.71	830.29	P	229.15	211.14	212.13	115.08	2
17					L	132.10	114.09	115.07	66.55	1

Q|F|G|V|N|V|n|A|A|D|S|D|G|W|T|P|L
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.07	111.06	112.04	65.03	Q					17
2	276.13	258.12	259.11	138.57	F	1678.77	1660.76	1661.74	839.63	16
3	333.16	315.15	316.13	167.08	G	1531.70	1513.69	1514.68	766.28	15
4	432.26	414.22	415.56	216.61	V	1474.68	1456.59	1457.65	737.98	14
5	546.32	528.60	529.31	273.63	N	1375.60	1357.68	1358.59	688.18	13
6	645.28	627.23	628.29	323.17	V	1261.67	1243.56	1244.58	631.28	12
7	775.31	757.46	758.35	388.19	N(+15.99)	1162.40	1144.50	1145.61	581.75	11
8	846.41	828.40	829.56	423.71	A	1032.45	1014.58	1015.44	516.73	10
9	917.34	899.44	900.42	459.22	A	961.50	943.42	944.65	481.21	9
10	1032.45	1014.58	1015.45	516.74	D	890.31	872.01	873.73	445.69	8
11	1119.32	1101.50	1102.36	560.25	S	775.31	757.46	758.34	388.18	7
12	1234.53	1216.52	1217.70	617.31	D	688.18	670.38	671.30	345.11	6
13	1291.56	1273.54	1274.53	646.28	G	573.30	555.29	556.28	287.15	5
14	1477.68	1459.78	1460.72	739.60	W	516.28	498.27	499.25	258.64	4
15	1578.71	1560.67	1561.66	790.08	T	330.17	312.19	313.18	165.60	3
16	1675.74	1657.72	1658.71	838.18	P	229.15	211.14	212.13	115.08	2
17					L	132.10	114.09	115.07	66.55	1

Q|F|G|V|n|V|n|A|A|D|S|D|G|W|T|P|L
OH OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	129.07	111.06	112.04	65.03	Q					17
2	276.13	258.12	259.11	138.57	F	1694.77	1676.76	1677.74	848.28	16
3	333.16	315.36	316.13	167.08	G	1547.72	1529.69	1530.67	774.35	15
4	432.48	414.21	415.22	216.61	V	1490.68	1472.48	1473.85	746.11	14
5	562.26	544.25	545.24	281.13	N(+15.99)	1391.61	1373.69	1374.58	696.20	13
6	661.53	643.46	644.30	331.17	V	1261.72	1243.38	1244.54	631.07	12
7	791.43	773.36	774.34	396.18	N(+15.99)	1162.49	1144.65	1145.47	581.88	11
8	862.28	844.40	845.38	431.70	A	1032.40	1014.45	1015.09	516.76	10
9	933.47	915.34	916.06	467.22	A	961.43	943.42	944.40	481.21	9
10	1048.33	1030.55	1031.44	525.13	D	890.49	872.72	873.36	445.69	8
11	1135.45	1117.46	1118.47	568.25	S	775.45	757.42	758.34	388.18	7
12	1250.61	1232.81	1233.50	626.25	D	688.35	670.32	671.31	344.67	6
13	1307.68	1289.54	1290.52	654.28	G	573.34	555.29	556.41	287.15	5
14	1493.68	1475.89	1476.60	747.31	W	516.76	498.58	499.25	258.64	4
15	1594.68	1576.76	1577.65	797.63	T	330.24	312.19	313.18	165.60	3
16	1691.73	1673.72	1674.70	846.31	P	229.15	211.14	212.13	115.08	2
17					L	132.10	114.09	115.07	66.55	1

Figure 14 FIH mediated hydroxylation of ASPP2-V5 in U2OS cells. **A** MS-MS spectra of un-modified ASPP2 peptides, singly hydroxylated peptides at asparagine-986 and double hydroxylated peptides at asparagine-986 and asparagine-984. **B** Ion tables from the spectra above. Experimental Conditions: ASPP2-V5 in pcDNA3. The vectors were transfected and overexpressed for 24 h in U2OS and immunoprecipitated with anti-V5 antibody. Proteins were run on SDS-PAGE and the band correspondent to ASPP2 excised and the protein digested using elastase, 16 h at 37 °C.

6.2.5 FIH mediated hydroxylation of ASPP peptides

6.2.5.1 Characterisation of ASPP peptides as FIH substrates

ASPP1, ASPP2 and iASPP were all shown to be hydroxylated in an FIH dependent manner in U2OS cells. Of particular interest is the evidence for a second hydroxylation at the asparagine in the -2 position relative to the canonical hydroxylation site. To further investigate the potential double hydroxylation, peptides corresponding to iASPP (670-693), ASPP1 (932-954) and ASPP2 (969-991) were then tested as FIH substrates (Figure 15).

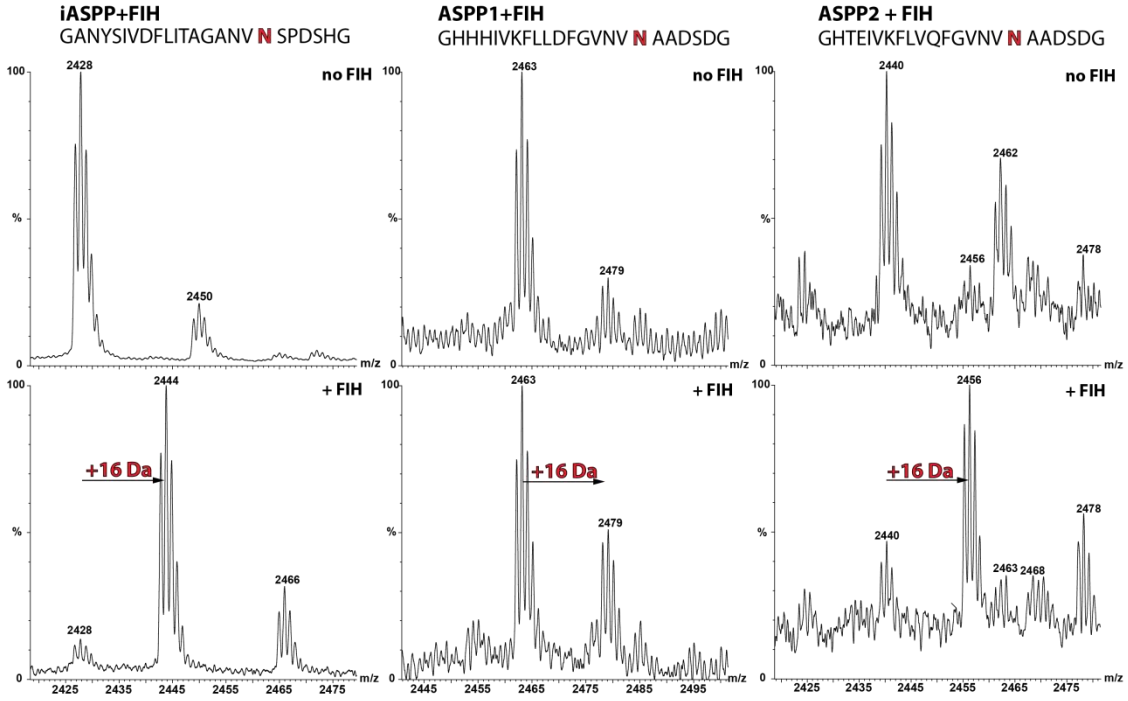
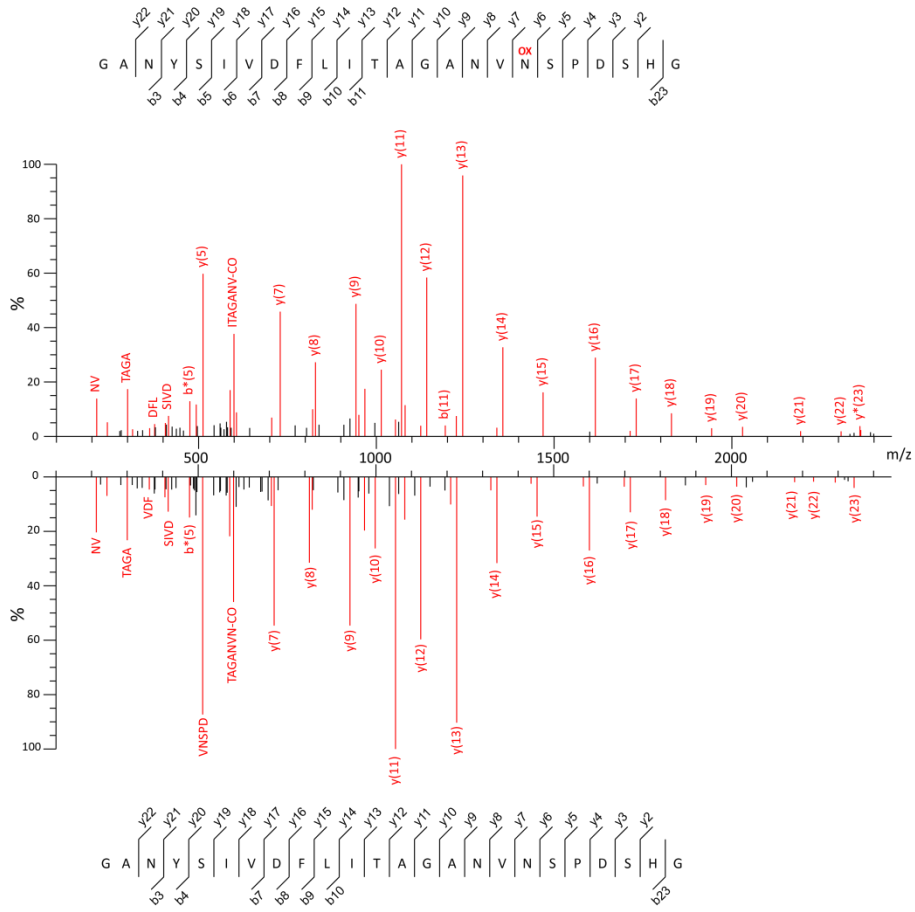


Figure 15 FIH mediated hydroxylation of ASPP peptides. End point MALDI-TOF based hydroxylation assays of an iASPP derived 24mer peptide (residues 670-693), an ASPP1 derived 23mer peptide (residues 932-954), and an ASPP2 derived 23mer peptide (residue 969-991). Reaction conditions: 50 µM peptide, 50 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), 50 mM Tris-HCl pH = 7.5 incubated for 30 min at 37 °C.

MALDI-MS analysis implies that all ASPP peptides are hydroxylated at one site in an FIH dependent manner (Figure 15). The hydroxylation site of iASPP (670-693) was confirmed by MALDI MS/MS. The results locate the hydroxylation site of the peptide to the residue corresponding to asparagine-687 in full-length iASPP (Figure 16). Asparagine-687 is located at the loop between ankyrin repeat 2 and ankyrin repeat 3 of the C-terminal domain and at a position corresponding to asparagine-986 of ASPP2 (Figure 8). Unlike the recombinant forms of iASPP and ASPP2, and the three full-length ASPP proteins in U2OS cells, none of the peptides showed evidence for double hydroxylation.



iASPP 24mer

#	b	Seq	y	#
1		G		24
2		A		23
3	243.11	N	2291.09	22
4	406.17	Y	2177.05	21
5		S	2013.99	20
6		I	1926.96	19
7	705.36	V	1813.872	18
8	820.38	D	1714.803	17
9	967.45	F	1599.776	16
10	1080.54	L	1452.708	15
11		I	1339.624	14
12		T	1226.54	13
13		A	1125.492	12
14		G	1054.455	11
15		A	997.4334	10
16		N	926.3962	9
17		V	812.3533	8
18		N	713.2849	7
19		S	599.242	6
20		P	512.21	5
21		D	415.1572	4
22		F	300.1302	3
23	2344.12	H	213.0982	2
24		G		1

iASPP-OH 24mer

#	b	Seq	y	#
1		G		24
2		A		23
3	243.11	N	2307.089	22
4	406.17	Y	2193.046	21
5	493.20	S	2029.983	20
6	606.29	I	1942.951	19
7	705.36	V	1829.866	18
8	820.38	D	1730.798	17
9	967.45	F	1615.771	16
10	1080.54	L	1468.703	15
11	1193.62	I	1355.619	14
12		T	1242.535	13
13		A	1141.487	12
14		G	1070.45	11
15		A	1013.428	10
16		N	942.3912	9
17		V	828.3482	8
18		N (OH)	729.2798	7
19		S	599.242	6
20		P	512.21	5
21		D	415.1572	4
22		F	300.1302	3
23	2360.12	H	213.0982	2
24		G		1

Figure 16 FIH mediated hydroxylation of ASPP peptides. MS-MS spectra from MALDI LIFT-TOF-TOF analysis of the FIH hydroxylation site of an iASPP peptide. Right: Ion tables from iASPP MS-MS spectra. Reaction conditions: 50 µM peptide, 50 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), 50 mM Tris-HCl pH = 7.5 incubated for 30 min at 37 °C.

To further investigate the binding mode of iASPP, ASPP1 and ASPP2 peptides to FIH, the peptides were co-crystallised with FIH and the crystal structures solved using molecular replacement (search model PDB code 1H2K[61], Figure 17). All of the peptides were found to bind to FIH in the crystalline state. A crystal structure of the FIH.Zn(II).NOG.iASPP (670-693) complex was solved to 1.97 Å, FIH.Zn(II).NOG.ASPP1 (932-954) to 2.03 Å, FIH.Zn(II).NOG.ASPP2 (969-991) to 1.89 Å (Table 1). Analysis of crystal structures of the FIH-ASPP complexes reveal a conserved binding mode of all three peptides to FIH, which was also observed for Notch1 and HIF-1α binding to FIH (Figure 17, for details on FIH substrate binding see chapter 1 and chapter 3).

An overall poor solubility of peptides derived from iASPP, ASPP1 and ASPP2 proved to be the limiting factor towards further investigations. In addition to the solubility problems, differences in the sequences of the ASPP proteins might also contribute to different reactivity with FIH. To investigate further whether the 'VNVN' motif was the only requirement for FIH mediated double hydroxylation of ankyrin repeats, the 'VNVN' motif was placed into the context of a consensus ankyrin peptide (the resulting peptides are labelled 1CA_VNVN, 1CA_ANVN, 1CA_VNAN; results with them and are described below).

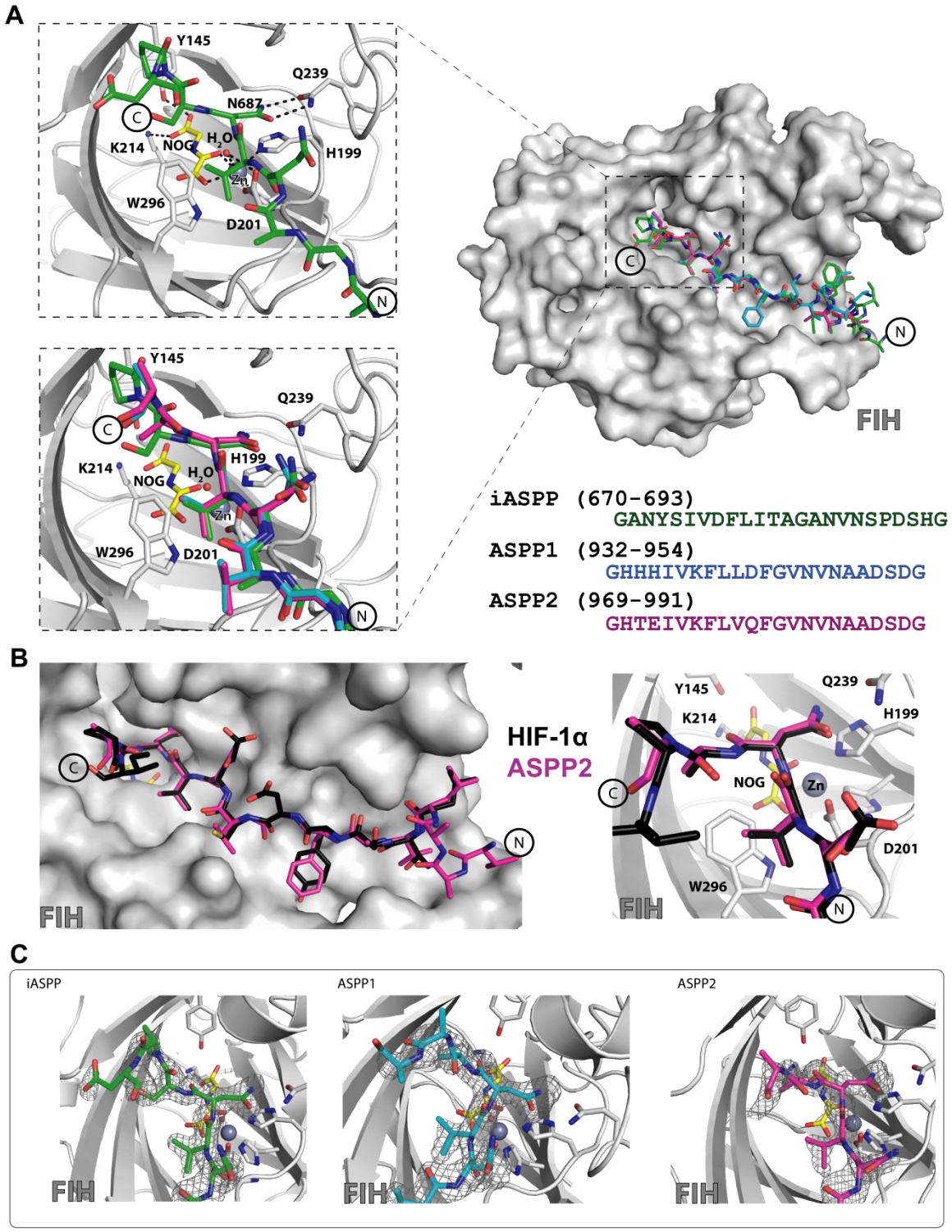


Figure 17 Views from crystal structures of FIH in complex with peptide sequences derived from all three human ASPP family members. A Overlay of the crystal structures of FIH in complex with ASPP1, ASPP2 and iASPP peptides, illustrating similarities in binding mode. **B** Views from a crystal structure of ASPP2 in complex with FIH overlayed with HIF-1α peptide (generated from PDB code 1H2K[61]) **C** Views from crystal structures of FIH in complex with ASPP1, ASPP2 and iASPP with mFo-Fc OMIT map contoured to 3σ.

Table 1: Data collection and refinement statistics of FIH complexes with ASPP peptides.

Datasets	**FIH**.Zn(II).NOG iASPP	**FIH**.Zn(II).NOG ASPP1	**FIH**.Zn(II).NOG ASPP2
Data Collection			
Beamline (Wavelength, Å)	DLS I04-1 (0.9159)	DLS I02 (0.9795)	DLS I03 (0.9793)
Detector	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F
Data Processing	Xia2/ DIALS	Xia2/ DIALS	Xia2/ DIALS
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.1, 86.1, 148.1	86.2, 86.2, 147.1	86.1, 86.1, 148.8
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	40372 (1980)*	39456 (2880)*	45613 (2207)*
Resolution (Å)	74.51-1.97 (1.97-2.00)*	56.31-1.98 (2.03-1.98)*	74.51-1.89 (1.93-1.89)*
R_{merge}	0.079 (2.105)*	0.080 (2.35)*	0.165 (2.841)*
I/σI	14.8 (1.3)*	17.0 (1.4)*	9.3 (1.1)*
Completeness (%)	100 (98.9)*	100.0 (100.0)*	100.0 (98.3)*
Multiplicity	8.5 (7.8)*	18.9 (17.9)*	18.7 (18.9)*
Wilson B value (Å^2)	33.8	40.7	35.0
Refinement	PHENIX	PHENIX	PHENIX
R_{work}/R_{free}	0.1745/ 0.1995	0.1830/0.2083	0.1879/0.2039
No. atoms			
-Enzyme (A/B)	2705	2711	2719
-Ligand	NOG (10)	NOG (10)	NOG (10)
-Substrate	123	90	98
-Water	218	116	153
Avg. B-factors			
Enzyme	45.2	57.4	51.6
Ligand	32.1	41.3	35.7
Water	49.0	53.6	49.9
Substrate	69.2	84.7	97.0
R.m.s deviations			
-Bond lengths (Å)	0.011	0.006	0.005
-Bond angles (°)	0.977	0.809	0.761

* Highest resolution shell in parentheses.

†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

6.2.5.2 Evaluation of the "VNVN" motif in peptide substrates

Consensus ankyrin peptides 1CA_VNVN, 1CA_ANVN, and 1CA_VNAN (see above) were subjected to FIH hydroxylation assays. The resulting MALDI based end-point hydroxylation assays imply that all three peptides are hydroxylated by FIH only at one site (Figure 18). Using MALDI-MS/MS the hydroxylation site could be assigned to the canonical FIH hydroxylation site within the consensus ankyrin sequence (Figure 19).

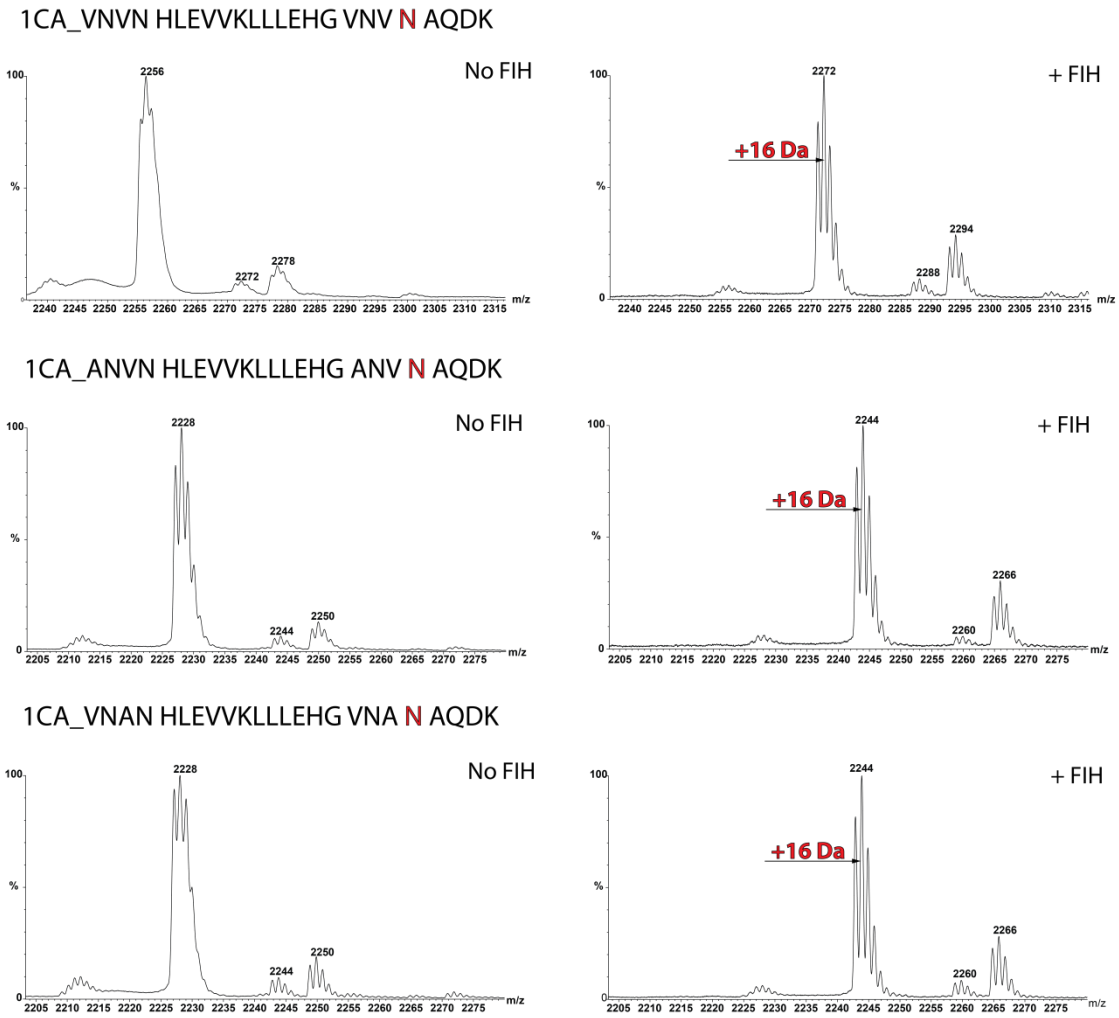
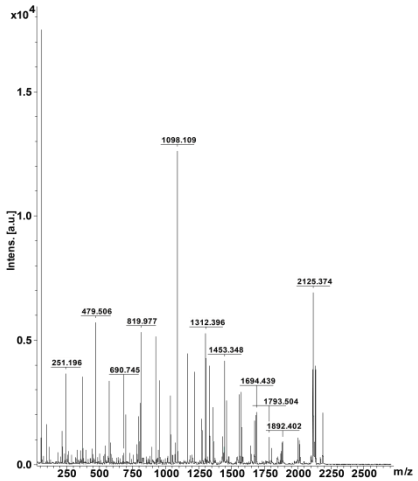


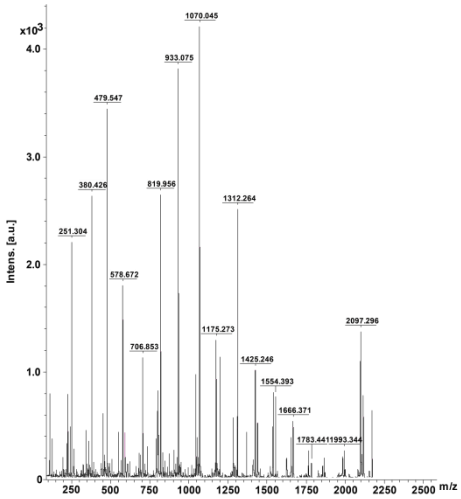
Figure 18 FIH hydroxylation assays of 1CA derived peptides containing "VNVN", "ANVN" and "VNAN" motifs. A End point hydroxylation assays of 1CA_VNVN, 1CA_ANVN and 1CA_VNAN. Reaction conditions: 50 µM peptide, 50 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), Tris-HCl pH = 7.5 incubated for 30 min at 37 °C.

1CA_VNVN HLEVVKLLLEHG VNV **N** AQDK



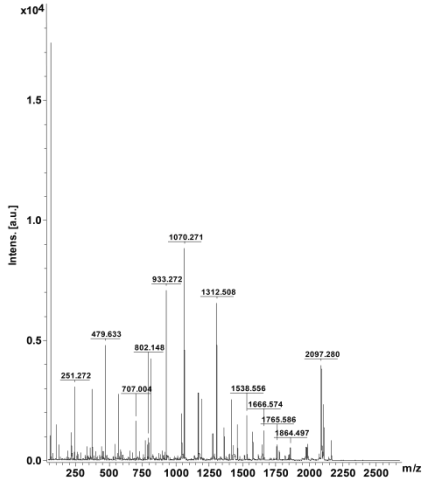
#	b	a	Seq	y	#
1			H		20
2	251.20		L	2134.46	19
3	380.35		E	2021.45	18
4	479.51		V	1892.40	17
5	578.65		V	1793.5	16
6	706.88		K	1694.44	15
7	819.98		L	1566.42	14
8	933.14		L	1453.35	13
9	1046.26		L		12
10	1175.30		E	1227.20	11
11	1312.40	1284.34	H	1098.11	10
12	1369.42	1340.28	G	961.00	9
13	1468.45	1440.45	V		8
14	1582.47		N	804.93	7
15	1681.45	1653.52	V		6
16	1811.47		N (OH)		5
17	1882.40		A		4
18	2010.44	1854.58	Q		3
19	2125.37		D		2
20			K		1

1CA_ANVN HLEVVKLLLEHG ANV **N** AQDK



#	b	a	Seq	y	#
1			H		20
2	251.3		L	2106.35	19
3	380.43		E	1993.34	18
4	479.55	451.54	V	1864.35	17
5	578.67	550.69	V	1765.38	16
6	706.85		K	1666.37	15
7	819.96	791.91	L	1538.32	14
8	933.07		L	1215.25	13
9	1046.19		L	1312.26	12
10	1175.27		E	1199.12	11
11		1284.34	H	1070.05	10
12			G		9
13	1440.32	1412.37	A		8
14	1554.39		N	1369.33	7
15	1653.38		V		6
16	1783.44		N (OH)	591.64	5
17	1854.43		A		4
18	1982.41		Q		3
19	2097.3		D		2
20			K		1

1CA_VNAN HLEVVKLLLEHG VNA **N** AQDK



#	b	a	Seq	y	#
1			H		20
2	251.21		L	2106.34	19
3	380.46		E	1993.42	18
4	479.63		V	1864.50	17
5	578.81		V		16
6	707.00		K	1765.59	15
7	820.18		L	1538.56	14
8	933.27		L	1425.52	13
9	1046.45		L		12
10	1175.51		E	1199.36	11
11	1312.51	1284.63	H	1070.27	10
12	1369.60		G		9
13	1468.62	1440.71	V		8
14	1582.63		N		7
15		1625.65	A		6
16	1783.51		N (OH)		5
17	1854.57		A		4
18	1982.52		Q		3
19	2097.28		D		2
20			K		1

Figure 19 FIH hydroxylation assays of 1CA derived peptides containing "VNVN", "ANVN" and "VNAN" motifs. MALDI-TOF MS/MS analysis of end point hydroxylation assays of 1CA_VNVN, 1CA_ANVN and 1CA_VNAN. Reaction conditions: 50 µM peptide, 50 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), Tris-HCl pH = 7.5 incubated for 30 min at 37 °C.

The results of MALDI-MS based time course assays imply a preference of FIH for the 1CA_ANVN and 1CA_VNVN over 1CA_VNAN (Figure 20). This is consistent with previously observed substrate specificity of FIH (see chapter 1). The consensus FIH substrate sequence is “ADVN”, N-terminal to the hydroxylation site (chapter 1). 1CA_ANVN is the closest related peptide to this consensus, so is predicted to be the best substrate. 1CA_VNAN lacks the valine in the -1 position to the hydroxylation site, which makes it significantly less preferable as an FIH substrate.

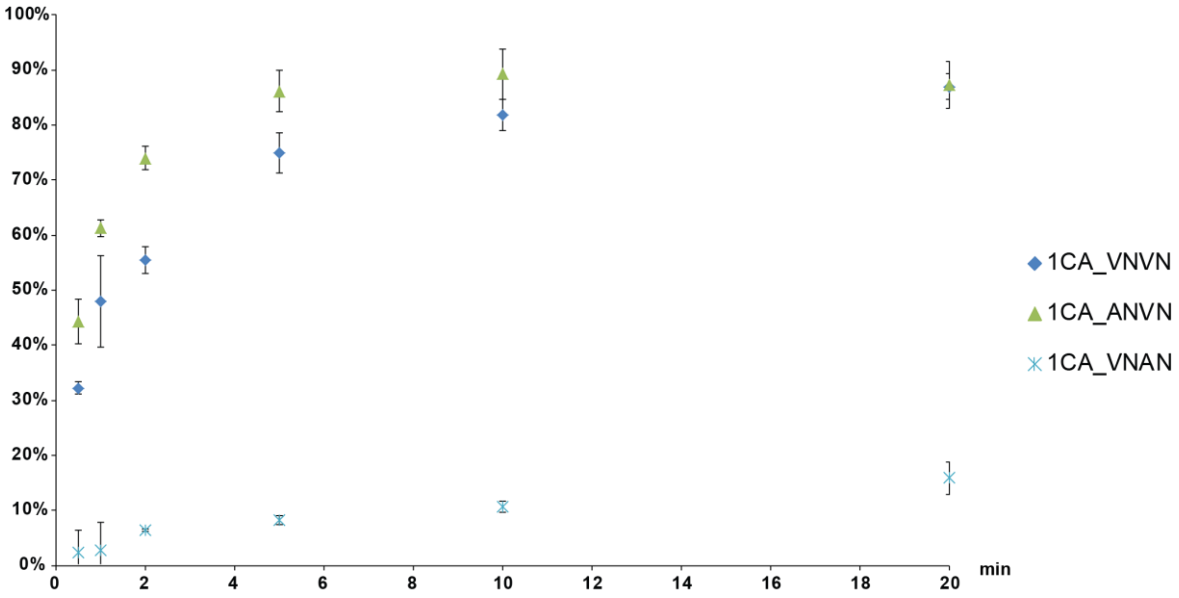


Figure 20 MALDI-MS based time course assay for FIH mediated hydroxylation of the 1CA derived peptides. Experimental conditions: 50 µM peptide, 1 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), Tris-HCl pH = 7.5 incubated at 37 °C (n=3, ± standard deviation).

The in-solution substrate specificity of FIH for the consensus ankyrin constructs was supported by X-ray crystallographic analysis. The 1CA_VNVN, 1CA_ANVN and 1CA_VNAN peptides were co-crystallised with FIH and the structures solved (in collaboration with Dr Rasheduzzaman Chowdhury, University of Oxford). The crystal structures revealed an overall conserved binding mode between the synthetic substrates (Figure 21). Notably, the B-factors for the 1CA_VNAN peptide are higher (~30%) than the B-factors for 1CA_VNVN and 1CA_ANVN peptides (Table 2), which could suggest lower occupancy, *i.e.* less binding in the crystal state.

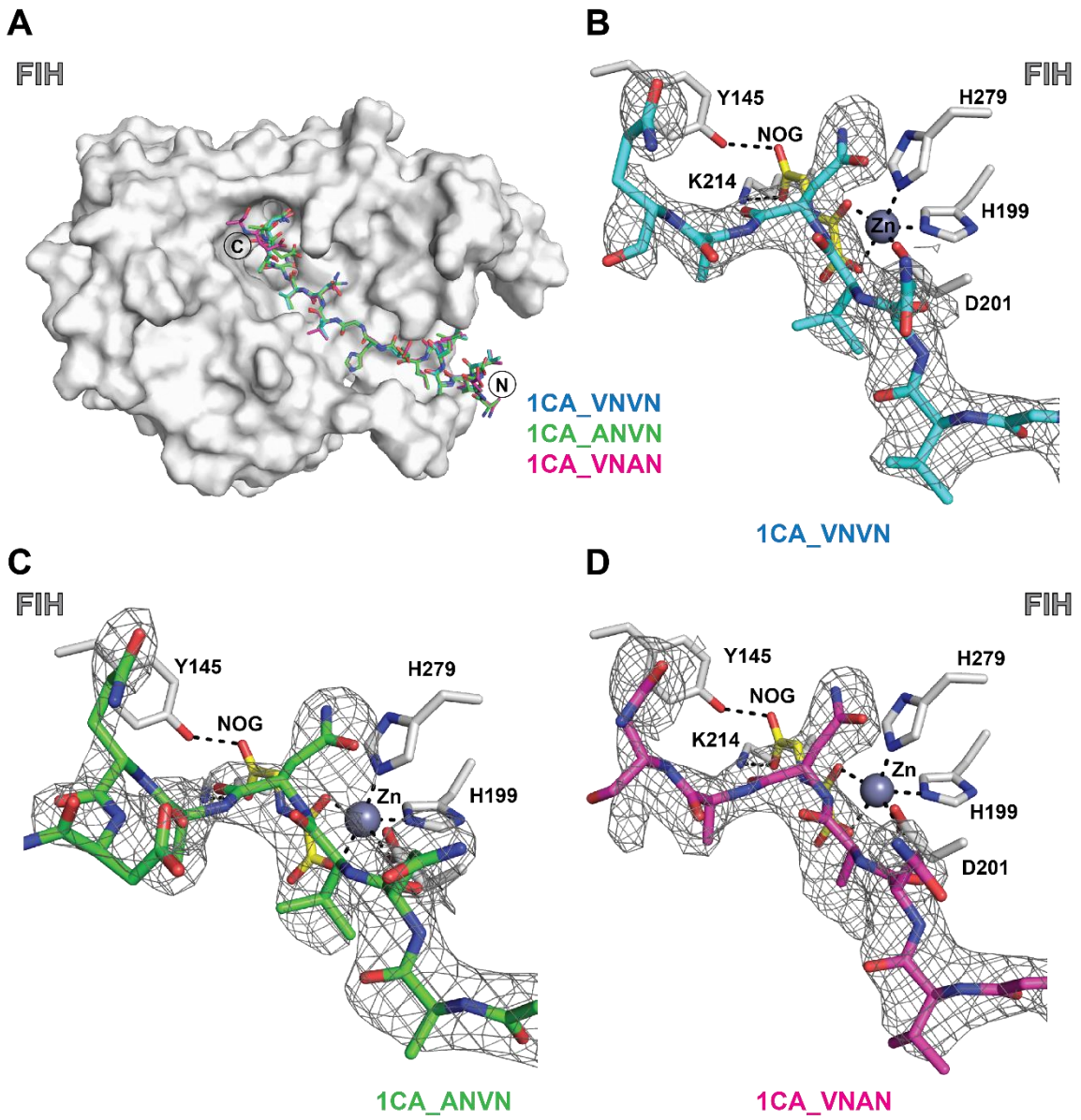


Figure 21 Views from crystal structures of FIH in complex with modified consensus ankyrin peptides. **A** View from the overlay of crystal structures of FIH in complex with 1CA_VNVN, 1CA_ANVN and 1CA_VNAN illustrates conserved binding. **B** Views from a crystal structure of FIH in complex with 1CA_VNVN, mFo-Fc OMIT maps contoured to 3σ around the substrate. **C** Views from a crystal structure of FIH in complex with 1CA_ANVN, mFo-Fc OMIT maps contoured to 3σ around the substrate. **D** Views from a crystal structure of FIH in complex with 1CA_VNAN, mFo-Fc OMIT maps contoured to 3σ around the substrate.

In summary, the results imply that the "VNVN" motif is not sufficient for double hydroxylation of short FIH peptide substrates. Neither ASPP derived peptides nor consensus ankyrin peptides containing a "VNVN" motif showed evidence for hydroxylation of 2 asparagines. The substrate specificity of consensus ankyrin derived peptides is of the order ANVN>VNVN>VNAN, which is in accordance to the previously reported substrate specify of FIH[62].

Table 2: Data collection and refinement statistics of FIH complexes with 1CA peptides substrates.

Datasets	**FIH**.Zn(II).NOG 1CA_ANVN	**FIH**.Zn(II).NOG 1CA_VNAN	**FIH**.Zn(II).NOG 1CA_VNVN
Data Collection			
Beamline (Wavelength, Å)	DLS I03 (0.9763)	DLS I03 (0.9763)	DLS I03 (0.9763)
Detector	Pilatus 6M-F	Pilatus 6M-F	Pilatus 6M-F
Data Processing	HKL2000[63]	HKL2000[63]	HKL2000[63]
Space group	$P4_12_12$	$P4_12_12$	$P4_12_12$
Cell dimensions			
a,*b*,*c* (Å)	86.3, 86.3, 143.1	86.6, 86.9, 147.6	86.3, 86.3, 145.7
α,β,γ(°)	90, 90, 90	90, 90, 90	90, 90, 90
No. of molecules/ASU	1	1	1
No. reflections	12485 (1220)*	19993 (1973)*	20146 (1976)*
Resolution (Å)	37.58-2.83 (2.93-2.83)*	74.87-2.50 (2.59-2.50)*	50-2.50 (2.59-2.50)*
R_{merge}	0.144 (0.869)*	0.057 (0.771)*	0.095 (0.959)*
I/σI	8.8 (2.0)*	27.9 (2.3)*	22.4(3.1)*
Completeness (%)	91.8 (92.5)*	98.0 (98.0)*	100 (99.8)*
Multiplicity	5.3 (5.0)*	6.2 (6.0)*	12.4 (11.9)*
Wilson B value (Å^2)	36.5	63.7	61.9
Refinement	PHENIX	PHENIX	PHENIX
R_{work}/R_{free}	0.2313/0.2491	0.2312/0.2584	0.2144/0.2480
No. atoms			
-Enzyme (A/B)	2715	2721	2587
-Ligand	NOG (10)	NOG(10)	NOG (10)
-Substrate	135	135	135
-Water	58	43	38
Avg. B-factors			
Enzyme	77.2	66.7	63.6
Ligand	76.5	79.2	57.1
Water	65.6	93.9	58.5
Substrate	80.9	126.2	92.0
R.m.s deviations			
-Bond lengths (Å)	0.003	0.003	0.009
-Bond angles (°)	0.566	0.571	0.625

* Highest resolution shell in parentheses.

†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

6.3 FIH mediated hydroxylation of recombinant ASPP and other VNVN protein constructs

6.3.1 Selection of human ARD proteins for FIH hydroxylation assays

On a protein level, all three ASPP proteins are hydroxylated twice by FIH on two different asparagine residues within the same ankyrin repeat. Estimating the abundance of the double hydroxylation of ASPP2 and iASPP, based on spectral and peptide counts, suggested that double hydroxylation appears to be present to a higher degree in ASPP2 compared with iASPP. The "VNVN" motif of ASPP2 might play are crucial role in FIH mediated double hydroxylation (see above). Since the -1 valine is important for FIH mediated hydroxylation, the "VNVN" motif is the most probable protein sequence to be doubly hydroxylated by FIH. Consequently, it was of interest to investigate if double hydroxylation could be observed for other human proteins.

To identify human proteins containing a "VNVN" motif in their ARD a database compiled by Schmierer et al[52] (generated from annotated ankyrins from the SMART[64], PFAM[65] and Uniprot[30] databases annotating 1505 ankyrin repeats from 252 distinct ARD proteins) was searched. Notably, the annotation of ankyrin repeats is often inaccurate, as for example ASPP1 and iASPP are annotated with only 2 ankyrin repeats, whereas the crystal structure, at least for iASPP, shows 4 ankyrin repeats. The actual number of ankyrin repeats in the human proteome is therefore likely to be higher than annotated in current databases. Within the set of 1505 annotated ankyrin repeats only 6 repeats from 6 distinct proteins were found to contain a 'VNVN' motif. Next to ASPP1 and ASPP2, ASB11 (ankyrin repeat and SOCS box protein 11), CDNK2C (cyclin-dependent kinase 4 inhibitor C, P18-INK4C), myosin 16 and GABPB2 (GA-binding protein subunit beta-2) were identified (Figure 22). According to the Kyoto Encyclopaedia of Genes and Genomes (KEGG)[66-68] pathway analysis, no significant enrichment for any pathway was found for the 6 proteins. This would suggest that the proteins have independent functions in the cells.

The domain structure of the "VNVN" proteins was then investigated. ASPP1 and ASPP2 are closely related to each other, but none of the other human "VNVN" proteins share common domain architecture besides the ARD (according to InterPro protein sequence and classification[31], Figure 22A). This raises the question as to whether the "VNVN" motif is a regulatory element for a specific pathway or rather a comparably rare structural motif for a small set of Ankyrin repeats.

FIH can accept many ankyrin repeat domain proteins as substrates (see chapter 1). A "consensus" recognition sequence has been identified (LLxxGADV**N**A), these include two key residues: -1 valine and -8 leucine with respect to the hydroxylation site (see Chapter 1). Alignment of the sequences of the 6 ankyrin repeat proteins containing the "VNVN" motif reveals high sequence homology for ASPP1/2, ASB11, CDN2C and MYO16 to the "FIH consensus" sequence (Figure 22B). All of the identified "VNVN" proteins have the "VNVN" motif in the same position within a particular ankyrin repeat, but not necessarily on the same repeat, i.e. ASPP1 and ASPP2 have the "VNVN" motif on the 2nd Ankyrin, whereas ASB11 and P18-INK4C incorporate the "VNVN" motif on the 1st Ankyrin repeat. Notably, of the 6 human "VNVN" proteins, only GABPB2 does not contain a leucine at the -8 position relative to the putative hydroxylation site. Based on the primary structure, P18-INK4C, ASB11 and myosone-16 are likely to be FIH substrates.

To investigate if the identified "VNVN" proteins are indeed FIH substrates and if the "VNVN" motif can be subject to double hydroxylation, a small screen of ankyrin protein hydroxylation based on the proteins analysed above, was envisaged. Crystal structures of ASB11 (PDB code 4UUC) and P18-INK4C have been reported (PDB code 1HIB[69]) and the availability of bacterial expression plasmids made them ideal candidates for the *in vitro* screen. Additionally, the role of the C-terminal SH3 domain of ASPP2 was investigated. To accomplish this, a version of the ASPP2 C-terminus, missing the C-terminal SH3 domain, was included in the screen (Figure 22D).

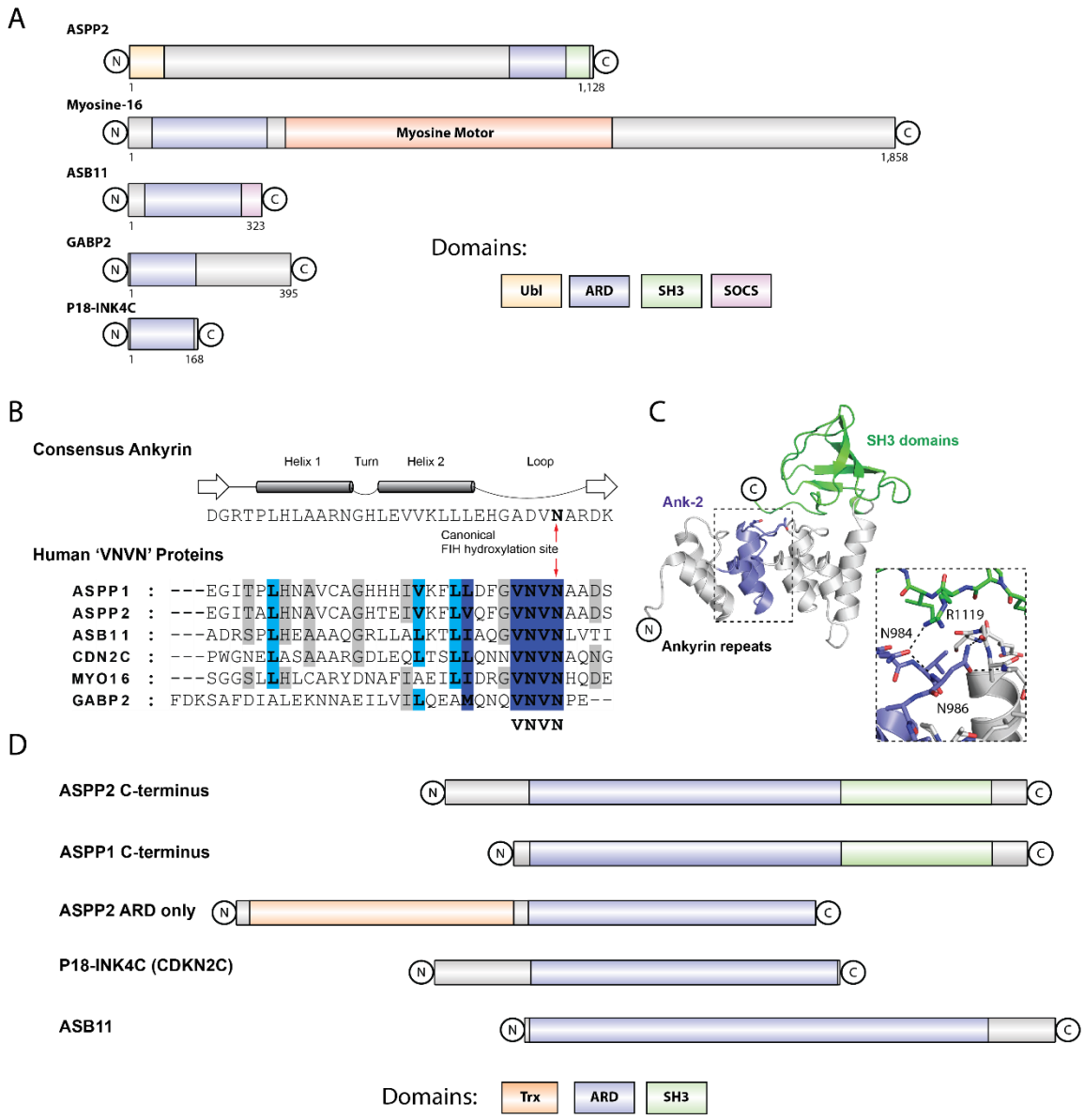


Figure 22 Sequence alignments of human ankyrin repeats bearing a "VNVN" motif. **A** Domain architectures of human "VNVN" proteins (domains according to InterPro[31] protein sequence analysis and classification). **B** Sequence alignment of human "VNVN" ankyrin repeats illustrates the position of the putative hydroxylation sites within the ankyrin sequence. **C** View from a crystal structure of the C-terminus of ASPP2 (2.27 Å, PDB code 4A63[29]) illustrates inter-domain and intra-domain interactions of the putative hydroxylate residues within the ASPP2 protein. **D** Domain architecture of protein constructs used for protein hydroxylation screen.

6.3.2 FIH hydroxylation assays of human VNVN proteins

As previously described (see section 6.2), $ASPP2_{920\text{-}1128}$ was readily hydroxylated in a FIH dependent manner at asparagine-986. Accurate quantification of modified peptides compared to non-modified peptides was (at least with current experimental set-up) not possible but indications of relative abundance could be estimated from the number of identified peptide spectra (see also section 6.2.3). Using spectral counts, 16 out of 25 peptide spectra covered both presumed hydroxylation sites were found to show evidence for asparagine-986 hydroxylation however only 2 showed evidence for asparagine-984 hydroxylation. This number is lower than what was previously observed and may be due to a lower enzyme concentration compared to previous experiments (section 6.2.3). None of the peptides covering asparagine-984 and asparagine-986 showed any evidence of hydroxylation in the no-FIH control (a total of 17 peptides), supporting FIH dependency of the modification. Peptides covering only one of the hydroxylation sites have been disregarded for this analysis.

The SH3 domains of ASPP proteins have been shown to be important for structural integrity of the C-terminal fold[33]. Structural stability, however, is assumed to be a hindrance to FIH mediated ARD hydroxylation[53,70]. Therefore, it was of interest to investigate how the truncation of the SH3 domains could change the hydroxylation pattern of ASPP2. Analysis of the FIH mediated hydroxylation of $ASPP2_{920\text{-}1057}$-Trx however did not yield conclusive results regarding the importance of the SH3 domain.

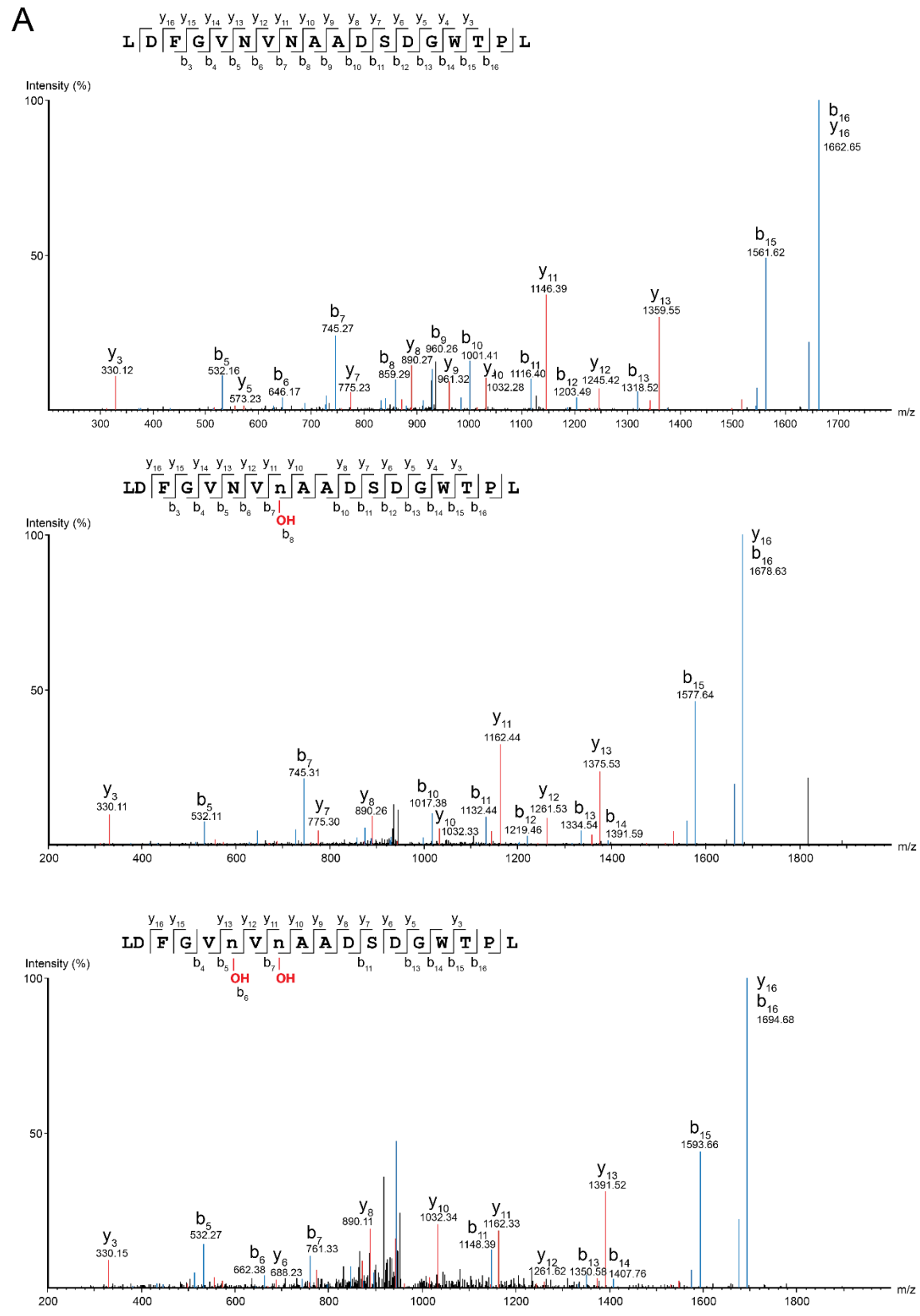
The C-terminus of $ASPP1_{883\text{-}1090}$ was found to be hydroxylated in an FIH dependent manner (Figure 23) this is consistent with the cell-based hydroxylation assays (see section 6.2.4). Comparing spectral counts of peptides covering both hydroxylation sites, 7/11 peptides showed evidence for hydroxylation at asparagine-948 (corresponding to asparagine-986 in ASPP2) and 2/11 peptides showed evidence for hydroxylation at asparagine-948 and asparagine-946.

In addition to the ASPP proteins, evidence for FIH mediated hydroxylation was observed for the full-length recombinant form of P18-INK4C at asparagine-32 (equivalent to asparagine-986 in ASPP2) and to a much lower level at asparagine-30 (equivalent to asparagine-984 in ASPP2, Figure 24). There was no evidence for hydroxylation in the no-FIH control, supporting the FIH dependency of the modification.

Evidence for FIH mediated hydroxylation was further observed for $ASB11_{64\text{-}287}$. Hydroxylations at asparagine-92 and asparagine-90 (corresponding to ASPP2 asparagine-986 and asparagine-984) were observed (Figure 26) in an FIH dependent manner. Notably, the levels of doubly hydroxylated peptides appear to be higher

compared to the singly hydroxylated peptides as judged by spectral count analysis. The results imply that in addition to the “VNVN” motif in the 1st ankyrin repeat, ASB11 can also be hydroxylated by FIH in the 2nd ankyrin repeat at asparagine-125 (Figure 26). It is noteworthy that there is evidence for an oxidation on the histidine at the -2 position relative to the hydroxylation site within the “AHVN” sequence. Histidine oxidation however, is a common artefact for MS/MS experiments and was also observed in the no-FIH control.

In summary, all proteins tested containing a “VNVN” motif, $ASPP2_{920\text{-}1128}$, $ASPP1_{883\text{-}1090}$, P18-INK4C and $ASB11_{64\text{-}287}$ were accepted by FIH as substrate, at least in their recombinant form. All of the proteins show some evidence for an FIH mediated double hydroxylation of the “VNVN” motif; however, the levels of double hydroxylation appear to vary. Quantification of the double hydroxylation was not accurate and further experiments are needed for accurate quantification such as intact protein mass-spectrometry.



(Figure continues on next page)

B

L D F G V N V N A A D S D G W T P L

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					18
2	229.12	211.11	212.09	115.06	D	1777.80	1759.79	1760.78	889.62	17
3	376.22	358.30	359.16	188.59	F	1662.65	1644.67	1645.75	832.19	16
4	433.07	415.25	416.18	217.10	G	1515.73	1497.62	1498.59	758.26	15
5	532.16	514.13	515.25	266.64	V	1458.49	1440.68	1441.66	729.84	14
6	646.17	628.15	629.26	323.66	N	1359.55	1341.57	1342.54	680.31	13
7	745.27	727.28	728.20	373.05	V	1245.42	1227.56	1228.08	623.29	12
8	859.29	841.42	842.04	430.22	N	1146.39	1128.50	1129.48	573.75	11
9	930.26	912.45	913.40	465.73	A	1032.28	1014.35	1015.53	517.20	10
10	1001.41	983.46	984.45	501.25	A	961.32	943.42	944.38	481.16	9
11	1116.40	1098.52	1099.45	558.39	D	890.27	872.42	873.36	446.07	8
12	1203.49	1185.46	1186.62	602.34	S	775.23	757.11	758.26	388.18	7
13	1318.52	1300.74	1301.56	659.80	D	688.25	670.32	671.32	344.67	6
14	1375.44	1357.60	1358.59	688.25	G	573.23	555.30	556.28	287.15	5
15	1561.62	1543.50	1544.55	781.35	W	516.34	498.27	499.31	258.64	4
16	1662.65	1644.67	1645.71	832.19	T	330.12	312.42	313.18	165.60	3
17	1759.79	1741.78	1742.77	880.52	P	229.15	211.14	212.13	115.08	2
18					L	132.10	114.09	115.07	66.55	1

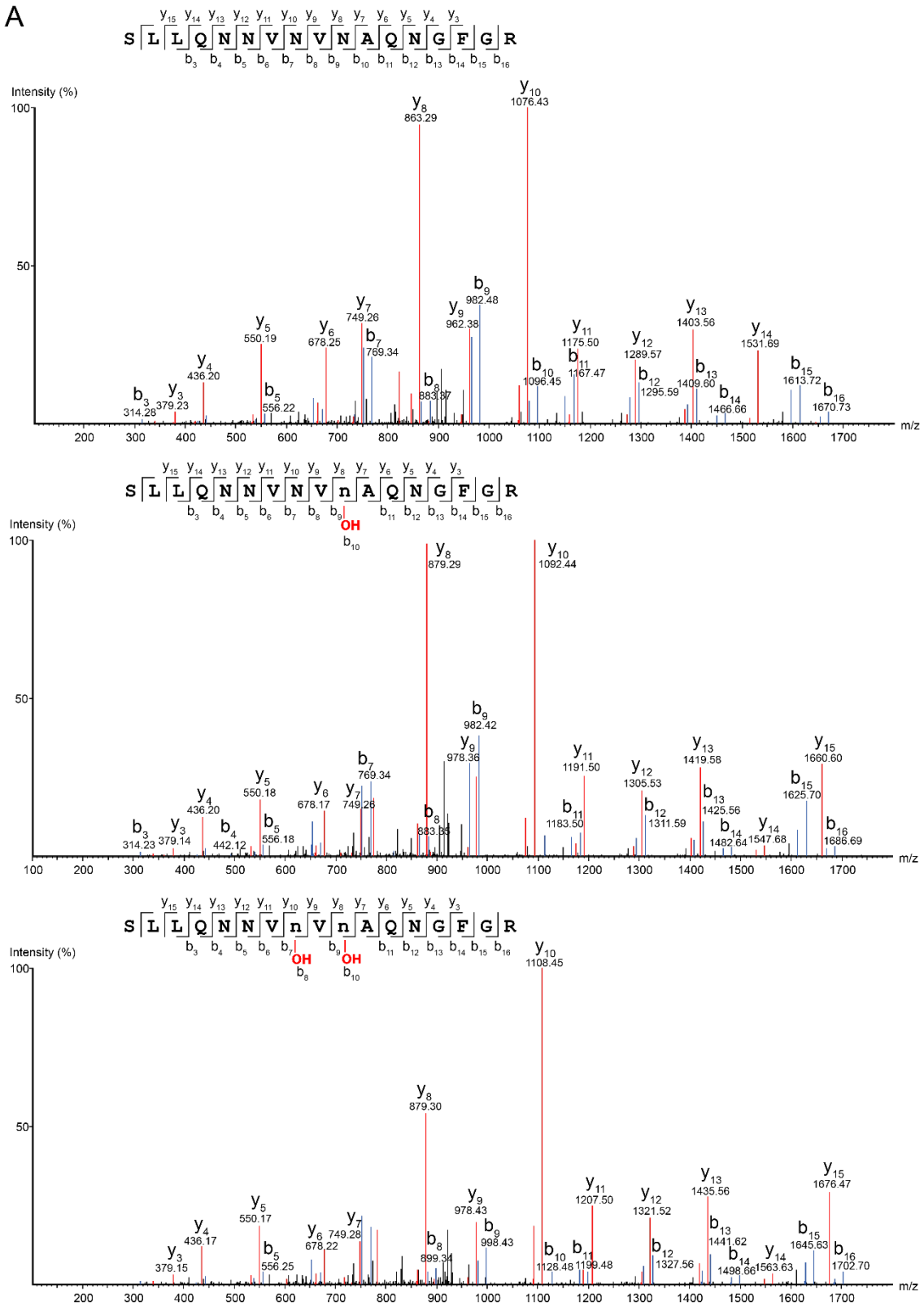
L D F G V N V n A A D S D G W T P L
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					18
2	229.12	211.11	212.09	115.06	D	1793.80	1775.79	1776.77	897.40	17
3	376.12	358.18	359.16	188.59	F	1678.63	1660.62	1661.74	839.89	16
4	433.23	415.20	416.09	217.10	G	1531.66	1513.61	1514.68	766.27	15
5	532.11	514.26	515.04	266.64	V	1474.51	1456.79	1457.65	737.84	14
6	646.25	628.14	629.17	323.66	N	1375.53	1357.55	1358.59	688.22	13
7	745.31	727.29	728.28	373.19	V	1261.53	1243.55	1244.54	631.20	12
8	875.40	857.22	858.90	438.21	N(+15.99)	1162.44	1144.39	1145.47	581.75	11
9	946.46	928.21	929.22	473.73	A	1032.33	1014.45	1015.44	516.73	10
10	1017.38	999.44	1000.47	509.01	A	961.43	943.27	944.40	481.06	9
11	1132.44	1114.06	1115.50	566.79	D	890.26	872.16	873.36	446.17	8
12	1219.46	1201.62	1202.37	610.74	S	775.30	757.39	758.34	388.18	7
13	1334.54	1316.61	1317.56	667.79	D	688.22	670.32	671.23	344.67	6
14	1391.59	1373.60	1374.58	696.30	G	573.21	555.40	556.28	287.15	5
15	1577.64	1559.72	1560.66	789.67	W	516.18	498.27	499.72	258.64	4
16	1678.63	1660.62	1661.71	839.87	T	330.11	312.12	313.18	165.60	3
17	1775.79	1757.78	1758.76	888.41	P	229.15	211.14	212.13	115.08	2
18					L	132.10	114.09	115.07	66.55	1

L D F G V n V n A A D S D G W T P L
OH OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					18
2	229.12	211.11	212.09	115.06	D	1809.79	1791.78	1792.77	905.40	17
3	376.19	358.18	359.16	188.59	F	1694.68	1676.68	1677.74	847.93	16
4	433.35	415.20	416.18	217.10	G	1547.74	1529.69	1530.81	774.28	15
5	532.27	514.18	515.25	266.64	V	1490.68	1472.67	1473.65	745.84	14
6	662.38	644.01	645.29	331.66	N(+15.99)	1391.52	1373.60	1374.58	696.30	13
7	761.33	743.75	744.54	381.08	V	1261.62	1243.56	1244.28	631.28	12
8	891.42	873.41	874.61	446.02	N(+15.99)	1162.33	1144.49	1145.10	581.75	11
9	962.46	944.34	945.43	481.73	A	1032.34	1014.80	1015.44	516.73	10
10	1033.50	1015.44	1016.25	517.33	A	961.48	943.37	944.34	481.21	9
11	1148.39	1130.51	1131.50	574.76	D	890.11	872.18	873.36	446.02	8
12	1235.55	1217.54	1218.53	618.28	S	775.27	757.35	758.45	388.18	7
13	1350.58	1332.87	1333.55	675.79	D	688.23	670.32	671.74	344.67	6
14	1407.76	1389.59	1390.58	704.30	G	573.22	555.18	556.28	287.15	5
15	1593.66	1575.49	1576.66	797.00	W	516.28	498.27	499.22	258.64	4
16	1694.68	1676.68	1677.70	847.93	T	330.15	312.19	313.18	165.60	3
17	1791.78	1773.77	1774.76	896.84	P	229.15	211.14	212.13	115.08	2
18					L	132.10	114.09	115.07	66.55	1

Figure 23 FIH mediated hydroxylation of recombinant ASPP1 N948 and N946. A MS-MS spectra of non-modified ASPP1 peptides, single hydroxylated peptides at asparagine-948 and double hydroxylated peptides at asparagine-946. **B** Ion tables from the spectra above. Reaction conditions 1 µM ASPP1, 0.25 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), Tris-HCl pH = 7.5 incubated for 3 h at 37 °C.



(Figure continues on next page)

B

S L L Q N N V N V N A Q N G F G R

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	88.04	70.03	71.01	44.52	S					17
2	201.12	183.11	184.10	101.06	L	1757.90	1739.89	1740.88	879.78	16
3	314.28	296.20	297.18	157.60	L	1644.82	1626.81	1628.00	822.77	15
4	442.20	424.37	425.24	221.63	Q	1531.69	1513.69	1514.64	766.37	14
5	556.22	538.18	539.22	278.65	N	1403.56	1385.76	1386.61	702.40	13
6	670.22	652.23	653.24	335.68	N	1289.57	1271.46	1272.52	645.32	12
7	769.34	751.34	752.32	385.21	V	1175.50	1157.70	1158.48	588.30	11
8	883.37	865.45	866.34	442.20	N	1076.43	1058.41	1059.46	539.22	10
9	982.48	964.52	965.39	491.77	V	962.38	944.42	945.39	481.67	9
10	1096.45	1078.56	1079.40	548.79	N	863.29	845.33	846.28	432.21	8
11	1167.47	1149.50	1150.51	584.31	A	749.26	731.27	732.22	375.66	7
12	1295.59	1277.62	1278.59	648.23	Q	678.25	660.21	661.21	340.09	6
13	1409.60	1391.62	1392.58	705.36	N	550.19	532.28	533.15	275.64	5
14	1466.66	1448.71	1449.59	734.27	G	436.20	418.24	419.29	218.62	4
15	1613.72	1595.67	1596.68	807.40	F	379.23	360.81	362.11	190.10	3
16	1670.73	1652.70	1653.69	835.91	G	232.14	214.13	215.11	116.57	2
17					R	175.12	157.11	158.09	88.06	1

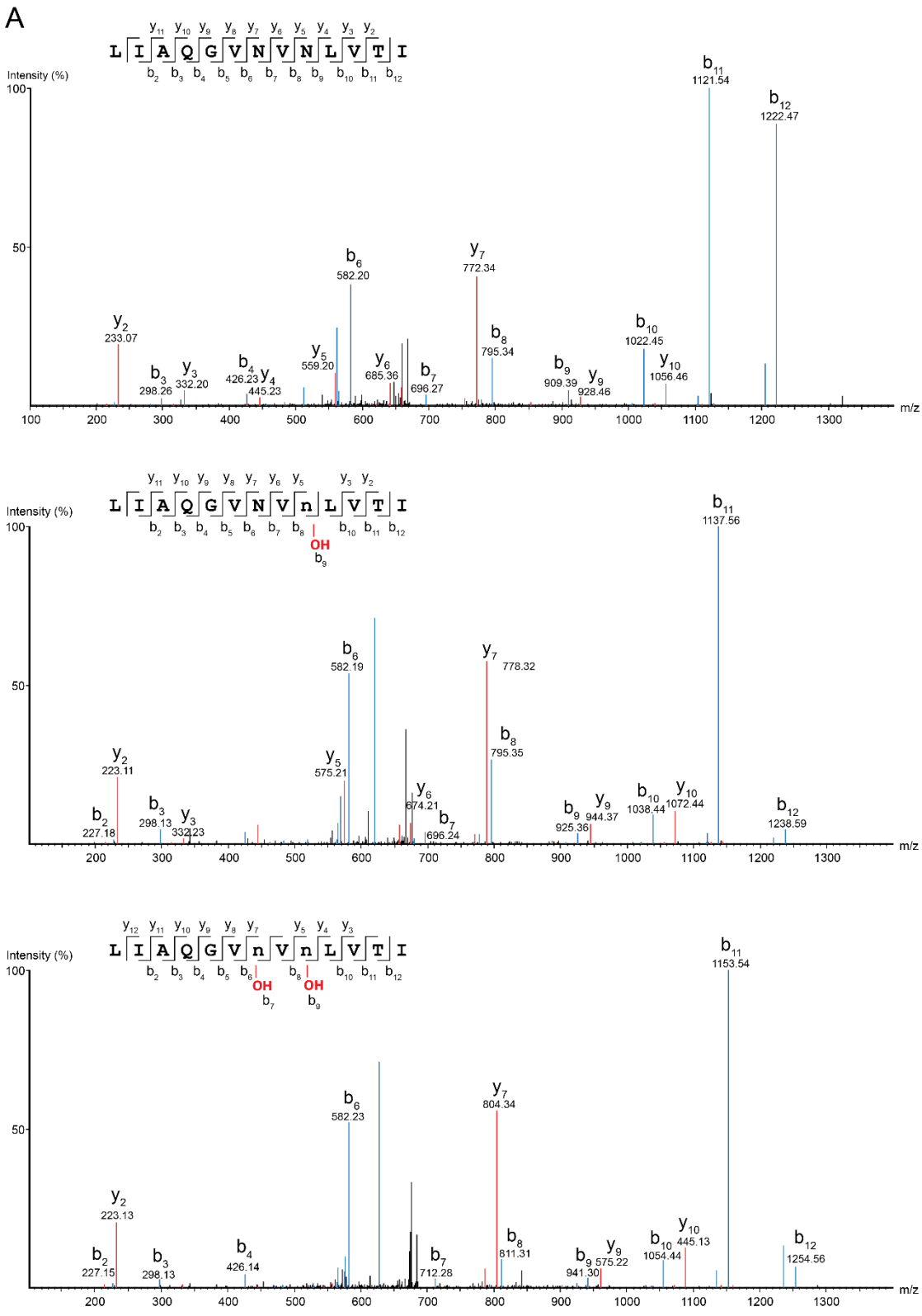
S L L Q N N V N V n A Q N G F G R
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	88.04	70.03	71.01	44.52	S					17
2	201.12	183.11	184.10	101.06	L	1773.90	1755.89	1756.87	887.42	16
3	314.23	296.23	297.18	157.60	L	1660.60	1642.80	1643.70	830.91	15
4	442.12	424.22	425.35	221.63	Q	1547.68	1529.65	1530.75	774.47	14
5	556.18	538.30	539.29	278.65	N	1419.58	1401.45	1402.52	710.38	13
6	670.20	652.19	653.19	335.68	N	1305.53	1287.62	1288.61	653.19	12
7	769.34	751.23	752.25	385.21	V	1191.50	1173.58	1174.48	596.36	11
8	883.35	865.35	866.28	442.12	N	1092.44	1074.56	1075.40	546.76	10
9	982.42	964.38	965.39	492.24	V	978.36	960.41	961.32	489.74	9
10	1112.53	1094.56	1095.54	556.78	N(+15.99)	879.29	861.41	862.32	440.20	8
11	1183.50	1165.53	1166.58	592.30	A	749.26	731.18	732.22	375.12	7
12	1311.59	1293.57	1294.55	656.36	Q	678.17	660.15	661.25	340.11	6
13	1425.56	1407.58	1408.68	713.35	N	550.18	532.16	533.12	275.64	5
14	1482.64	1464.70	1465.69	741.87	G	436.20	418.22	419.23	218.62	4
15	1629.70	1611.82	1612.70	815.48	F	379.14	361.20	361.72	190.10	3
16	1686.69	1668.62	1669.67	844.29	G	232.14	214.13	215.11	116.57	2
17					R	175.12	157.11	158.09	88.06	1

S L L Q N N V n V n A Q N G F G R
OH OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	88.04	70.03	71.01	44.52	S					17
2	201.12	183.11	184.10	101.06	L	1789.89	1771.88	1772.87	895.59	16
3	314.22	296.16	297.18	157.60	L	1676.47	1658.67	1659.78	838.90	15
4	442.16	424.15	425.11	221.63	Q	1563.63	1545.61	1546.58	782.50	14
5	556.25	538.19	539.23	278.65	N	1435.56	1417.63	1418.65	718.36	13
6	670.29	652.26	653.21	335.68	N	1321.52	1303.59	1304.54	661.20	12
7	769.34	751.30	752.28	385.21	V	1207.50	1189.67	1190.44	604.28	11
8	899.34	881.45	882.33	450.36	N(+15.99)	1108.45	1090.36	1091.38	555.19	10
9	998.43	980.52	981.36	499.76	V	978.43	960.57	961.49	489.25	9
10	1128.48	1110.55	1111.54	564.78	N(+15.99)	879.30	861.32	862.30	440.20	8
11	1199.48	1181.49	1182.58	600.30	A	749.28	731.27	732.32	375.18	7
12	1327.56	1309.64	1310.63	664.33	Q	678.22	660.24	661.20	340.13	6
13	1441.62	1423.56	1424.53	721.03	N	550.17	532.08	533.07	275.64	5
14	1498.66	1480.66	1481.67	749.86	G	436.17	418.27	419.16	218.62	4
15	1645.63	1627.67	1628.69	823.40	F	379.15	361.20	362.44	190.10	3
16	1702.70	1684.77	1685.76	851.91	G	232.14	214.13	215.11	116.57	2
17					R	175.12	157.11	158.09	88.06	1

Figure 24 FIH mediated hydroxylation of recombinant P18-INK4 N32/30. A MS-MS spectra for non-modified P18-INK4C peptides, single hydroxylated peptides at asparagine-32 and double hydroxylated peptides at asparagine-30. **B** Ion tables from the spectra above. Reaction conditions: 1 µM P18-INK4, 0.25 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), Tris-HCl pH = 7.5 incubated for 3 h at 37 °C.



(Figure continues on next page)

B

L I A Q G V N V N L V T I

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					13
2	227.16	209.17	210.15	114.09	I	1240.73	1222.47	1223.70	621.00	12
3	298.26	280.20	280.99	149.61	A	1127.50	1109.48	1110.73	564.22	11
4	426.23	408.26	409.32	213.64	Q	1056.46	1038.59	1039.37	529.09	10
5	483.22	465.28	466.14	242.15	G	928.46	910.54	911.52	464.77	9
6	582.20	564.22	565.09	291.68	V	871.30	853.12	854.50	436.26	8
7	696.27	678.39	679.38	348.70	N	772.34	754.26	755.43	386.73	7
8	795.34	777.16	778.19	397.98	V	658.36	640.48	641.24	329.71	6
9	909.39	891.29	892.49	455.11	N	559.20	541.33	542.41	280.17	5
10	1022.45	1004.66	1005.47	511.89	L	445.23	427.29	428.07	223.15	4
11	1121.54	1103.48	1104.52	561.39	V	332.20	314.29	315.38	166.61	3
12	1222.47	1204.73	1205.69	611.86	T	233.07	215.04	216.08	117.07	2
13					I	132.10	114.09	115.07	66.55	1

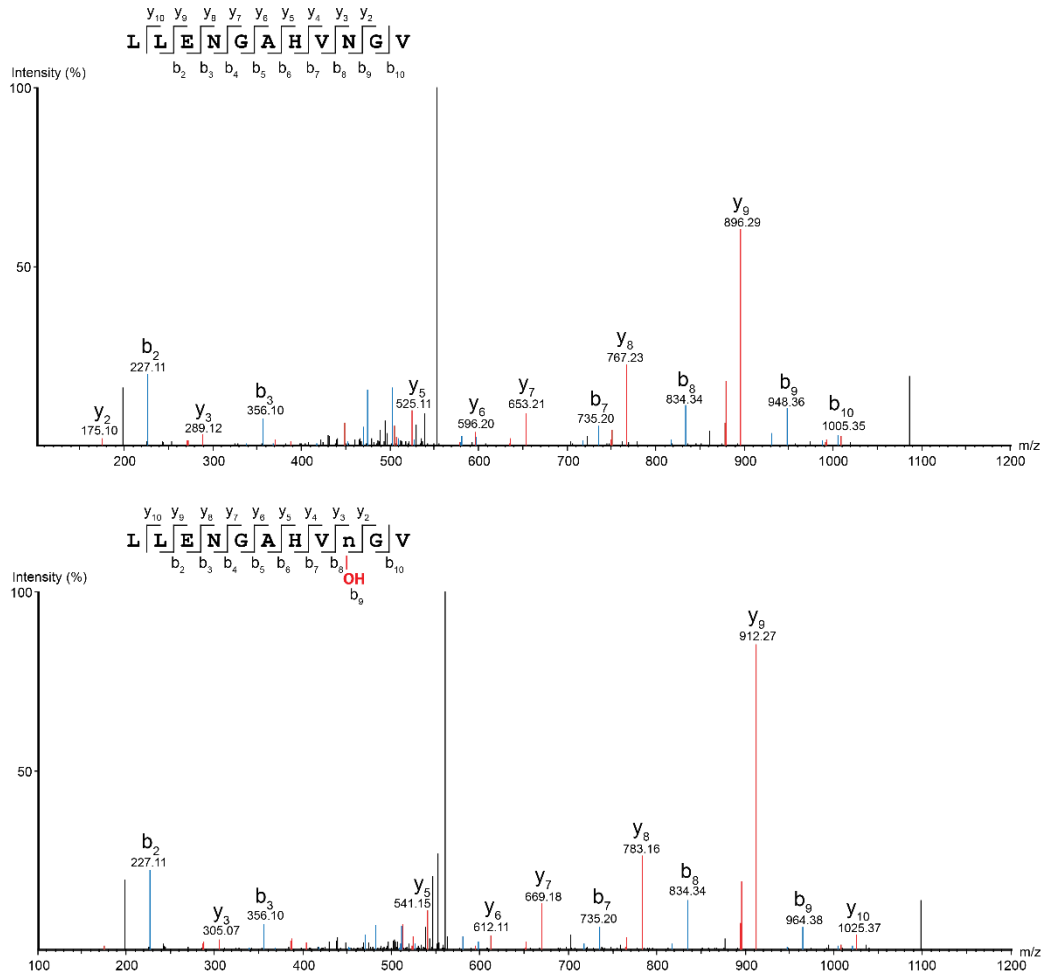
L I A Q G V N V n L V T I
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					13
2	227.18	209.17	210.15	114.09	I	1256.72	1238.59	1239.69	629.03	12
3	298.13	280.20	281.19	149.61	A	1143.52	1125.55	1126.61	572.32	11
4	426.11	408.26	409.22	213.64	Q	1072.44	1054.29	1055.48	536.80	10
5	483.24	465.25	466.25	242.15	G	944.37	926.53	927.51	472.77	9
6	582.19	564.20	565.22	291.68	V	887.39	869.66	870.49	444.12	8
7	696.24	678.04	679.03	348.95	N	788.32	770.43	771.42	394.73	7
8	795.35	777.46	778.29	398.11	V	674.21	655.95	657.23	338.02	6
9	925.36	907.16	908.48	463.10	N(+15.99)	575.21	557.33	558.31	288.17	5
10	1038.44	1020.78	1021.47	519.87	L	445.30	427.29	428.28	223.15	4
11	1137.56	1119.39	1120.60	569.38	V	332.23	314.29	315.08	166.61	3
12	1238.59	1220.67	1221.68	619.94	T	233.11	215.10	216.12	117.07	2
13					I	132.10	114.09	115.07	66.55	1

L I A Q G V n V n L V T I
OH OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					13
2	227.15	209.36	210.15	114.09	I	1272.72	1254.56	1255.69	636.86	12
3	298.13	280.05	281.19	149.61	A	1159.48	1141.44	1142.60	580.32	11
4	426.14	408.35	409.36	214.13	Q	1088.47	1070.58	1071.47	544.64	10
5	483.13	464.90	466.28	242.11	G	960.40	942.53	943.51	481.20	9
6	582.23	564.15	565.20	291.68	V	903.43	885.58	886.49	452.27	8
7	712.28	694.39	695.37	356.70	N(+15.99)	804.34	786.23	787.42	403.10	7
8	811.31	793.35	794.44	406.60	V	674.41	656.40	657.07	338.14	6
9	941.30	923.50	924.43	470.92	N(+15.99)	575.22	557.20	558.31	288.17	5
10	1054.44	1036.42	1037.56	528.22	L	445.13	427.29	428.28	223.15	4
11	1153.54	1135.57	1136.63	577.31	V	332.14	314.21	315.11	166.61	3
12	1254.56	1236.44	1237.68	627.98	T	233.13	215.10	216.12	117.07	2
13					I	132.10	114.09	115.07	66.55	1

Figure 25 FIH mediated Hydroxylation of recombinant ASB11 at N92/N90. A MS-MS spectra for non-modified ASB11 peptide, single hydroxylated peptide at asparagine-92 and double hydroxylated peptide at asparagine-92 and asparagine-90. **B** Ion tables from the spectra above. Reaction conditions: 1 µM ASB11, 0.25 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), Tris-HCl pH = 7.5 incubated for 3 h at 37 °C.



L L E N G A H V N G V

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					11
2	227.11	209.18	210.15	114.09	L	1009.35	991.37	992.38	505.32	10
3	356.10	338.23	339.19	178.61	E	896.29	878.26	879.21	448.83	9
4	470.14	452.29	453.23	235.63	N	767.23	749.22	750.19	384.38	8
5	527.12	509.27	510.12	264.14	G	653.21	635.22	636.26	327.15	7
6	598.10	580.20	581.12	299.24	A	596.20	578.28	579.18	298.66	6
7	735.20	717.32	718.28	368.31	H	525.11	507.27	508.09	263.14	5
8	834.34	816.44	817.29	417.92	V	388.22	370.11	371.07	194.61	4
9	948.36	930.37	931.35	474.83	N	289.12	271.09	272.10	145.08	3
10	1005.35	987.29	988.47	503.41	G	175.10	157.10	158.08	88.05	2
11					V	118.09	100.08	101.06	59.54	1

L L E N G A H V n G V
OH

#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	114.09	96.08	97.06	57.55	L					11
2	227.08	209.14	210.07	114.09	L	1025.37	1007.34	1008.40	513.26	10
3	356.10	338.17	339.16	178.99	E	912.27	894.26	895.29	456.71	9
4	470.14	452.08	453.10	235.63	N	783.16	765.14	766.24	392.32	8
5	527.18	509.15	510.11	263.94	G	669.18	651.09	652.17	335.17	7
6	598.21	580.35	581.18	299.66	A	612.11	594.59	595.24	306.32	6
7	735.21	717.14	718.14	368.21	H	541.15	523.06	524.08	271.14	5
8	834.34	816.44	817.38	417.73	V	404.04	386.12	387.10	202.21	4
9	964.38	946.47	947.39	482.76	N(+15.99)	305.07	287.11	288.09	153.07	3
10	1021.43	1003.50	1004.38	511.23	G	175.09	157.10	158.08	88.05	2
11					V	118.09	100.08	101.06	59.54	1

Figure 26 FIH mediated hydroxylation of ASB11 at N125. Top: MS-MS spectrum of non-hydroxylated ASB11 peptide; **Middle:** MS-MS spectrum of ASB11 peptide hydroxylated at N125; **Bottom:** Ion-tables from the MS-MS spectra above. Reaction conditions: 1 µM ASB11, 0.25 µM FIH, 1 mM ascorbate, 1 mM 2OG, 200 µM Fe(II), Tris-HCl pH = 7.5 incubated for 3 h at 37 °C.

6.4 Conclusions and outlook

The work performed on members of the ASPP protein family implys that FIH can accept all three members of the ASPP protein family as substrates at the peptide and, the isolated recombinant protein levels as well as with full-length proteins in cells. Crystal structures of FIH in complex with ASPP1, ASPP2 and iASPP derived peptides were solved and ASPP proteins were shown to be strong competitors towards FIH mediated HIF-α hydroxylation. The exact mechanisms by which ASPP proteins can inhibit FIH mediated HIF-α hydroxylation remains to be elucidated.

Hydroxylation studies of FIH with full-length ankyrin protein constructs *in vitro* showed evidence for an unprecedented second hydroxylation site next to the canonical FIH hydroxylation site.

Studies on ASPP2 and iASPP revealed a “VNVN” motif favouring double hydroxylation by FIH. Biochemical and structural studies with consensus ankyrin peptide sequences bearing a “VNVN”, “ANVN” and “VNAN” motif show that the “VNVN” is not sufficient for FIH mediated double hydroxylation. FIH mediated double hydroxylation is only observed for intact ARD proteins and not for shorter peptides. This observation underlines the importance of long-range protein-substrate interactions in FIH-ARD complexes. Subsequent FIH hydroxylation studies with recombinant forms of ASB11 and P18-INK4C proteins reveal that FIH mediated double hydroxylation is not limited to the ASPP protein family. No obvious link regarding sequence homology, domain structure or biological function of human “VNVN” proteins was found; leaving the physiological role of FIH mediated double hydroxylation to be elucidated.

6.5 References

1 Finlay, C. A., Hinds, P. W. & Levine, A. J. The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57**, 1083-1093 (1989).

2 Baker, S. J. *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244**, 217-221 (1989).

3 Hinds, P., Finlay, C. & Levine, A. J. Mutation is required to activate the p53 gene for cooperation with the ras oncogene and transformation. *J Virol* **63**, 739-746 (1989).

4 Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307-310, doi:10.1038/35042675 (2000).

5 Slee, E. A., O'Connor, D. J. & Lu, X. To die or not to die: how does p53 decide? *Oncogene* **23**, 2809-2818, doi:10.1038/sj.onc.1207516 (2004).

6 Rouault, J. P. *et al.* Identification of BTG2, an antiproliferative p53-dependent component of the DNA damage cellular response pathway. *Nat Genet* **14**, 482-486, doi:10.1038/ng1296-482 (1996).

7 el-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817-825 (1993).

8 Hermeking, H. *et al.* 14-3-3sigma is a p53-regulated inhibitor of G2/M progression. *Mol Cell* **1**, 3-11 (1997).

9 Hollander, M. C. *et al.* Analysis of the mammalian gadd45 gene and its response to DNA damage. *J Biol Chem* **268**, 24385-24393 (1993).

10 Tanaka, H. *et al.* A ribonucleotide reductase gene involved in a p53-dependent cell-cycle checkpoint for DNA damage. *Nature* **404**, 42-49, doi:10.1038/35003506 (2000).

11 Hsieh, J. K. *et al.* RB regulates the stability and the apoptotic function of p53 via MDM2. *Mol Cell* **3**, 181-193 (1999).

12 Ryan, K. M. & Vousden, K. H. Characterization of structural p53 mutants which show selective defects in apoptosis but not cell cycle arrest. *Mol Cell Biol* **18**, 3692-3698 (1998).

13 Caelles, C., Helmberg, A. & Karin, M. p53-dependent apoptosis in the absence of transcriptional activation of p53-target genes. *Nature* **370**, 220-223, doi:10.1038/370220a0 (1994).

14 Ding, H. F. *et al.* Essential role for caspase-8 in transcription-independent apoptosis triggered by p53. *J Biol Chem* **275**, 38905-38911, doi:10.1074/jbc.M004714200 (2000).

15 Lane, D. P. Cancer. p53, guardian of the genome. *Nature* **358**, 15-16, doi:10.1038/358015a0 (1992).

16 Vousden, K. H. & Lu, X. Live or let die: the cell's response to p53. *Nat Rev Cancer* **2**, 594-604, doi:10.1038/nrc864 (2002).

17 Cadwell, C. & Zambetti, G. P. The effects of wild-type p53 tumor suppressor activity and mutant p53 gain-of-function on cell growth. *Gene* **277**, 15-30 (2001).

18 Sullivan, A. & Lu, X. ASPP: a new family of oncogenes and tumour suppressor genes. *Br J Cancer* **96**, 196-200, doi:10.1038/sj.bjc.6603525 (2007).

19 Trigiante, G. & Lu, X. ASPP [corrected] and cancer. *Nat Rev Cancer* **6**, 217-226, doi:10.1038/nrc1818 (2006).

20 Naumovski, L. & Cleary, M. L. The p53-binding protein 53BP2 also interacts with Bc12 and impedes cell cycle progression at G2/M. *Mol Cell Biol* **16**, 3884-3892 (1996).

21 Iwabuchi, K., Bartel, P. L., Li, B., Marraccino, R. & Fields, S. Two cellular proteins that bind to wild-type but not mutant p53. *Proc Natl Acad Sci U S A* **91**, 6098-6102 (1994).

22 Samuels-Lev, Y. *et al.* ASPP proteins specifically stimulate the apoptotic function of p53. *Mol Cell* **8**, 781-794 (2001).

23 Vives, V. *et al.* ASPP2 is a haploinsufficient tumor suppressor that cooperates with p53 to suppress tumor growth. *Genes Dev* **20**, 1262-1267, doi:10.1101/gad.374006 (2006).

24 Vives, V., Slee, E. A. & Lu, X. ASPP2: a gene that controls life and death in vivo. *Cell Cycle* **5**, 2187-2190 (2006).

25 Bergamaschi, D. *et al.* ASPP1 and ASPP2: common activators of p53 family members. *Mol Cell Biol* **24**, 1341-1350 (2004).

26 Kas, S. M. *et al.* Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. *Nat Genet* **49**, 1219-1230, doi:10.1038/ng.3905 (2017).

27 Bergamaschi, D. *et al.* iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. *Nat Genet* **33**, 162-167, doi:10.1038/ng1070 (2003).

28 Lettre, G. *et al.* Genome-wide RNAi identifies p53-dependent and -independent regulators of germ cell apoptosis in C. elegans. *Cell Death Differ* **11**, 1198-1203, doi:10.1038/sj.cdd.4401488 (2004).

29 Canning, P., von Delft, F. & Bullock, A. N. Structural basis for ASPP2 recognition by the tumor suppressor p73. *J Mol Biol* **423**, 515-527, doi:10.1016/j.jmb.2012.08.005 (2012).

30 The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).

31 Mitchell, A. *et al.* The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**, D213-221, doi:10.1093/nar/gku1243 (2015).

32 Gorina, S. & Pavletich, N. P. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science* **274**, 1001-1005 (1996).

33 Robinson, R. A., Lu, X., Jones, E. Y. & Siebold, C. Biochemical and structural studies of ASPP proteins reveal differential binding to p53, p63, and p73. *Structure* **16**, 259-268, doi:10.1016/j.str.2007.11.012 (2008).

34 Tidow, H., Andreeva, A., Rutherford, T. J. & Fersht, A. R. Solution structure of ASPP2 N-terminal domain (N-ASPP2) reveals a ubiquitin-like fold. *J Mol Biol* **371**, 948-958, doi:10.1016/j.jmb.2007.05.024 (2007).

35 Espanel, X. & Sudol, M. Yes-associated protein and p53-binding protein-2 interact through their WW and SH3 domains. *J Biol Chem* **276**, 14514-14523, doi:10.1074/jbc.M008568200 (2001).

36 Liu, C. Y. *et al.* PP1 cooperates with ASPP2 to dephosphorylate and activate TAZ. *J Biol Chem* **286**, 5558-5566, doi:10.1074/jbc.M110.194019 (2011).

37 Royer, C. *et al.* ASPP2 links the apical lateral polarity complex to the regulation of YAP activity in epithelial cells. *PLoS One* **9**, e111384, doi:10.1371/journal.pone.0111384 (2014).

38 Vigneron, A. M., Ludwig, R. L. & Vousden, K. H. Cytoplasmic ASPP1 inhibits apoptosis through the control of YAP. *Genes Dev* **24**, 2430-2439, doi:10.1101/gad.1954310 (2010).

39 Llanos, S. *et al.* Inhibitory member of the apoptosis-stimulating proteins of the p53 family (iASPP) interacts with protein phosphatase 1 via a noncanonical binding motif. *J Biol Chem* **286**, 43039-43044, doi:10.1074/jbc.M111.270751 (2011).

40 Sottocornola, R. *et al.* ASPP2 binds Par-3 and controls the polarity and proliferation of neural progenitors during CNS development. *Dev Cell* **19**, 126-137, doi:10.1016/j.devcel.2010.06.003 (2010).

41 Cong, W. *et al.* ASPP2 regulates epithelial cell polarity through the PAR complex. *Curr Biol* **20**, 1408-1414, doi:10.1016/j.cub.2010.06.024 (2010).

42 Helps, N. R., Barker, H. M., Elledge, S. J. & Cohen, P. T. Protein phosphatase 1 interacts with p53BP2, a protein which binds to the tumour suppressor p53. *FEBS Lett* **377**, 295-300, doi:10.1016/0014-5793(95)01347-4 (1995).

43 Nakagawa, H. *et al.* APCL, a central nervous system-specific homologue of adenomatous polyposis coli tumor suppressor, binds to p53-binding protein 2 and translocates it to the perinucleus. *Cancer Res* **60**, 101-105 (2000).

44 Chatr-Aryamontri, A. *et al.* The BioGRID interaction database: 2017 update. *Nucleic Acids Res* **45**, D369-D379, doi:10.1093/nar/gkw1102 (2017).

45 Stark, C. *et al.* BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **34**, D535-539, doi:10.1093/nar/gkj109 (2006).

46 Janke, K. *et al.* Factor inhibiting HIF-1 (FIH-1) modulates protein interactions of apoptosis-stimulating p53 binding protein 2 (ASPP2). *J Cell Sci* **126**, 2629-2640, doi:10.1242/jcs.117564 (2013).

47 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, doi:10.1074/jbc.M804999200 (2008).

48 Lancaster, D. E. *et al.* Disruption of dimerization and substrate phosphorylation inhibit factor inhibiting hypoxia-inducible factor (FIH) activity. *Biochem J* **383**, 429-437, doi:10.1042/BJ20040735 (2004).

49 Cockman, M. E. *et al.* Hypoxia inducible factor-alpha binding and ubiquitylation by the von Hippel-Lindau tumor suppressor protein. *J Biol Chem* **275**, 25733-25741, doi:10.1074/jbc.M002740200 (2000).

50 Cockman, M. E., Webb, J. D. & Ratcliffe, P. J. FIH-dependent asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Ann N Y Acad Sci* **1177**, 9-18, doi:10.1111/j.1749-6632.2009.05042.x (2009).

51 Coleman, M. L. *et al.* Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038, doi:10.1074/jbc.M704102200 (2007).

52 Schmierer, B., Novák, B. & Schofield, C. J. Hypoxia-dependent sequestration of an oxygen sensor by a widespread structural motif can shape the hypoxic response--a predictive kinetic model. *BMC Syst Biol* **4**, 139, doi:10.1186/1752-0509-4-139 (2010).

53 Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-Inducible Factor (FIH) as an Oxygen Sensor. *J Biol Chem*, doi:10.1074/jbc.M115.653014 (2015).

54 Yang, M. *et al.* Asparagine and aspartate hydroxylation of the cytoskeletal ankyrin family is catalyzed by factor-inhibiting hypoxia-inducible factor. *J Biol Chem* **286**, 7648-7660, doi:10.1074/jbc.M110.193540 (2011).

55 Yang, M. *et al.* Factor-inhibiting hypoxia-inducible factor (FIH) catalyses the post-translational hydroxylation of histidinyl residues within ankyrin repeat domains. *FEBS J* **278**, 1086-1097, doi:10.1111/j.1742-4658.2011.08022.x (2011).

56 Yang, M. *et al.* Substrate selectivity analyses of factor inhibiting hypoxia-inducible factor. *Angew Chem Int Ed Engl* **52**, 1700-1704, doi:10.1002/anie.201208046 (2013).

57 Wilkins, M. R. *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-552 (1999).

58 Keil, B. Proteolysis Data Bank: specificity of alpha-chymotrypsin from computation of protein cleavages. *Protein Seq Data Anal* **1**, 13-20 (1987).

59 Rietschel, B. *et al.* Elastase digests: new ammunition for shotgun membrane proteomics. *Mol Cell Proteomics* **8**, 1029-1043, doi:10.1074/mcp.M800223-MCP200 (2009).

60 Davis, S. *et al.* Expanding Proteome Coverage with CHarge Ordered Parallel Ion aNalysis (CHOPIN) Combined with Broad Specificity Proteolysis. *J Proteome Res* **16**, 1288-1299, doi:10.1021/acs.jproteome.6b00915 (2017).

61 Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806, doi:10.1074/jbc.C200644200 (2003).

62 Linke, S. *et al.* Substrate requirements of the oxygen-sensing asparaginyl hydroxylase factor-inhibiting hypoxia-inducible factor. *J Biol Chem* **279**, 14391-14397, doi:10.1074/jbc.M313614200 (2004).

63 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr D Biol Crystallogr* **66**, 133-144, doi:10.1107/S0907444909047374 (2010).

64 Letunic, I. *et al.* SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res* **34**, D257-260, doi:10.1093/nar/gkj079 (2006).

65 Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279-285, doi:10.1093/nar/gkv1344 (2016).
66 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
67 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**, D457-462, doi:10.1093/nar/gkv1070 (2016).
68 Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* **45**, D353-D361, doi:10.1093/nar/gkw1092 (2017).
69 Venkataramani, R., Swaminathan, K. & Marmorstein, R. Crystal structure of the CDK4/6 inhibitory protein p18INK4c provides insights into ankyrin-like repeat structure/function and tumor-derived p16INK4 mutations. *Nat Struct Biol* **5**, 74-81 (1998).
70 Hardy, A. P., Prokes, I., Kelly, L., Campbell, I. D. & Schofield, C. J. Asparaginyl beta-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J Mol Biol* **392**, 994-1006, doi:10.1016/j.jmb.2009.07.070 (2009).

7 Biological roles of ASPP2-FIH interactions

This Chapter describes studies on the interactomes of ASPP2 in U2OS and AGS cells. Several potential ASPP2 interaction partners were identified using V5-tagged ASPP2 and were verified by western blots. Two novel interaction partners were found that bound exclusively to a PP1 binding-deficient ASPP2 variant. A robust and reproducible methodology for the immunoprecipitation of endogenous ASPP2 was developed. The results of interaction studies of endogenous ASPP2 were compared with those for overexpressed ASPP2. The effects of FIH-mediated hydroxylation and the impact of alanine variations of the FIH hydroxylation sites on ASPP2 interactions were investigated.

7.1 Introduction

FIH is a promiscuous hydroxylase with a well-defined function in the HIF pathway (chapter 1.3). In addition to the HIF-α isoforms, FIH catalyses hydroxylation of many ARD containing proteins[1] (see also chapter 1.4). While the role of FIH in the HIF pathway is well established, the physiological role of ankyrin hydroxylation remains to be elucidated (see chapter 1.7). The studies presented in chapter 6 revealed evidence for an unprecedented double hydroxylation of a single ankyrin domain in ASPP proteins. Further, competition of ASPP proteins with HIF for FIH activity was observed *in vitro* (chapter 6). These results raise the possibility that FIH-mediated hydroxylation of ASPP proteins regulates ASPP function.

The HIF pathway is substantially regulated at the protein level, with the PHD catalysis negatively regulating HIF-1/2α stability and FIH negatively regulating HIF transcriptional activities[2,3] (see also chapter 1). To investigate the possible effects of FIH-mediated hydroxylation on ASPP2 activity and a possible role of ASPP2-FIH interactions on FIH activity in human cells, proteomic approaches were used to account for possible regulatory mechanisms at the protein level.

Firstly, the effects of CRISPR KO of FIH and ASPP2 on the proteome of U2OS cells was studied, followed by investigating the interactome of ASPP2 in the presence and absence of FIH, and compared to the interactome of an ASPP2 PP1 binding deficient variant.

7.2 Proteomic analysis of U2OS WT, U2OS FIH KO and U2OS ASPP2 KO cells

Initially, the effect of the CRISPR knock-out of ASPP2 and FIH on the proteome of U2OS cells was investigated using mass-spectrometry. U2OS WT, U2OS ASPP2 and U2OS FIH CRISPR KO cells were grown to near full confluency (>90%); the medium was then removed, and the cell monolayer was washed with PBS buffer. The cells were then harvested by scraping, and immediately frozen at -80 °C. In cooperation with Ms. Georgina Berridge (University of Oxford) the small molecule fraction was extracted for use in ongoing metabolomic analyses. The proteins were re-suspended in urea buffer prior to trypsin digestion for liquid chromatography- tandem mass spectrometry (LC-MS/MS) analysis (see chapter 10). Experiments were performed with three independent replicates. Data analysis was performed using PEAKS 7.5[4] for peptide annotation, Progenesis (Nonlinear Dynamics, Newcastle, UK), for label-free quantification and normalisation against housekeeping genes, and Perseus[5] for downstream analysis and visualisation. In total, >2,600 proteins were identified. Proteins only identified by a single peptide (~400) were excluded from further quantitative analysis. Therefore, in total, ~2,200 proteins were analysed. In principal component analysis (PCA), the U2OS FIH KO samples form a cluster distinct from the U2OS WT and U2OS ASPP2 KO cells (Figure 1), note that one sample from U2OS WT and U2OS ASPP2 KO did not cluster well with the other two.

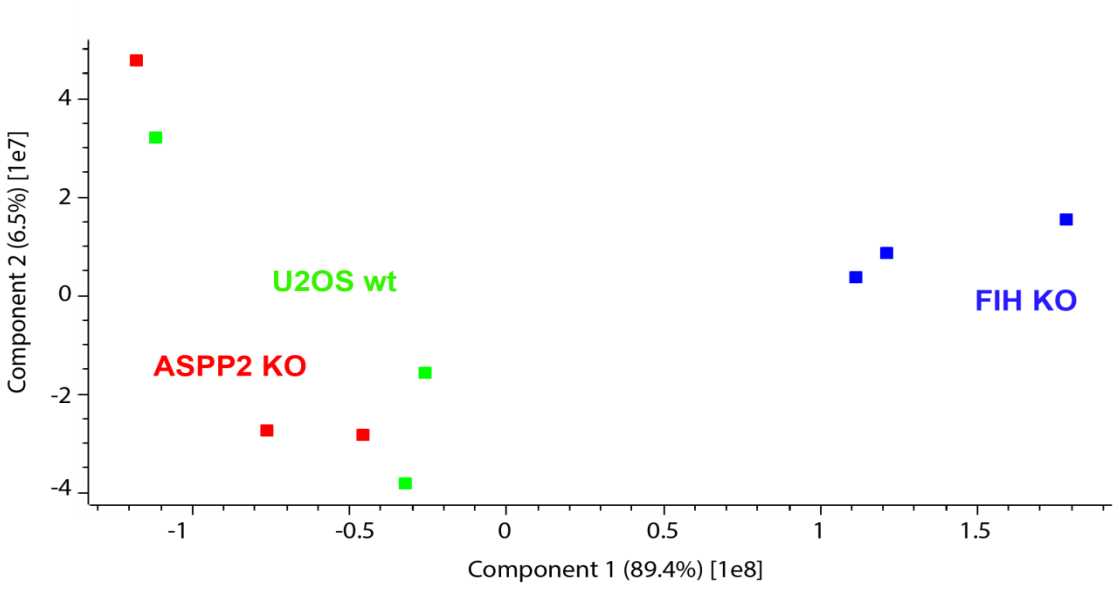


Figure 1 PCA of the proteomic analysis of U2OS WT, U2OS ASPP2 KO, and U2OS FIH KO cells. Data from proteomic analysis of U2OS WT, ASPP2 KO and FIH KO samples. Proteins identified based on only a single peptide were excluded from the analysis. PCA generated using Perseus software.

In total, 122 proteins were found to be significantly upregulated ($p<0.05$) in the U2OS FIH KO, as compared to U2OS WT, based on a *t-test* assuming a normal distribution (Figure 2). KEGG Pathway analysis revealed a cluster of proteins involved in oxidative phosphorylation (OXPHOS) were most significant upregulated (see supporting information for chapter 7, Appendix). Notably, this pathway is also of importance for neurodegenerative diseases such as Parkinson's disease, Alzheimer's disease, and Huntington's disease, as well as non-alcoholic fatty acid liver disease (see supporting information for chapter 7, Appendix). A link between FIH and these diseases has not been described (see also chapter 1).

These findings are in accordance with studies using FIH KO mouse models suggested that FIH may have a role in the regulation of mitochondrial function (see chapter 1.7). FIH KO mice show decreased body weight and an increased resistance to high fat diet[6], possibly resulting from alterations of mitochondrial function. Further investigations are needed to confirm this potential link and to examine the mechanism by which FIH regulates metabolism.

The relative abundance of mitochondrial proteins found in higher abundance in U2OS FIH KO cells is shown in Figure 3. The impact on the protein levels involved in cell metabolism is most dominant in FIH KO cells compared to U2OS WT, which is in accordance to FIHs role as negative transcriptional regulator[7,8]. However, some proteins, like the eukaryotic translation initiation factor 3 and adenylosuccinate lyase, were found in lower abundance in U2OS FIH KO as compared to U2OS WT (Figure 3). Further experiments are needed to confirm these findings.

HIF is known to have a profound impact on cell metabolism, by upregulating glycolysis and downregulating the TCA cycle and OXPHOS (see chapter 1). It is therefore likely that these effects are HIF-α transcriptionally induced, but an HIF-independent mechanism is also possible.

Under the similar conditions, U2OS ASPP2 KO cells did not show significant enrichment of upregulated proteins within a particular pathway in comparison to U2OS WT cells (based on KEGG pathway analysis; data not shown). The effect of ASPP2 on expression of certain target genes was previously demonstrated for its interactions with p53[9-11]. U2OS cell have an intact p53 pathway but low basal p53 levels. Therefore, changes in the protein level of pro-apoptotic proteins would only be expected upon p53 induction. Notably, no correlation with the FIH phenotype was observed.

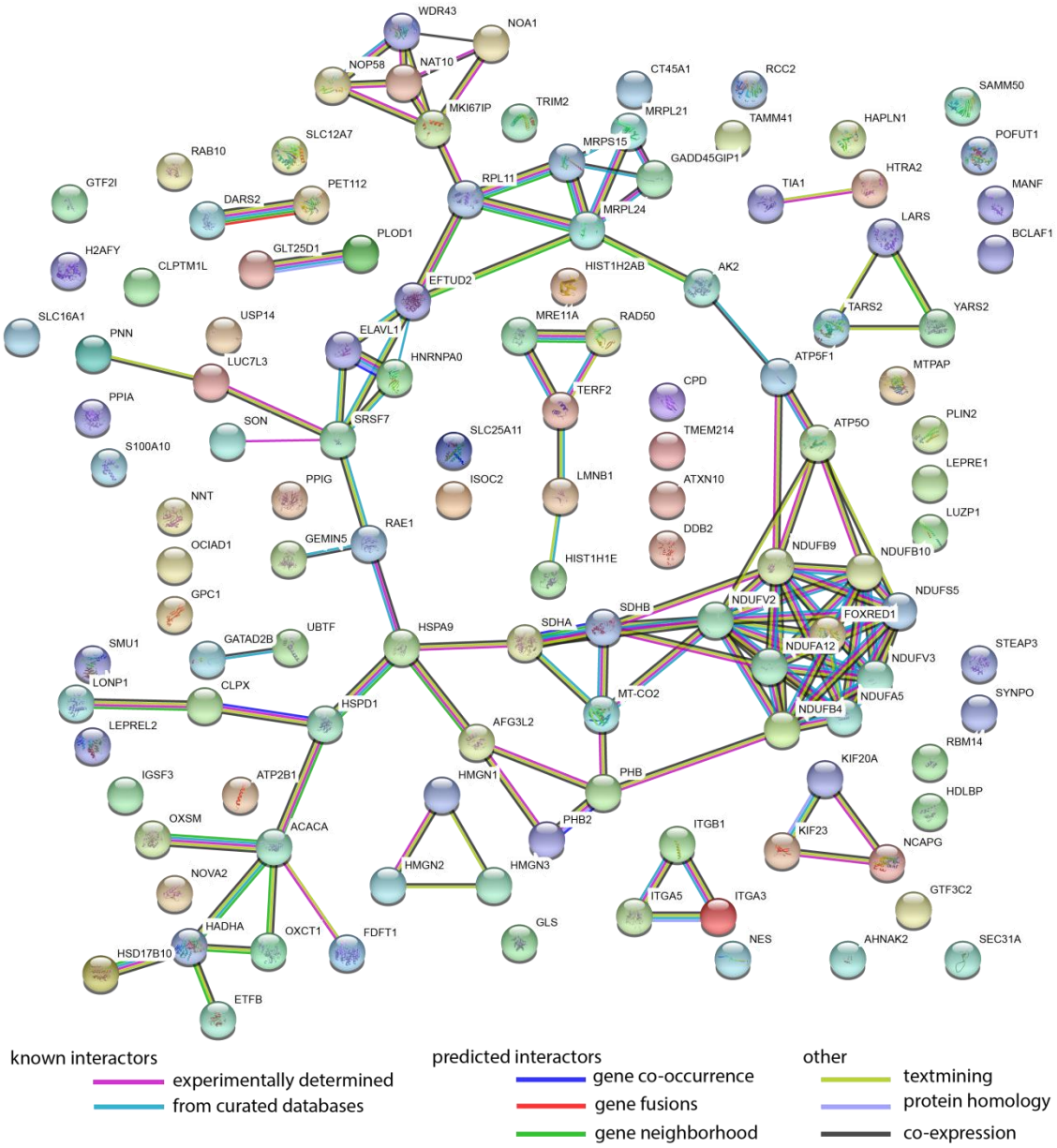


Figure 2 Proteins detected in higher abundance in U2OS FIH KO cells as compared to U2OS WT cells. Protein-protein interactions of proteins found in higher abundance in U2OS FIH KO cells compared to U2OS WT according to the STRING database[12,13]. Image generated using String10.5 web server.

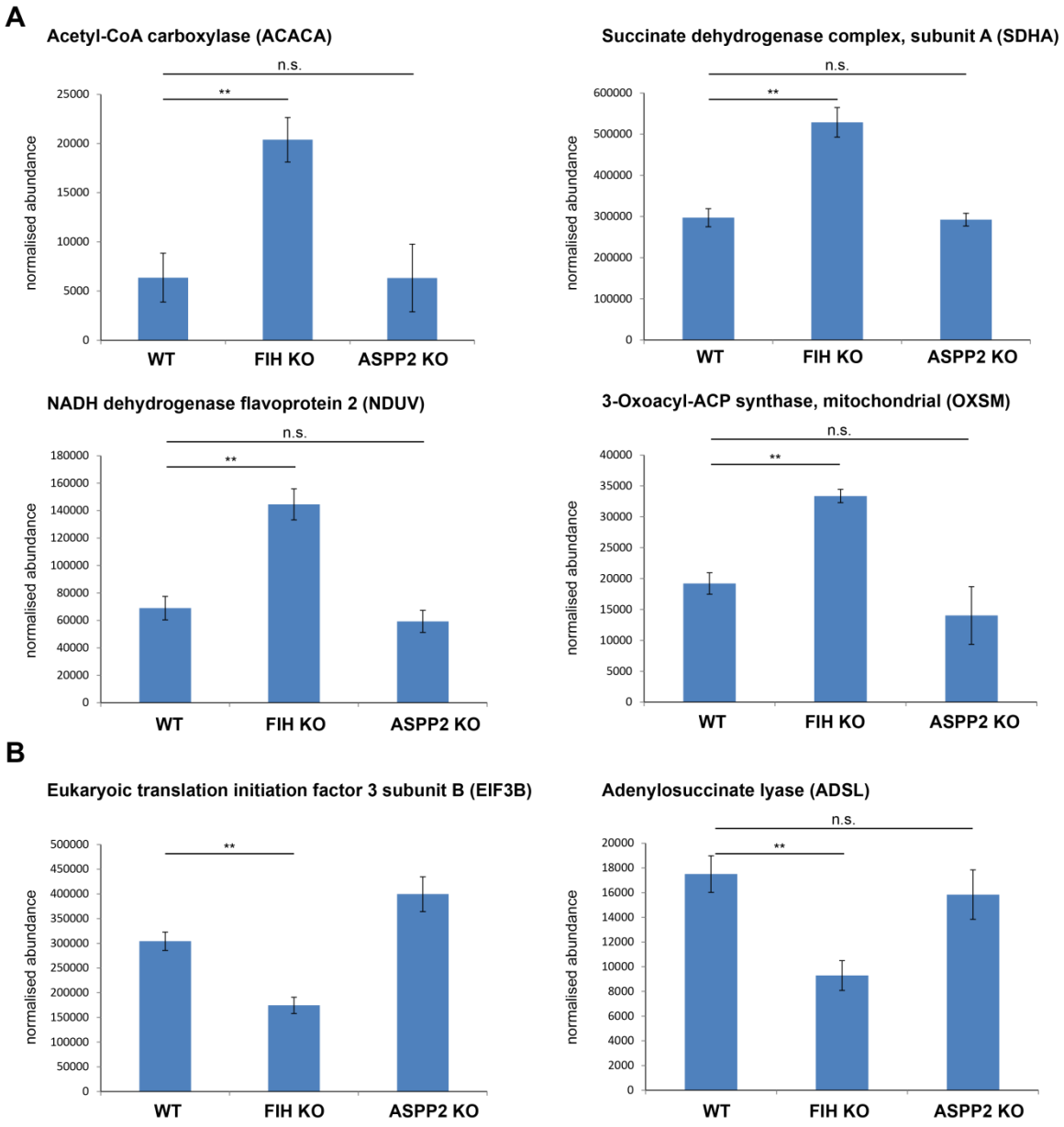


Figure 3 Proteins affected by FIH KO in U2OS cells. A Proteins detected in higher abundance in U2OS FIH KO cells compared to U2OS WT and U2OS ASPP2 KO cells. **B** Proteins detected in lower abundance in U2OS FIH KO cells compared to U2OS WT and U2OS ASPP2 KO cells. Protein abundance normalised to housekeeping proteins. Error bars represent the standard deviation based on three independent experiments. Statistical analysis: n.s. not statistically significant, ** $p<0.01$).

7.3 ASPP2-V5 interactomes in U2OS cells

Proteomic analysis confirmed that ASPP2 KO does not have a significant effect on the proteome of U2OS cells. This allows U2OS ASPP2 KO cells for interactions studies between endogenous ASPP2 WT and different variants of ASPP2. To test the effect of FIH mediated hydroxylations on ASPP2, ASPP2 tagged with the small epitope of P and V proteins of paramyxovirus (ASPP2-V5) was transfected in U2OS ASPP2 KO and U2OS ASPP2 KO/FIH KO (DKO) cells, alongside with alanine variants of the FIH hydroxylation sites of ASPP2-V5: N984A, N986A and N984A/N986A and the PP1-binding deficient variant ASPP2-V5 V922A/F924A (RAKA). Alanine variants of proteins are frequently used in the literature as a protein mimic that cannot be hydroxylated[14,15]. However, the position of the hydroxylation site within an ankyrin repeats is structurally important and raises questions whether the effect on protein stability may be predominantly responsible for effects observed (see also chapter 6). ASPP2-V5 RAKA, which was reported to be unable to co-immunoprecipitate PP1, was included as a control for differential substrate binding[16,17].

ASPP2-V5 and variants thereof were immunoprecipitated (24 h following transfection, with assistance of Mr James Chan, University of Oxford) under mild conditions using anti-V5 antibody (for detailed protocol see chapter 10.4). The samples were washed, and the proteins eluted from the beads using low pH glycine buffer. Proteins were prepared for in-solution digest, treated with trypsin, and the resulting peptides were analysed using LC-MS/MS (see chapter 10). The identified peptides were searched against the human proteome (UniProt[18] database) and assigned using PEAKS7[4] software. LC-MS/MS runs from all sampless were aligned, and proteins quantified using label-free quantification with Progenesis (Nonlinear Dynamics, Newcastle, UK). The different IPs were normalised against all proteins identified in the IP. Note that one sample from the ASPP2-V5 RAKA IP was removed from the analysis as an outlier (total protein amount >100x the average of all other IPs).

Comparison between the proteins identified in the V5-IP of ASPP2-V5 WT and the empty pcDNA3.1 vector control is shown as a Venn diagram in Figure 4. For the Venn diagram analysis, all proteins with >3 spectral counts were used in the analysis. Proteins present in all three repeats were counted as a hit. Samples present in both, the control and ASPP2 samples were disregarded as background. In total, 41 proteins were identified as potential ASPP2 interactors. Ribosomal proteins and keratins, common contaminants in IPs, were removed from the list, yielding 21 potential interactors (Table 1). The identified proteins include ASPP2, as well as known interactors including YAP1[16,17,19], PP1[16,17,20] and PARD3[21,22].

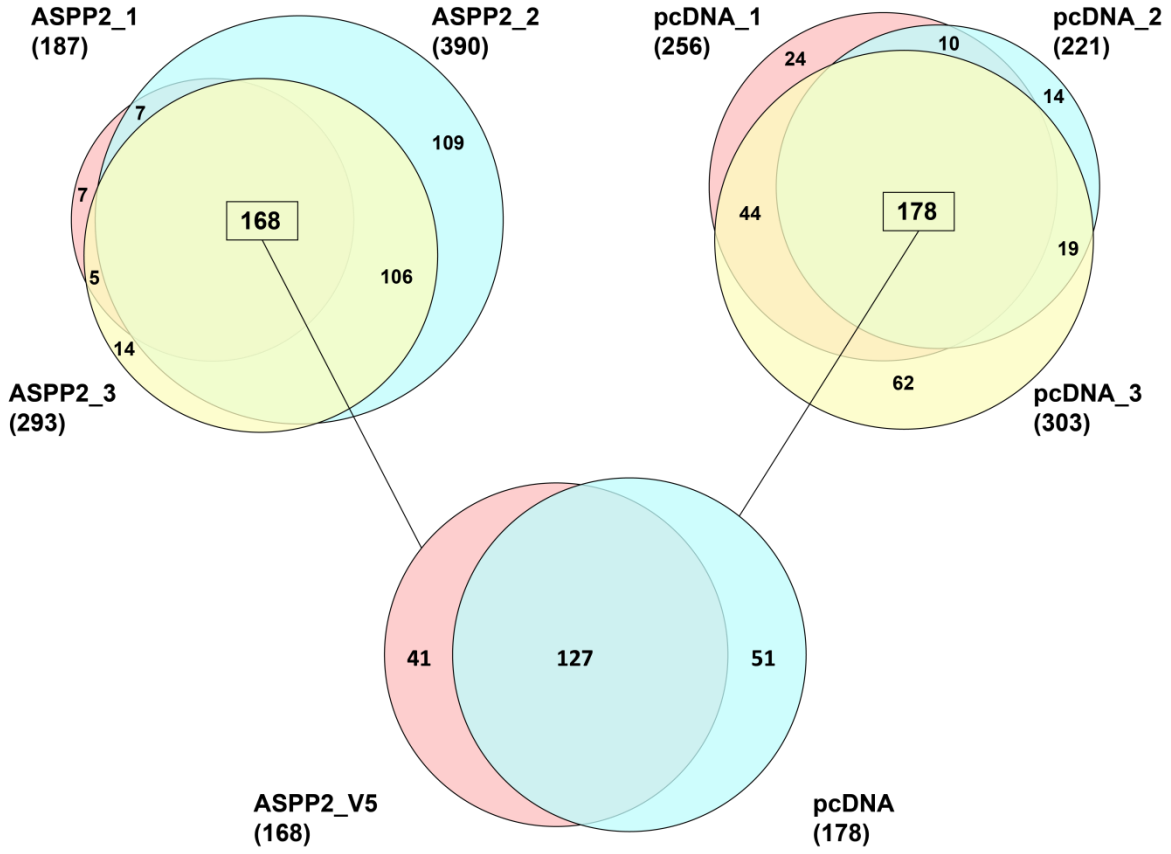


Figure 4 Venn diagram analysis of ASPP-V5 IP in U2OS cells. The numbers in brackets indicate the number of identified proteins. Proteins with <4 spectral counts were disregarded for this analysis. Only proteins identified in all repeats by PEAKS7 (see above) were counted as hits. Hits found in both the control empty pcDNA3.1 vector sample and the ASPP2 sample were disregarded from further analysis.

To investigate whether proteins interacting with ASPP2 cluster in particular cellular pathways, KEGG pathway analysis of the identified potential ASPP2 interactors was performed. The most highly enriched pathway of ASPP2 interacting proteins identified is the HIPPO pathway, consistent with the role of ASPP2 in the regulation of the HIPPO pathway[17,23]. In addition to proteins in the HIPPO pathway (YAP1, PP1s and PARD3, see above), several proteins involved in further cellular functions of ASPP2 were found including pericentriolar material 1 protein (PCM1; involved in the kinetochore complex[24,25]), caveolin-1 (CAV; a scaffolding protein of caveolar membranes[26,27] and previously described ASPP2 interactor; Lu and co-workers, unpublished data) and Ras associated domain-containing protein 8 (RASSF8) which is important for adherence junction maintenance[23,28,29]. Notably, due to the mild conditions employed in cell lysis, a bias towards cytoplasmic protein-protein interactions is possible.

Table 1 Potential ASPP2 interactors in U2OS cells according to Venn diagram analysis of ASPP2-V5 IP in U2OS cells.

Protein	Description
ASPP2	Apoptosis-stimulating of p53 protein 2
BAG2	BAG family molecular chaperone regulator 2
CAV1	Caveolin-1
CC85B	Coiled-coil domain-containing protein 85B
CC85C	Coiled-coil domain-containing protein 85C
TRI58	E3 ubiquitin-protein ligase TRIM58
GRN	Granulins
H14	Histone H1.4
HORN	Hornerin
PARD3	Partitioning defective 3 homolog
PCM1	Pericentriolar material 1 protein
PTRF	Polymerase I and transcript release factor
RASSF7	Ras association domain-containing protein 7
RASSF8	Ras association domain-containing protein 8
RCN1	Reticulocalbin-1
PP1A	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit
PP1B	Serine/threonine-protein phosphatase PP1-beta catalytic subunit
PP1G	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit
YAP1	Transcriptional coactivator YAP1
TBB5	Tubulin beta chain
ZBTB2	Zinc finger and BTB domain-containing protein 2
ZN639	Zinc finger protein 639

Using label-free quantification[30-32], abundances of proteins co-immunoprecipitated with ASPP2 WT were compared to the ASPP2 variants (Figure 5 and Figure 6). Notably, proteins found to be more abundant in the ASPP-V5 IP over the empty vector control were also identified in the Venn diagram analysis as potential interaction partners. Additionally, FERM domain-containing protein 6 (FRMD6), a structural constituent of the cytoskeleton found at the plasma membrane, previously reported as ASPP2 interactor[33,34], was found to be enriched in the ASPP2-V5 IP compared to the empty vector control.

IP-Western blot (IP-WB) validation ASPP2 interactions with PP1α, RASSF8, YAP1, and FIH is shown in Figure 7A.

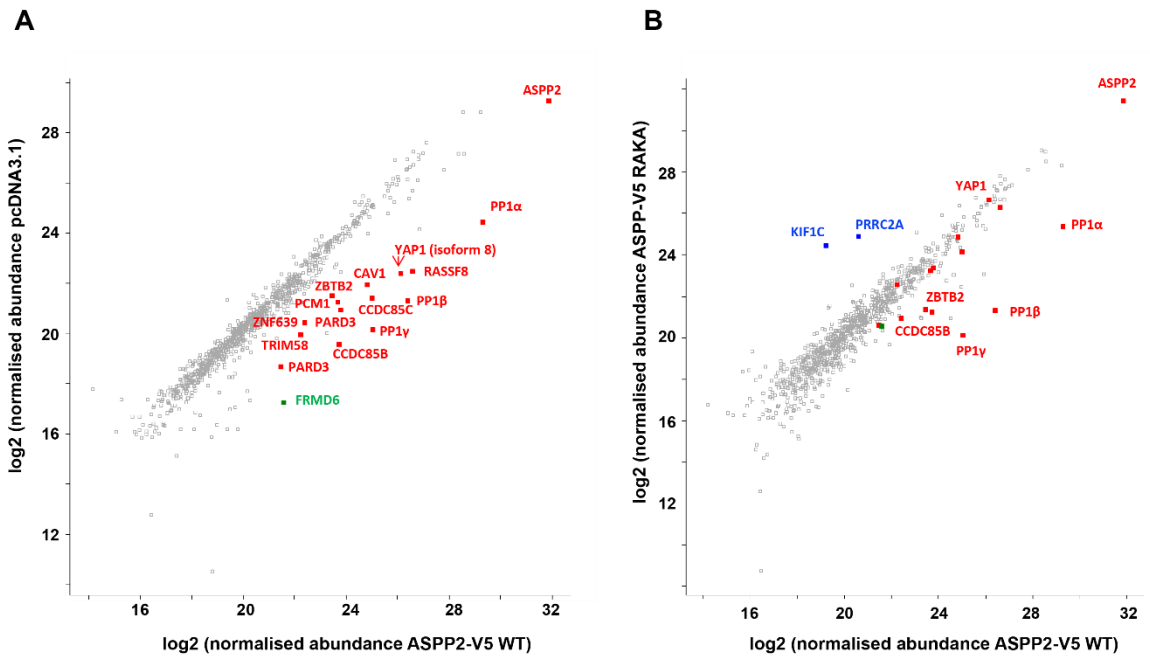


Figure 5 Normalised abundances of proteins co-immunoprecipitated with ASPP2-V5. **A** Proteins co-immunoprecipitated with ASPP2-V5 WT versus empty pcDNA3.1 vector. **B** Proteins co-immunoprecipitated with ASPP2-V5 WT versus ASPP2-V5 RAKA. Each dot represents an identified protein. Proteins of interest are coloured red for previously described proteins that are enriched in ASPP2-V5 WT IP, green for those enriched in ASPP2-V5 IP but not previously described, and blue for those enriched in ASPP2-V5 RAKA IP over ASPP2-V5 WT.

Comparison of the results for ASPP2-V5 WT with ASPP2-V5 RAKA indicate that the ability to pull-down all three isoforms of PP1 is impaired for the ASPP2 PP1 binding deficient ASPP variant at similar ASPP2 protein levels (Figure 6). Reduced levels of ZBTB2 and CCDC85B were co-immunoprecipitated with ASPP2-V5 RAKA compared to ASPP2-V5 WT. Western blot analysis did not yield conclusive results primarily due to poor quality of the antibodies available (Mr James Chan, University of Oxford). ZTBT2 was shown be phosphorylated[35], and CCDC85B has three predicted phosphorylation sites (based on modifications of orthologues according to Phosphosite Plus[36]). Whether these sites are targeted by the PP1-ASPP2 complex for dephosphorylation and thereby regulate their binding to ASPP2 remains unknown.

Strikingly, two proteins were observed in the ASPP2-V5 RAKA IP, but not in the ASPP2-V5 IP; kinesin-like protein KIF1C (KIF1C), a motor protein involved in the retrograde transport of Golgi vesicles to the endoplasmic reticulum[37] and protein PRRC2A (PRRC2A) which is associated with mRNA splicing[38].

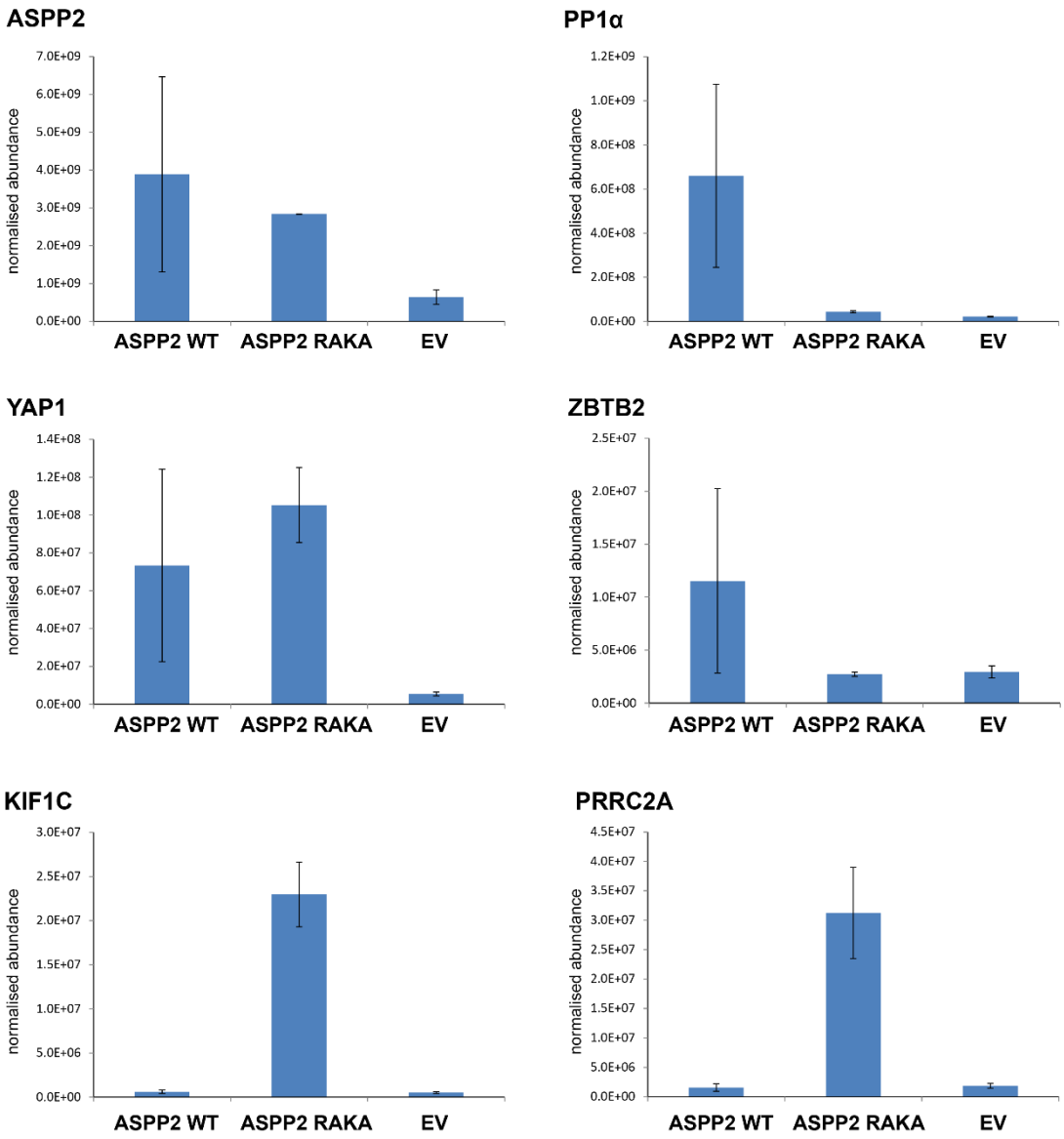


Figure 6 Normalised abundances of proteins co-immunoprecipitated with ASPP2-V5 WT and ASPP2-V5 RAKA from U2OS ASPP2 KO cells. Error bars represent the standard deviation from 2-3 independent experiments. Due to bias from normalisation and different sample numbers, no statistical analysis was performed. EV= empty pcDNA3.1 vector.

The differential co-immunoprecipitation of KIF1C with ASPP2-V5 WT or ASPP2-V5 RAKA was confirmed by IP-WB (Mr James Chan, University of Oxford, Figure 7B). Both, PRRC2A and KIF1C are phosphorylated on multiple serine and threonine residues[35,39-46], and could potentially be substrates for the ASPP2-PP1 phosphatase complex. Alternatively, increased binding to ASPP2-V5 RAKA over WT was observed in IP-WB for other proteins whose binding sites were reported overlap to with PP1 binding[17] (Figure 7A), potentially due to steric clashes; this could also be the case for KIF1C and PRRC2A. Further studies are needed to investigate the mechanism by which PP1 binding deficiency facilitates KIF1C and PRRC2A binding to ASPP2.

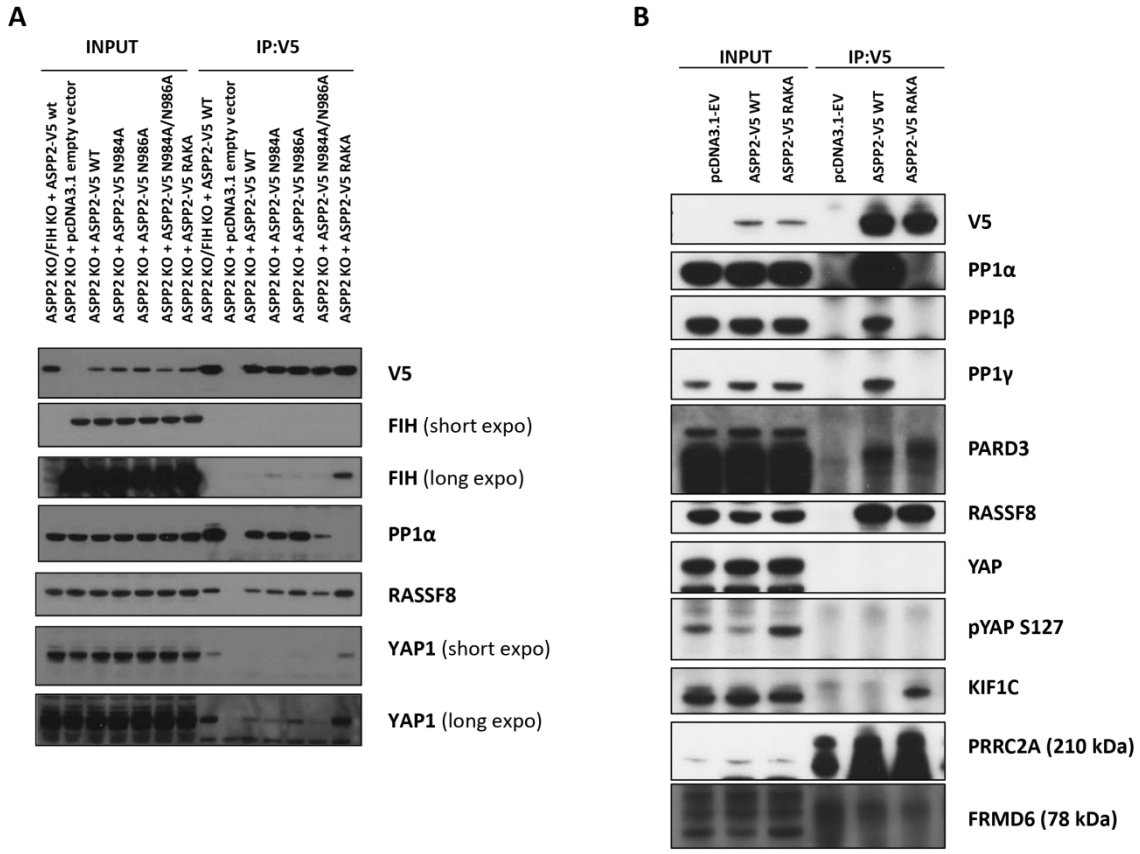


Figure 7 Validation of MS hits by IP-WB. A Western blot analysis of reported ASPP2 interactors that were co-immunoprecipitated with ASPP-V5 WT and ASPP2 variants from U2OS ASPP2 KO and DKO cells. **B** Validation for differential co-immunoprecipitation of ASPP2 interactors from U2OS ASPP2 KO cells (performed and blotted by Mr James Chan, University of Oxford).

ASPP2 and PP1 form an active diphosphatase complex with ASPP2 functioning as a regulatory subunit, involved in substrate recognition and binding, and PP1 as a catalytic subunit[20,47]. By disruption of the complex, an increase in phosphorylation of targets of the phosphatase complex would be expected, assuming no functional redundancy of PP1-phosphatase complexes.

ASPP2 has been shown to be phosphorylated on multiple residues[39,41,42,44,45]. Disruption of the ASPP2-PP1 complex likely has an impact on the ASPP2 phosphorylation status, which would lead to increased ASPP2 phosphorylation in the PP1-binding deficient RAKA variant.

In total, 319 different peptides from both ASPP2-V5 and ASPP2-V5 RAKA were identified in the IP; 23 of these peptides were phosphorylated (Figure 8A). 5 phosphopeptides corresponding to serine-714, serine-737 and serine-412 were enriched in the ASPP2-V5 RAKA samples compared to the ASPP2-V5 WT. Representative MS/MS spectra of phosphorylated peptides are shown in Figure 9. None of the phosphopeptides were enriched in the ASPP2-V5 WT samples as compared to the ASPP2-V5 RAKA samples. The normalised abundances of the phosphorylated and the corresponding unmodified peptides are shown in Figure 8.

A possible explanation for the enrichment of certain phosphorylation sites on the ASPP2-V5 RAKA variant is an auto-dephosphorylation mechanism catalysed by the ASPP2-PP1 phosphatase complex. While the phosphorylation sites on ASPP2 are several hundred amino acids from the PP1 binding site, the intramolecular interaction of the ASPP2 ARD and SH3 domains haven been reported[16,17,20], potentially bring the phosphorylation sites close to the phosphatase active site. However, an intermolecular mechanism is possible. Notably, only three phosphorylation sites on ASPP2 were found to be enriched in the PP1-binding deficient variant. This does not exclude further PP1-regulated phosphorylation sites on ASPP2 but suggests at least some degree of selectivity of PP1 mediated ASPP2 dephosphorylation. Whether auto-dephosphorylation could be a self-regulatory mechanism of the ASPP2-PP1 phosphatase complex remains to be elucidated.

The physiological role of the identified phosphorylation sites on ASPP2 has not yet been identified. Notably, phosphorylation sites serine-737 and serine-714 are close to the binding site of *Helicobacter pylori* (*H. pylori*) cytotoxin-associated gene A (CagA) with ASPP2 (746-765)[48]. ASPP2-CagA interactions are believed to be important for CagA-mediated p53 degradation and its oncogenic role in *H. pylori* infections.

Following from the analysis of the ASPP2 post-translational modifications (PTMs), PTMs on ASPP2 interacting proteins were analysed. Amongst the strongest ASPP2 interacting proteins are the PP1s. The enzymatic activity of PP1 is modulated by binding of regulatory subunits and post-translational modifications of PP1[47,49-51]. None of the reported PP1 PTMs were however detected for any of the PP1 isoforms in the IP (data not shown). Whether the reason for the lack of PP1 PTMs is the low frequency of the modification, incomplete sequence coverage in the MS/MS experiment, or direct influence on the PP1-ASPP2 complex formation or stability remains unclear.

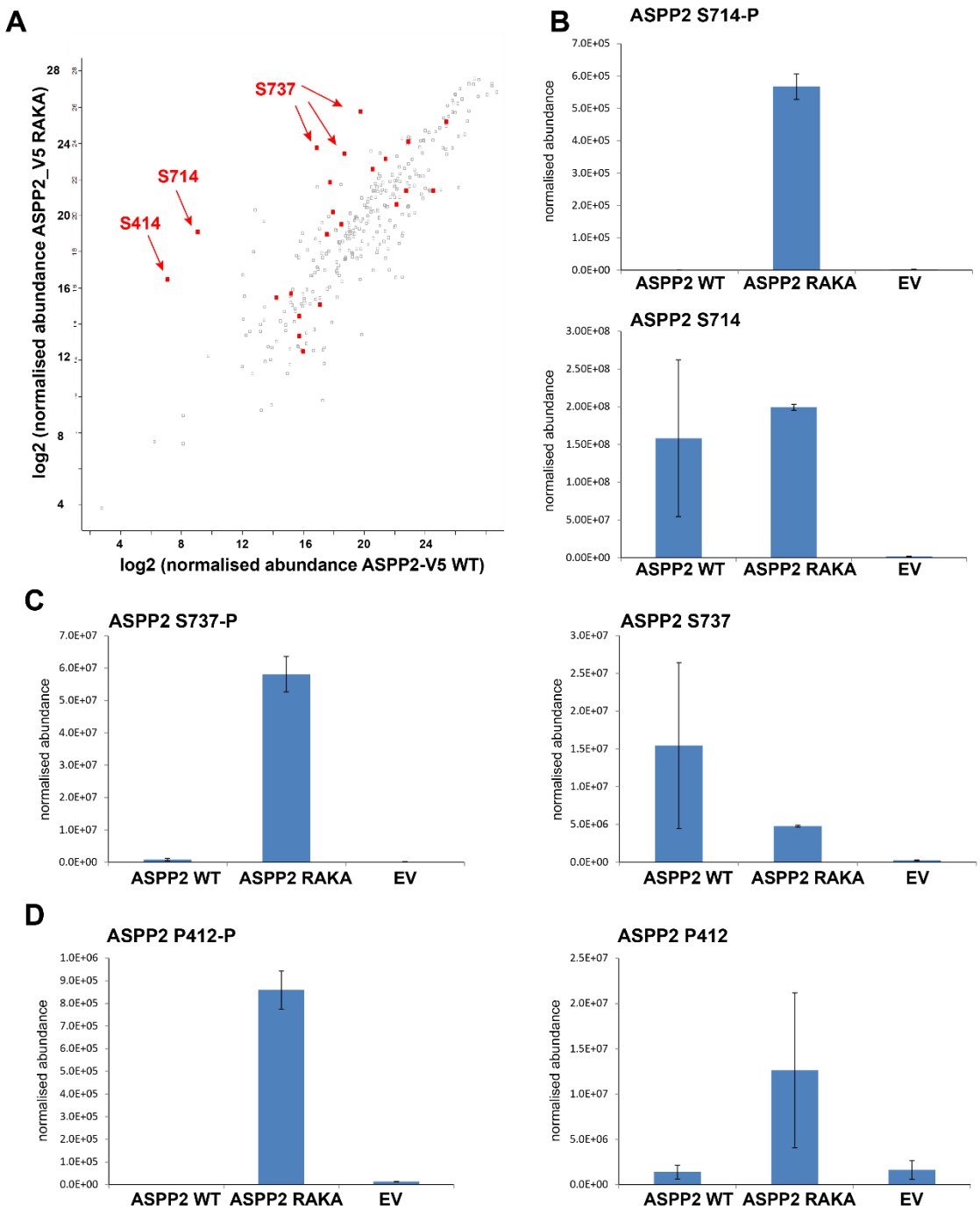
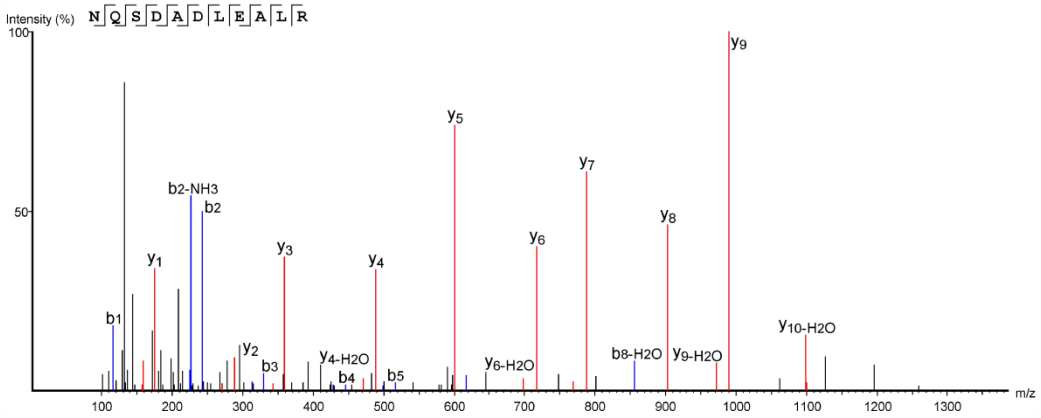


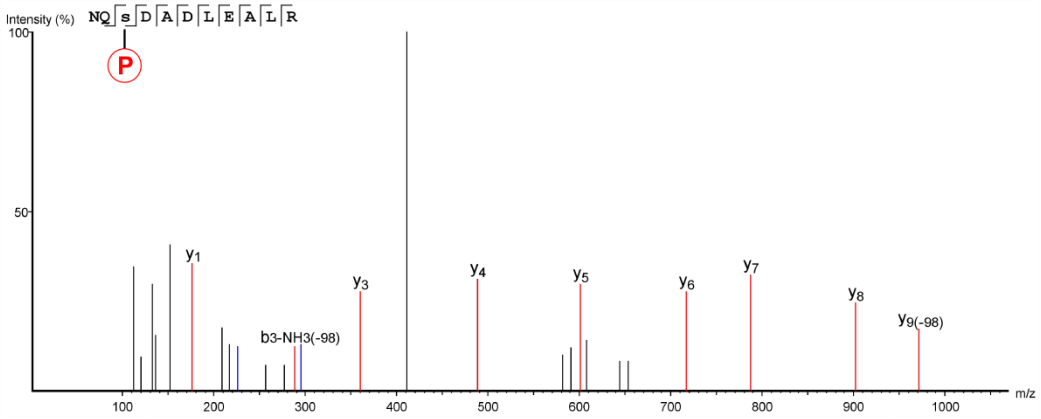
Figure 8 Differential phosphorylation of ASPP-V5 and ASPP2-V5 RAKA variant. A Normalised abundance of peptides corresponding to ASPP2 immunoprecipitated from ASPP2-V5 WT versus ASPP2-V5 RAKA. Each dot represents a single peptide. Phosphorylated peptides are marked in red. **B-D** Normalised abundance of selected phosphorylated ASPP2 peptides and their corresponding unmodified peptides from immunoprecipitated ASPP2-V5 WT versus ASPP2-V5 RAKA. Error bars represent the standard deviation from 2-3 independent experiments. Peptide abundance normalised to all identified proteins (see above).

A



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	115.05	97.04	98.02	58.03	N					11
2	243.11	225.10	226.08	122.05	Q	1117.55	1099.53	1100.52	559.27	10
3	330.14	312.13	313.11	165.57	S	989.49	971.48	972.46	495.24	9
4	445.17	427.16	428.13	223.08	D	902.46	884.45	885.43	451.73	8
5	516.24	498.19	499.17	258.60	A	787.43	769.41	770.40	394.22	7
6	631.23	613.22	614.21	316.12	D	716.39	698.38	699.37	358.70	6
7	744.32	726.31	727.29	372.66	L	601.36	583.36	584.34	301.18	5
8	873.36	855.43	856.33	437.18	E	488.28	470.27	471.26	244.64	4
9	944.40	926.39	927.37	472.70	A	359.24	341.23	342.21	180.12	3
10	1057.48	1039.47	1040.45	529.24	L	288.20	270.19	271.18	144.60	2
11					R	175.12	157.13	158.09	88.06	1

B



#	b	b-H2O	b-NH3	b (2+)	Seq	y	y-H2O	y-NH3	y (2+)	#
1	115.05	97.04	98.02	58.03	N					11
2	243.11	225.10	226.08	122.05	Q	1197.51	1179.50	1180.49	599.26	10
3	410.11	392.10	393.08	205.55	S(+79.97)	1069.46	1051.45	1052.43	535.23	9
4	525.13	507.12	508.11	263.07	D	902.46	884.45	885.43	451.73	8
5	596.17	578.16	579.14	298.59	A	787.42	769.42	770.40	394.22	7
6	711.20	693.19	694.17	356.10	D	716.39	698.38	699.37	358.70	6
7	824.28	806.27	807.26	412.64	L	601.36	583.36	584.34	301.18	5
8	953.33	935.31	936.30	477.16	E	488.28	470.27	471.26	244.64	4
9	1024.36	1006.35	1007.34	512.68	A	359.24	341.23	342.21	180.12	3
10	1137.45	1119.44	1120.42	569.22	L	288.20	270.19	271.18	144.60	2
11					R	175.12	157.11	158.09	88.06	1

Figure 9 MS/MS spectra for ASPP2 serine-714 phosphorylation. A MS/MS spectrum and ion table for an unmodified peptide containing serine-714. **B** MS/MS spectrum and ion table for peptide containing phosphorylated serine-714.

In addition to PP1, YAP1 is an established ASPP2 interactor (see above). In total, 40 peptides corresponding to YAP1 were found detected, of which five were found to be phosphorylated on serine-109, serine-366 and serine-289, in the MS/MS experiment. No sequence coverage was achieved for serine-127, a phosphorylation site reported to be dephosphorylated by the ASPP2-PP1 phosphatase complex[17] (see also Figure 7). The overall levels of endogenous YAP1 co-immunoprecipitated with ASPP2-V5 are similar between the ASPP2-V5 WT and the ASPP2-V5 RAKA variant samples. However, the data quality was insufficient to draw conclusions about potential PTM enrichment in either the ASPP2-V5 WT or ASPP2-V5 RAKA sample.

Overall, the ASPP2-V5 IP has demonstrated that differentially co-immunoprecipitated proteins can be detected and quantified, at least in the case of ASPP2 and the ASPP2 RAKA variant, using label-free quantitative mass spectrometry. While PTMs of endogenous interaction partners could be detected, the levels were too low to compare quantities between different samples.

Following from the results of the ASPP-V5 IP in U2OS ASPP2 KO cells, it was then tested whether differences in the ASPP2 interactome for unhydroxylated ASPP2 and hydroxylation site variants could be observed. ASPP2 is hydroxylated in an FIH-dependent manner at asparagine-984 and asparagine-986, the latter of which is the preferred hydroxylation site (see chapter 6). To investigate the effect of FIH-mediated hydroxylation, ASPP2-V5 WT, ASPP2-V5 N984A, ASPP2-V5 N986A and ASPP2-V5 N984A/N986A variants were transfected into U2OS ASPP2 KO cells, and ASPP2-V5 WT was transfected into U2OS ASPP2 KO / FIH KO (DKO) cells. ASPP2 was immunoprecipitated using an anti-V5 antibody as previously described.

Interestingly, both ASPP2-V5 N984A and ASPP2 N986A showed similar interactomes compared to ASPP2-V5 WT (Figure 10). The two alanine variant, ASPP2-V5 N984A/N986A showed reduced binding to some interacting proteins, including PP1α and YAP1, but possibly not PARD3 (Figure 11). Both, YAP1 and PP1α interact with the C-terminal domains of ASPP2[19,20], PARD3 binds to the N-terminus of ASPP2[21] and appears to be less effected by the mutation at the C-terminus (Figure 11). This could result from destabilisation of the C-terminal fold caused by the alanine mutations (for discussion of the structural importance of asparagine-984 and asparagine-986, see chapter 6). Interactions between FIH and ASPP2 were not detected for any of the ASPP2 variants or WT samples.

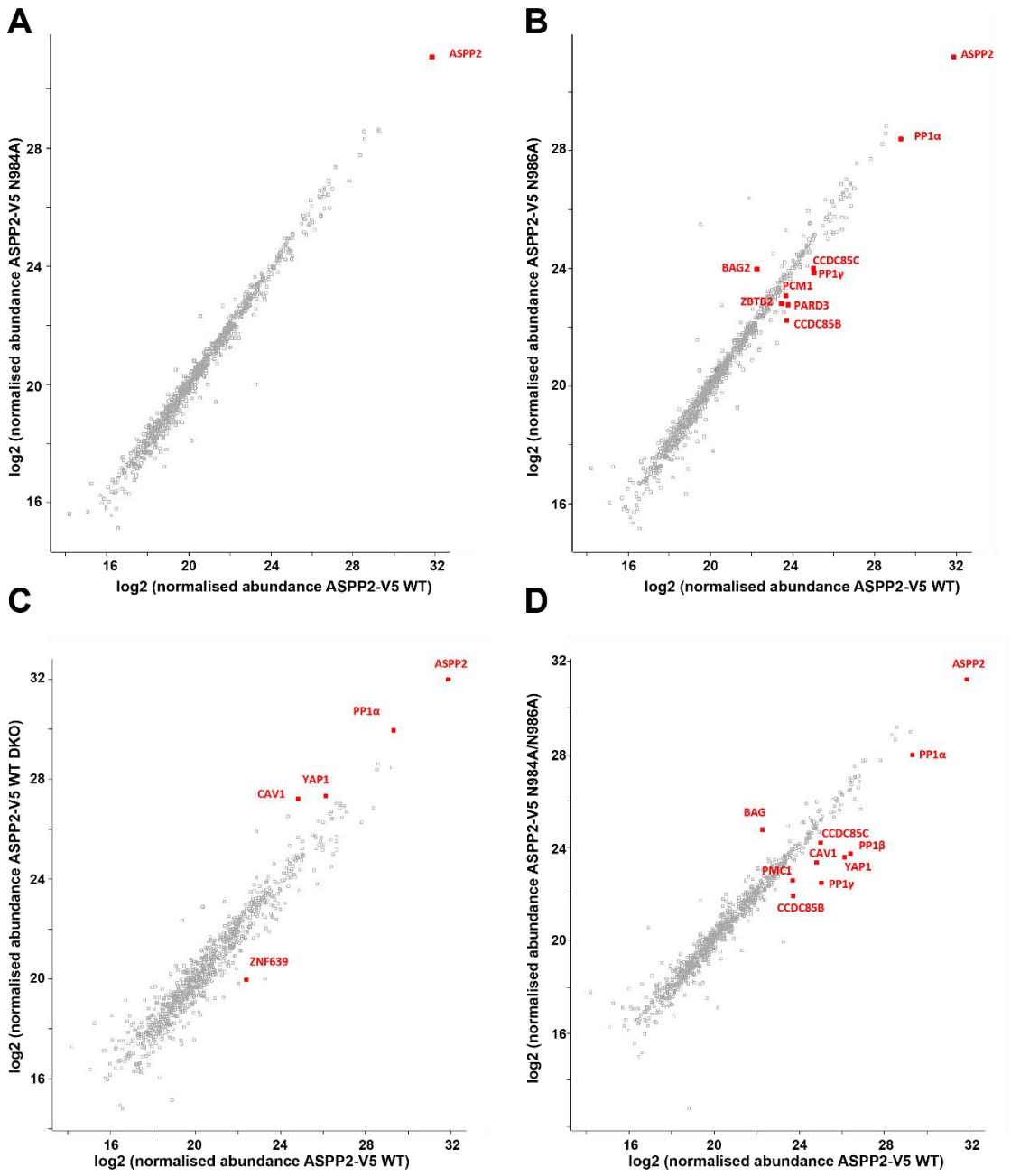


Figure 10 Normalised abundances of proteins co-immunoprecipitated with ASPP2 variants compared to ASPP2-V5 WT. Each dot represents the average normalised abundance of identified proteins from three independent experiments; proteins of interest are coloured red. **A** Comparison of the IP of ASPP2-V5 WT versus ASPP-V5 N984A in U2OS ASPP2 KO cells. **B** Comparison of the IP of ASPP2-V5 WT versus ASPP-V5 N986A in U2OS ASPP2 KO cells. **C** Comparison of the IP of ASPP2-V5 WT in U2OS ASPP2 KO cells versus ASPP-V5 WT in U2OS DKO cells. **D** Comparison of the IP of ASPP2-V5 WT versus ASPP-V5 N984A/N986A in U2OS ASPP2 KO cells.

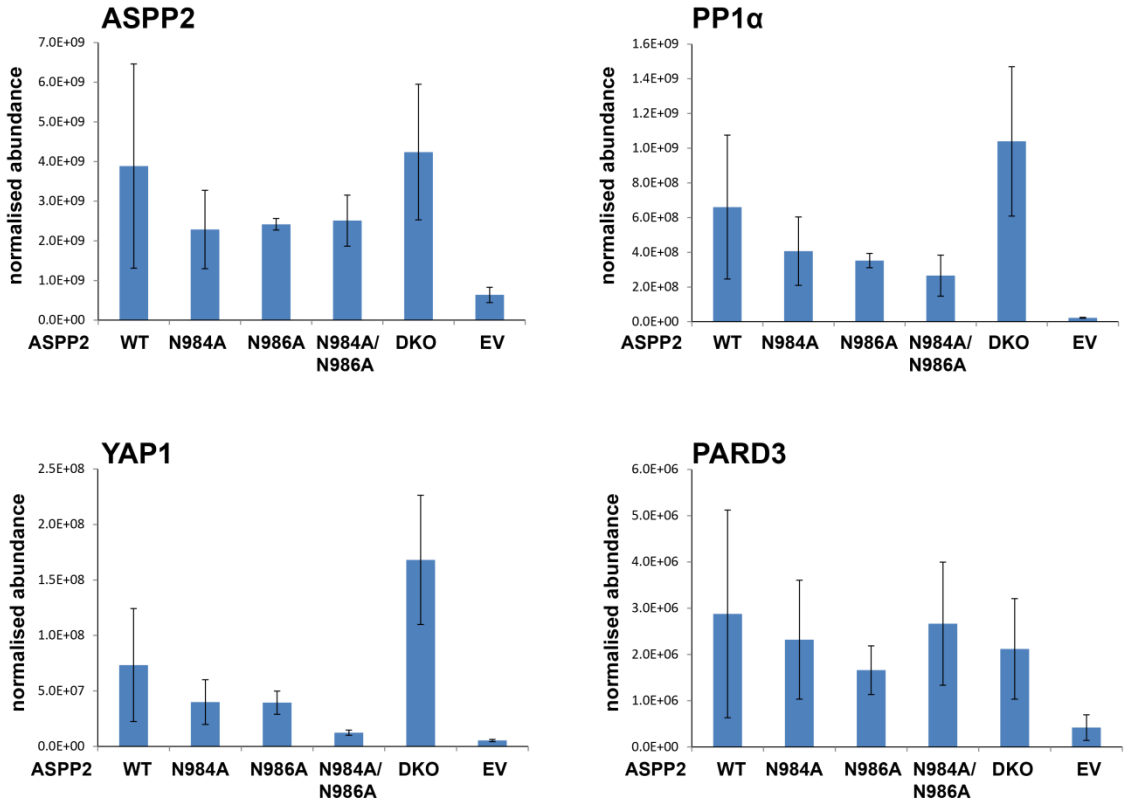


Figure 11 Normalised abundances of proteins co-immunoprecipitated with ASPP2-V5 and ASPP2-V5 alanine variants of the FIH hydroxylation sites from U2OS ASPP2 KO cells, and ASPP2-V5 WT from U2OS DKO cells. Error bars represent the standard deviations from 3 independent experiments.

Comparison of the abundances of interacting proteins co-immunoprecipitated with different ASPP2 mutants showed no direct evidence for the involvement of FIH in the regulation of ASPP2 interaction partners. More importantly, ASPP2-V5 WT immunoprecipitated from U2OS DKO cells showed increased abundance of C-terminal interaction partners PP1α and YAP1 compared to ASPP2-V5 N984A/N986A (Figure 11). Taken together, this would suggest that the alanine variants have a greater effect than the hydroxylations, and are an unsuitable model for non-hydroxylated ASPP2 in a cellular context.

No direct effect of FIH-mediated hydroxylation on ASPP2 interactions has been observed on global levels. ASPP2 interactions are highly regulated, in part due to its subcellular localisation. However, effects of FIH-mediated hydroxylation of ASPP2 on its protein-protein interactions could only affect sub-populations of ASPP2 complexes not detected in the MS/MS experiments, or could require external stimuli, (*e.g.* oxidative stress or infections).

7.4 ASPP2-V5 interactome in AGS cells

Following from the interaction studies of ASPP2-V5 in U2OS cells, it was of interest whether the same trends could be observed in a different epithelial cell line. The gastric adenocarcinoma cell line AGS was chosen because this cell line can be readily infected with *H. pylori* and therefore could be used for the further proteomic studies of the ASPP2 interactome in the context of *H. pylori* infection. Due to the lack of impact of FIH mediated hydroxylation on ASPP2 interactions further studies on the ASPP2-V5 interactions with FIH were not pursued.

Following from the striking effects of the PP1-binding deficient ASPP2-V5 RAKA variant on the interaction with KIF1C and PRRC2A in U2OS cells, it was further tested if the effect could also be observed in AGS cells. AGS ASPP2 KO cells were transfected with ASPP2-V5, ASPP2-V5 RAKA, or an empty vector control and processed as previously described.

Analysis of the anti-V5 IP shows strong enrichment of ASPP2 compared to the empty vector control. PP1α and PP1β as well as RASSF7 were shown to be enriched in the ASPP2-V5 IP compared to the empty vector control (Figure 12).

Comparison of the samples co-immunoprecipitated from ASPP2-V5 WT and the PP1 binding deficient ASPP2-V5 RAKA variant confirmed that no detectable levels of PP1α or PP1β were co-immunoprecipitated with this ASPP2 variant. Further, galactocerebrosidase (GALC) appeared to be enriched in the ASPP WT sample over the ASPP-V5 RAKA sample but this effect was not significant (Figure 12).

The proteins PRRC2A and KIF1C were not detected in the ASPP2-V5 IP from AGS cells. The overall protein levels of PRRC2A and KIF1C were relatively low in the ASPP2-V5 IP from U2OS cells, and the amount of ASPP2 immunoprecipitated from AGS cells was approximately half of the amount of ASPP2 immunoprecipitated from U2OS cells according to spectral and peptide counts (data not shown). This could have caused the low abundant binding proteins to fall under the detection limit. Repeat of the interactions studies with higher amounts of starting material may improve results.

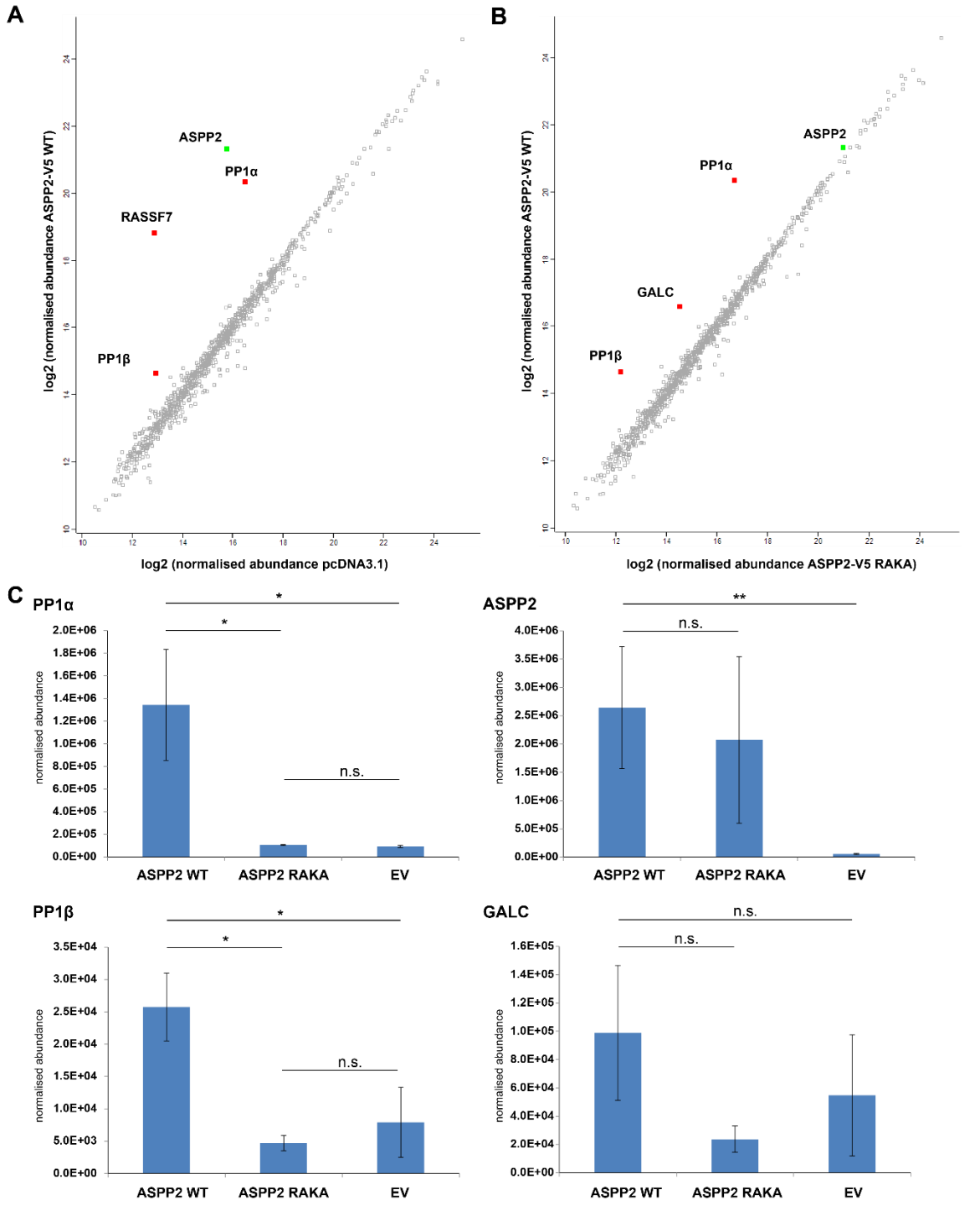


Figure 12 ASPP-V5 interactomes in AGS cells. A Normalised abundance of proteins co-immunoprecipitated with ASPP2-V5 WT versus empty vector control demonstrates that ASPP2-V5 and known interactors were immunoprecipitated. Each dot represents the average normalised abundance of identified proteins from three independent experiments; proteins of interest are coloured red. **B** Normalised abundance of proteins co-immunoprecipitated with ASPP2-V5 WT versus ASPP2-V5 RAKA. **C** Relative abundances of selected proteins co-immunoprecipitated in the V5-IP. Error bars represent the standard deviation from three independent experiments. Statistical analysis: n.s. not statistically significant, ** p<0.01). Data analysis performed with the Perseus proteomic software.

7.5 Endogenous ASPP2 IP

Following from the interactome studies of overexpressed ASPP2-V5 in U2OS ASPP2 KO cells, it was aimed to study the interactome of ASPP2 WT at the endogenous level. As the immunoprecipitation of ASPP2-V5 in U2OS cells yielded promising results, further studies were performed in this cell line.

A set of available antibodies were tested for their ability to immunoprecipitate endogenous ASPP2 with potential interactors (Figure 13). In total five antibodies (three mouse monoclonal and two rabbit polyclonal antibodies) were tested. The rabbit polyclonal antibody bp77 and the mouse monoclonal antibody Dx54.7 immunoprecipitated the greatest amounts of ASPP2 and PP1s. Following further optimisation, bp77 showed higher yields of immunoprecipitated ASPP2 and PP1s compared to Dx54.7.

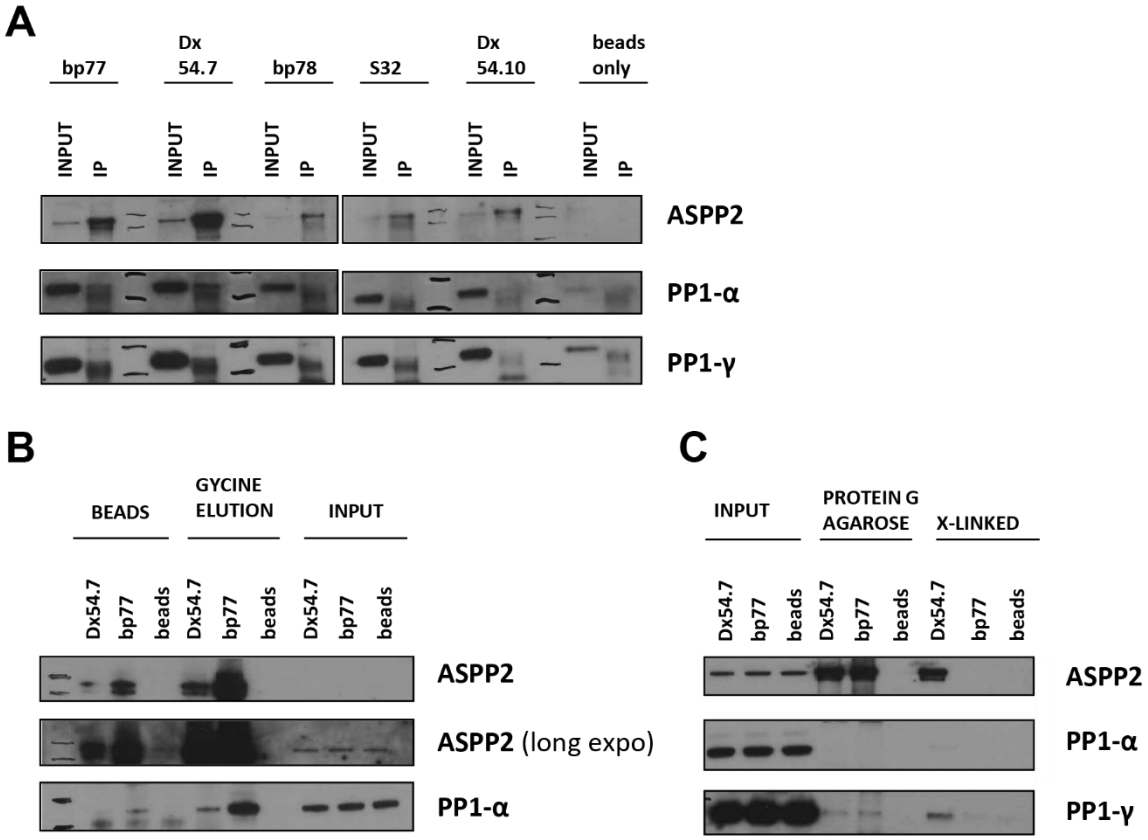


Figure 13 Optimisation of endogenous ASPP2 immunoprecipitation from U2OS cells. A IPs from U2OS cells using different antibodies. Antibodies Dx54.7, Dx54.10 and S32 are mouse monoclonal antibodies and bp77 and bp78 are rabbit polyclonal anti-ASPP2 antibodies. **B** Comparison of elution methods for ASPP2 IPs using bp77 and Dx54.7. For glycine elution, beads were incubated with low pH glycine (pH = 2.5). After glycine elution the beads were boiled in sample buffer and run alongside the glycine elution. **C** Comparison of endogenous ASPP2 immunoprecipitation using cross-linked and not crosslinked antibodies. Samples were eluted using low pH glycine buffer.

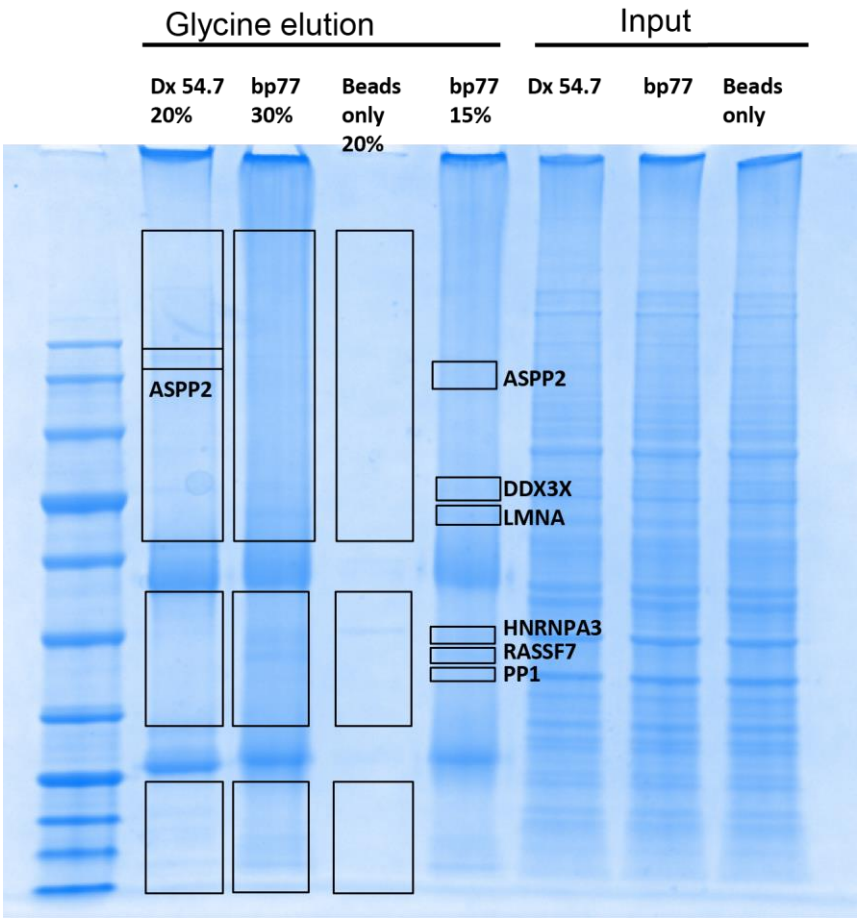


Figure 14 SDS PAGE gels of large scale endogenous ASPP2 IP using non-cross-linked antibodies. Percentages given are the fraction used of from the total IP. Boxes indicate bands that were excised for analysis by mass spectrometry. Protein names from highest scoring proteins identified in the bands using MASCOT[52] are indicated next to the bands.

Both Dx57.7 and bp77 were shown to be capable of immunoprecipitating endogenous ASPP2 and potential interactors from U2OS cells (Figure 14). However, compared to the anti-V5 antibody, the overall yield of immunoprecipitated ASPP2 was very low, therefore requiring substantial amounts of antibody. High antibody contamination was found to be a major obstacle for the use of both antibodies (Figure 14). Bands for ASPP2 and its potential interactors could be observed when the IP was run on an SDS PAGE gel, but the IP had to be diluted to prevent overloading of the gel, resulting in low protein coverage and protein confidence scores. The overall sequence coverage of ASPP2 was ~30-40% and did not cover the hydroxylation site (data not shown).

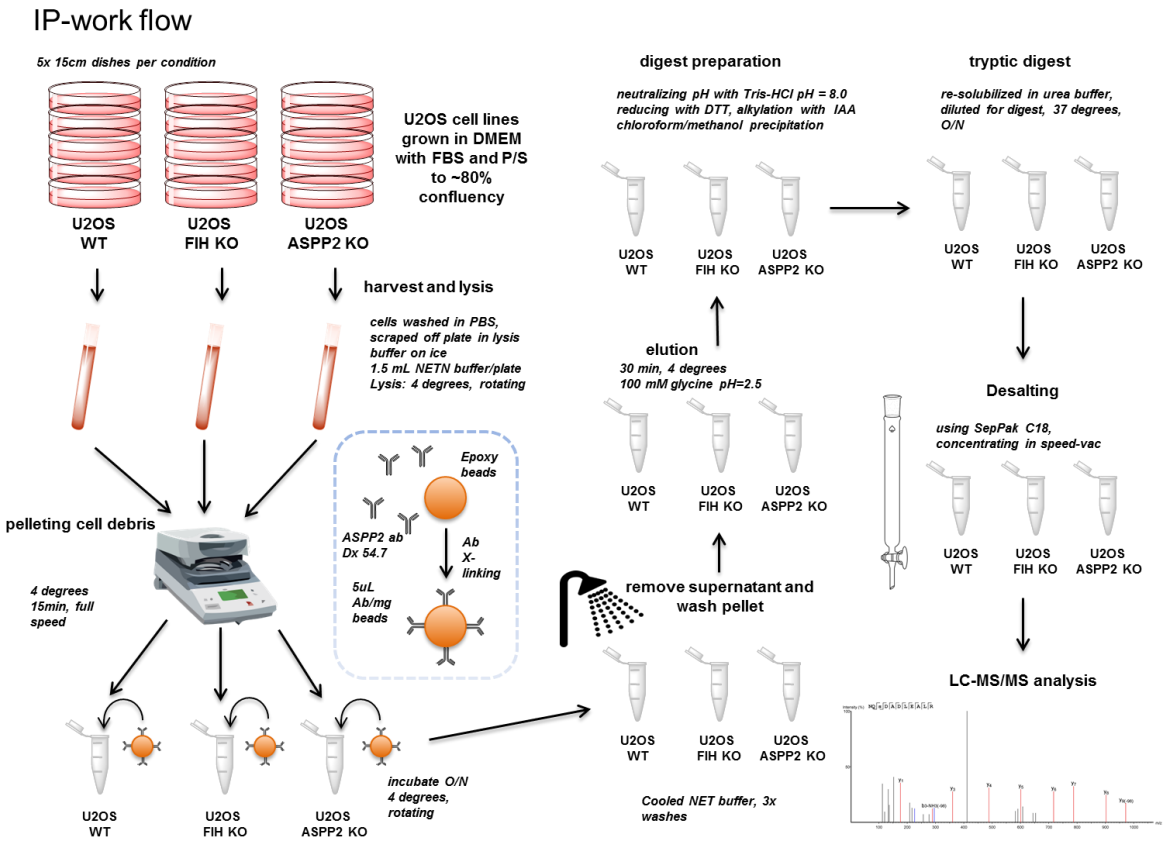


Figure 15 Workflow for endogenous ASPP2-IP. Optimized workflow for endogenous ASPP2 immunoprecipitation using cross-linked DX54.7 antibody.

To overcome challenges with antibody contamination in the ASPP2 IPs, an antibody cross-linking approach was chosen. By covalently attaching the antibody to beads, the amount of antibody in the sample could be reduced significantly. For the cross-linking, Dynabeads epoxy-resin was chosen, and the antibodies were linked to the beads according the manufacturers' protocol (for experimental procedures see chapter 10.4). A brief overview of the workflow is shown in Figure 15.

Using the above described workflow, ASPP2 was immunoprecipitated from U2OS WT and U2OS FIH KO cells alongside a control from U2OS ASPP2 KO cells. MS/MS data were collected in technical duplicates and samples from U2OS WT and U2OS ASPP2 KO cells were normalised to those from U2OS FIH KO using all identified proteins. In total, ~2100 proteins were identified. Further data analysis was performed as described previously.

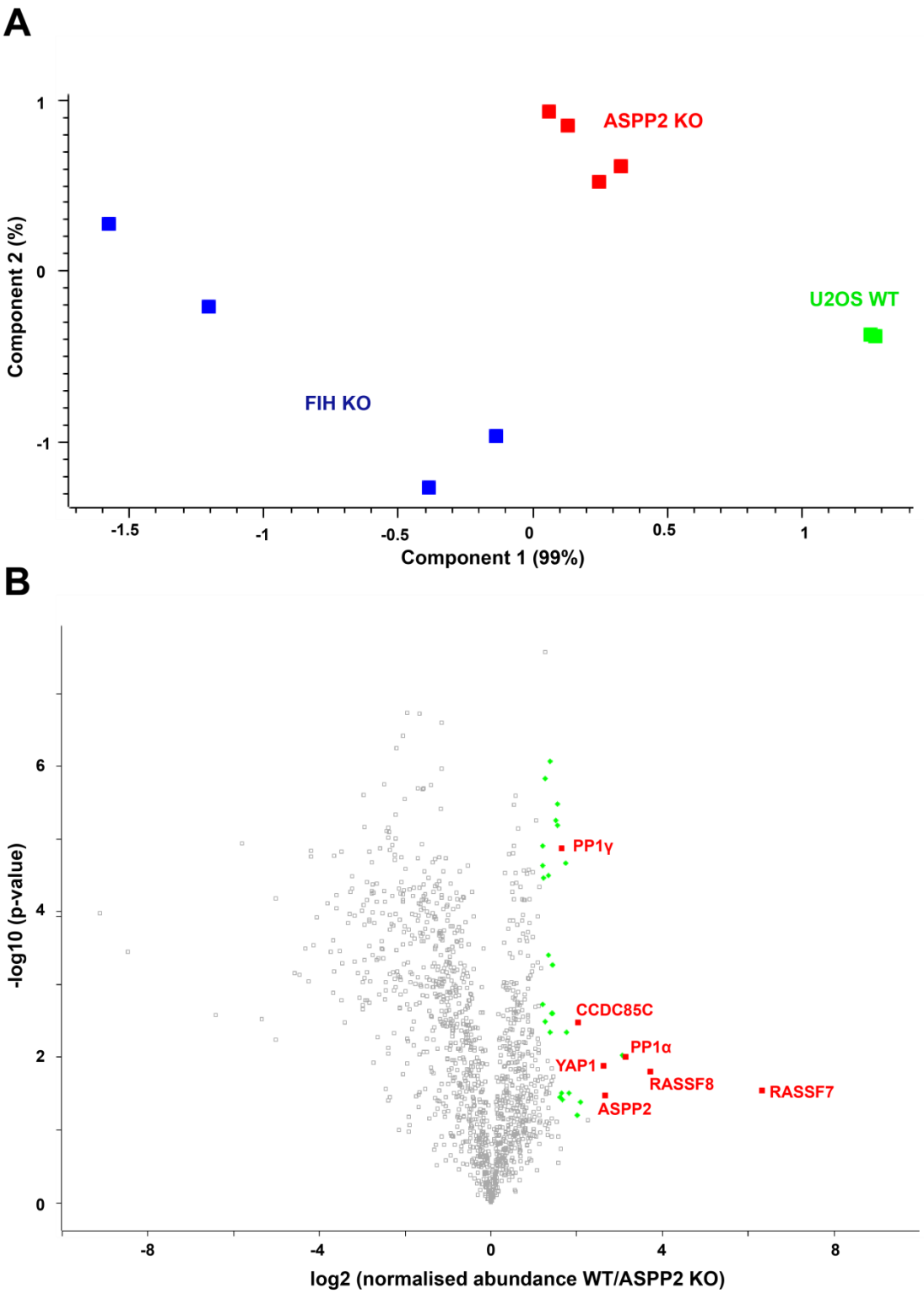
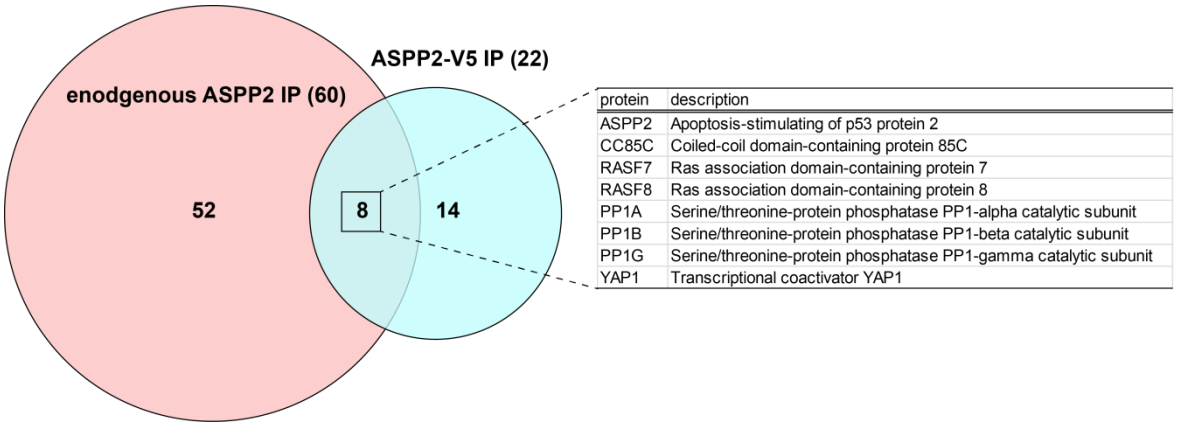


Figure 16 Endogenous ASPP2-IP from U2OS WT and U2OS ASPP2 KO cells. A PCA of the IPs of endogenous ASPP2 from U2OS WT, U2OS ASPP2 KO, and U2OS FIH KO cells. These data are from proteomic analysis of U2OS WT, ASPP2 KO and FIH KO samples. Proteins identified only by a single peptide were excluded from the analysis. PCA generated using Perseus MS software. **B** Volcano plot for the endogenous IP of ASPP2 from U2OS WT and U2OS ASPP2 KO cells. Each dot represents a single identified protein. Proteins significantly enriched are marked in green, and proteins of interest are marked in red.

In the principal component analysis, U2OS WT samples cluster closely and distinctively from the U2OS ASPP2 KO samples and the U2OS FIH KO samples (see Figure 16A). A volcano plot of the proteins identified from the ASPP2 IP from U2OS WT and U2OS ASPP2 KO cells illustrates the enrichment of ASPP2 in the samples from U2OS WT (Figure 16B). Several previously observed ASPP2 interacting proteins were found to be enriched in the IP of endogenous ASPP2 compared to the control, including RASSF7/8, YAP1 and PP1α (Table 1).

Proteins enriched >2 fold in the ASPP2 IP from U2OS WT over the control with a p-value <0.05 were marked as potential interactors. In total 52 proteins, excluding commonly found contaminants in IPs (*e.g.* ribosomal proteins and keratins) were identified (see supporting information for chapter 7, Appendix).

The list of all potential ASPP2 interactors from the co-IP of endogenous ASPP2 was compared to the potential interactors identified using overexpressed ASPP2 from the ASPP2-V5 IP (Figure 17). Seven proteins were identified in both samples as potential interactors, including the verified interactors PP1α/β/γ and YAP1 as well as the previously described RASSF7/8 and CCDC85C.



protein	description
ASPP2	Apoptosis-stimulating of p53 protein 2
CC85C	Coiled-coil domain-containing protein 85C
RASF7	Ras association domain-containing protein 7
RASF8	Ras association domain-containing protein 8
PP1A	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit
PP1B	Serine/threonine-protein phosphatase PP1-beta catalytic subunit
PP1G	Serine/threonine-protein phosphatase PP1-gamma catalytic subunit
YAP1	Transcriptional coactivator YAP1

Figure 17 Comparison of endogenous ASPP2 WT IP and ASPP2-V5 IP from U2OS cells. Venn diagram of proteins enriched in the endogenous ASPP2 IP over ASPP2 KO in U2OS WT cells and proteins enriched in the ASPP2-V5 IP enriched over the empty vector control in U2OS ASPP2 KO cells.

Following the verification of the endogenous ASPP IP it was further tested whether any significant differences between the interactomes of ASPP2 from U2OS and U2OS FIH KO cells could be observed (Figure 18). Overall, no specific protein could be identified who's binding to ASPP2 was potentially regulated by the hydroxylation status of ASPP2. This is in accordance with the previously described ASPP2-V5 IPs (see above).

Overall, the data presented shows no evidence for a regulatory role of FIH on ASPP2 interactions. However, a potential effect of FIH on ASPP2 interactions and functions cannot be excluded. ASPP2 has been shown to be able to shuttle between the nucleus and cytoplasm interacting with distinct sets of proteins[53]. Also, changes in interaction partners were observed at various stages of the cell cycle are possible[54]. At least for FIH, interactions are most likely transient and are thus challenging to detect *via* IPs.

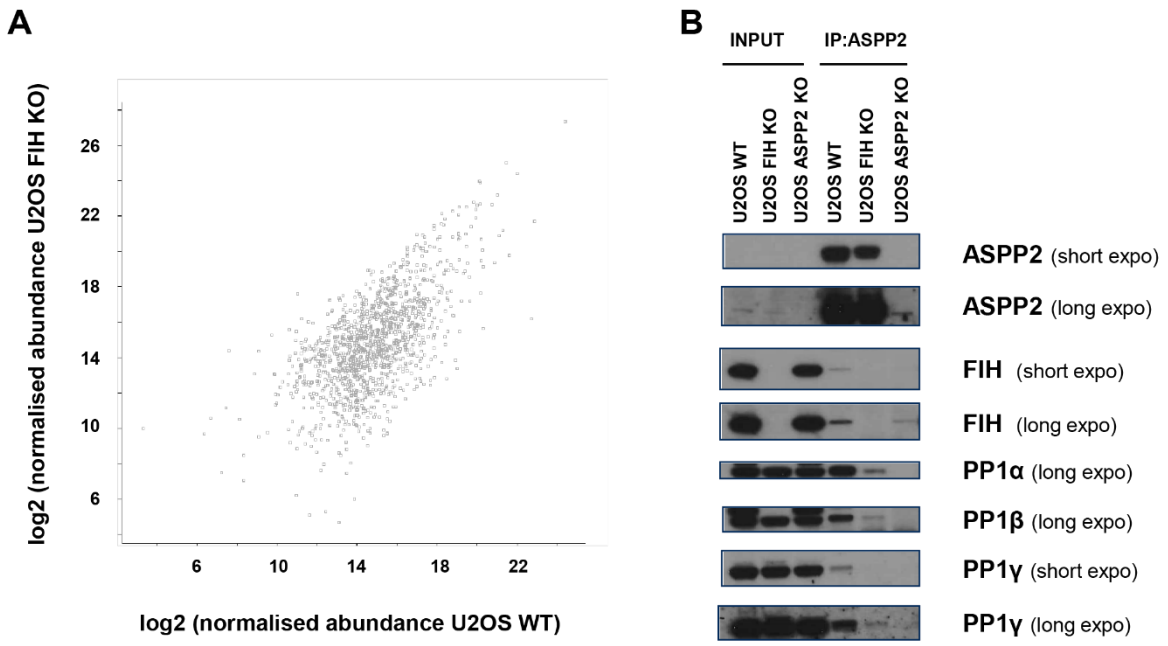


Figure 18 Comparison of the interactomes of ASPP2 WT from U2OS WT and U2OS FIH KO cells. A Normalised abundance of proteins co-immunoprecipitated with ASPP2 WT from U2OS WT and U2OS FIH KO cells. Each dot represents the average normalised abundance of identified proteins from two technical replicates. **B** Western blot analysis of proteins co-immunoprecipitated with ASPP2 from U2OS and U2OS FIH KO cells.

7.6 Conclusions and outlook

In summary, protocols for studying ASPP2 interactions using a label-free quantitative mass spectrometry based approach were established.

Distinct interactomes of ASPP2-V5 and ASPP2-V5 RAKA, a PP1-binding deficient variant, show different amounts of proteins co-immunoprecipitated with an anti-V5 IP. In addition to the PP1s, these include reduced amounts of ZBTB2 and CCDC85B in the ASPP2-V5 RAKA variant compared to the ASPP2-V5 WT. The proteins KIF1C and PRRC2A were exclusively found in the ASPP2-V5 RAKA IP. In mice, introduction of an ASPP2 PP1 binding deficient variant is at the embryonic stage lethal. The molecular mechanisms by which the ASPP2-PP1 complex is essential for development remain to be elucidated. Follow-up studies on ASPP2 and ASPP2-PP1 complex interactions using interactome studies, as presented here, will help to elucidate the role of ASPP2.

On a peptide level, increased phosphorylation of ASPP2-V5 RAKA at serine-737, serine-714 and serine-414 compared to ASPP2-V5 WT was observed. These are likely targets of an auto-dephosphorylation mechanism of the ASPP2-PP1 dephosphorylation complex. To date, little is known about the regulation of the ASPP2-PP1 complex, and potential auto-regulation by dephosphorylation of the complex is of interest.

The potential effect of FIH-mediated ASPP2 hydroxylation on ASPP2 interactomes was investigated using overexpressed ASPP2-V5 WT in U2OS ASPP2 KO and U2OS ASPP2 KO/FIH KO cells as well as ASPP2-V5 alanine variants and endogenous ASPP2 from U2OS WT and U2OS FIH KO cells. No direct link between ASPP2 hydroxylation status and changes in the interactome was observed. More importantly, the alanine variants of ASPP2 had a distinct influence on the ASPP2 interactome.

A robust and reproducible methodology for the interactome studies of endogenous ASPP2 was developed and the results were comparable with those from overexpressed ASPP2-V5. These methods give unique opportunities to investigate the interactome of ASPP2 under multiple conditions, and could be expanded to subcellular interactome studies, further elucidating the cellular role of ASPP2, particular with respect to interactions with PP1.

7.7 References

1 Cockman, M. E., Webb, J. D. & Ratcliffe, P. J. FIH-dependent asparaginyl hydroxylation of ankyrin repeat domain-containing proteins. *Ann N Y Acad Sci* **1177**, 9-18, doi:10.1111/j.1749-6632.2009.05042.x (2009).

2 Schofield, C. J. & Ratcliffe, P. J. Oxygen sensing by HIF hydroxylases. *Nat Rev Mol Cell Biol* **5**, 343-354, doi:10.1038/nrm1366 (2004).

3 Kaelin, W. G. & Ratcliffe, P. J. Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. *Mol Cell* **30**, 393-402, doi:10.1016/j.molcel.2008.04.009 (2008).

4 Zhang, J. *et al.* PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol Cell Proteomics* **11**, M111.010587, doi:10.1074/mcp.M111.010587 (2012).

5 Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740, doi:10.1038/nmeth.3901 (2016).

6 Zhang, N. *et al.* The asparaginyl hydroxylase factor inhibiting HIF-1alpha is an essential regulator of metabolism. *Cell Metab* **11**, 364-378, doi:10.1016/j.cmet.2010.03.001 (2010).

7 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, doi:10.1126/science.1068592 (2002).

8 Lando, D. *et al.* FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. *Genes Dev* **16**, 1466-1471, doi:10.1101/gad.991402 (2002).

9 Vives, V. *et al.* ASPP2 is a haploinsufficient tumor suppressor that cooperates with p53 to suppress tumor growth. *Genes Dev* **20**, 1262-1267, doi:10.1101/gad.374006 (2006).

10 Bergamaschi, D. *et al.* ASPP1 and ASPP2: common activators of p53 family members. *Mol Cell Biol* **24**, 1341-1350 (2004).

11 Samuels-Lev, Y. *et al.* ASPP proteins specifically stimulate the apoptotic function of p53. *Mol Cell* **8**, 781-794 (2001).

12 Jensen, L. J. *et al.* STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37**, D412-416, doi:10.1093/nar/gkn760 (2009).

13 Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**, D447-452, doi:10.1093/nar/gku1003 (2015).

14 Zheng, X. *et al.* Interaction with factor inhibiting HIF-1 defines an additional mode of cross-coupling between the Notch and hypoxia signaling pathways. *Proc Natl Acad Sci U S A* **105**, 3368-3373, doi:10.1073/pnas.0711591105 (2008).

15 Scholz, C. C. *et al.* FIH Regulates Cellular Metabolism through Hydroxylation of the Deubiquitinase OTUB1. *PLoS Biol* **14**, e1002347, doi:10.1371/journal.pbio.1002347 (2016).

16 Llanos, S. *et al.* Inhibitory member of the apoptosis-stimulating proteins of the p53 family (iASPP) interacts with protein phosphatase 1 via a noncanonical binding motif. *J Biol Chem* **286**, 43039-43044, doi:10.1074/jbc.M111.270751 (2011).

17 Royer, C. *et al.* ASPP2 links the apical lateral polarity complex to the regulation of YAP activity in epithelial cells. *PLoS One* **9**, e111384, doi:10.1371/journal.pone.0111384 (2014).

18 The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158-D169, doi:10.1093/nar/gkw1099 (2017).

19 Espanel, X. & Sudol, M. Yes-associated protein and p53-binding protein-2 interact through their WW and SH3 domains. *J Biol Chem* **276**, 14514-14523, doi:10.1074/jbc.M008568200 (2001).

20 Helps, N. R., Barker, H. M., Elledge, S. J. & Cohen, P. T. Protein phosphatase 1 interacts with p53BP2, a protein which binds to the tumour suppressor p53. *FEBS Lett* **377**, 295-300, doi:10.1016/0014-5793(95)01347-4 (1995).

21 Cong, W. *et al.* ASPP2 regulates epithelial cell polarity through the PAR complex. *Curr Biol* **20**, 1408-1414, doi:10.1016/j.cub.2010.06.024 (2010).

22 Sottocornola, R. *et al.* ASPP2 binds Par-3 and controls the polarity and proliferation of neural progenitors during CNS development. *Dev Cell* **19**, 126-137, doi:10.1016/j.devcel.2010.06.003 (2010).

23 Hauri, S. *et al.* Interaction proteome of human Hippo signaling: modular control of the co-activator YAP1. *Mol Syst Biol* **9**, 713, doi:10.1002/msb.201304750 (2013).

24 Balczon, R., Bao, L. & Zimmer, W. E. PCM-1, A 228-kD centrosome autoantigen with a distinct cell cycle distribution. *J Cell Biol* **124**, 783-793 (1994).

25 Dammermann, A. & Merdes, A. Assembly of centrosomal proteins and microtubule organization depends on PCM-1. *J Cell Biol* **159**, 255-266, doi:10.1083/jcb.200204023 (2002).

26 Lisanti, M. P., Tang, Z. L. & Sargiacomo, M. Caveolin forms a hetero-oligomeric protein complex that interacts with an apical GPI-linked protein: implications for the biogenesis of caveolae. *J Cell Biol* **123**, 595-604 (1993).

27 Glenney, J. R. & Soppet, D. Sequence and expression of caveolin, a protein component of caveolae plasma membrane domains phosphorylated on tyrosine in Rous sarcoma virus-transformed fibroblasts. *Proc Natl Acad Sci U S A* **89**, 10517-10521 (1992).

28 Zaessinger, S., Zhou, Y., Bray, S. J., Tapon, N. & Djiane, A. Drosophila MAGI interacts with RASSF8 to regulate E-Cadherin-based adherens junctions in the developing eye. *Development* **142**, 1102-1112, doi:10.1242/dev.116277 (2015).

29 Langton, P. F. *et al.* The dASPP-dRASSF8 complex regulates cell-cell adhesion during Drosophila retinal morphogenesis. *Curr Biol* **19**, 1969-1978, doi:10.1016/j.cub.2009.10.027 (2009).

30 Patel, V. J. *et al.* A comparison of labeling and label-free mass spectrometry-based proteomics approaches. *J Proteome Res* **8**, 3752-3759, doi:10.1021/pr900080y (2009).

31 Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* **389**, 1017-1031, doi:10.1007/s00216-007-1486-6 (2007).

32 Mann, M. Comparative analysis to guide quality improvements in proteomics. *Nat Methods* **6**, 717-719 (2009).

33 Huttlin, E. L. *et al.* The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* **162**, 425-440, doi:10.1016/j.cell.2015.06.043 (2015).

34 Huttlin, E. L. *et al.* Architecture of the human interactome defines protein communities and disease networks. *Nature* **545**, 505-509, doi:10.1038/nature22366 (2017).

35 Bian, Y. *et al.* An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. *J Proteomics* **96**, 253-262, doi:10.1016/j.jprot.2013.11.014 (2014).

36 Hornbeck, P. V. *et al.* PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res* **40**, D261-270, doi:10.1093/nar/gkr1122 (2012).

37 Dorner, C. *et al.* Characterization of KIF1C, a new kinesin-like protein involved in vesicle transport from the Golgi apparatus to the endoplasmic reticulum. *J Biol Chem* **273**, 20267-20275 (1998).

38 Lehner, B. *et al.* Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* **83**, 153-167 (2004).

39 Rigbolt, K. T. *et al.* System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal* **4**, rs3, doi:10.1126/scisignal.2001570 (2011).

40 Olsen, J. V. *et al.* Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635-648, doi:10.1016/j.cell.2006.09.026 (2006).

41 Dephoure, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc Natl Acad Sci U S A* **105**, 10762-10767, doi:10.1073/pnas.0805139105 (2008).
42 Mayya, V. *et al.* Quantitative phosphoproteomic analysis of T cell receptor signaling reveals system-wide modulation of protein-protein interactions. *Sci Signal* **2**, ra46, doi:10.1126/scisignal.2000007 (2009).
43 Beausoleil, S. A., Villén, J., Gerber, S. A., Rush, J. & Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **24**, 1285-1292, doi:10.1038/nbt1240 (2006).
44 Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal* **3**, ra3, doi:10.1126/scisignal.2000475 (2010).
45 Zhou, H. *et al.* Toward a comprehensive characterization of a human cancer cell phosphoproteome. *J Proteome Res* **12**, 260-271, doi:10.1021/pr300630k (2013).
46 Zahedi, R. P. *et al.* Phosphoproteome of resting human platelets. *J Proteome Res* **7**, 526-534, doi:10.1021/pr0704130 (2008).
47 Peti, W., Nairn, A. C. & Page, R. Structural basis for protein phosphatase 1 regulation and specificity. *FEBS J* **280**, 596-611, doi:10.1111/j.1742-4658.2012.08509.x (2013).
48 Nešić, D., Buti, L., Lu, X. & Stebbins, C. E. Structure of the Helicobacter pylori CagA oncoprotein bound to the human tumor suppressor ASPP2. *Proc Natl Acad Sci U S A* **111**, 1562-1567, doi:10.1073/pnas.1320631111 (2014).
49 Dohadwala, M. *et al.* Phosphorylation and inactivation of protein phosphatase 1 by cyclin-dependent kinases. *Proc Natl Acad Sci U S A* **91**, 6408-6412 (1994).
50 Kwon, Y. G., Lee, S. Y., Choi, Y., Greengard, P. & Nairn, A. C. Cell cycle-dependent phosphorylation of mammalian protein phosphatase 1 by cdc2 kinase. *Proc Natl Acad Sci U S A* **94**, 2168-2173 (1997).
51 Liu, C. W. *et al.* Inhibitory phosphorylation of PP1alpha catalytic subunit during the G(1)/S transition. *J Biol Chem* **274**, 29470-29475 (1999).
52 Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567, doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2 (1999).
53 Tordella, L. *et al.* ASPP2 suppresses squamous cell carcinoma via RelA/p65-mediated repression of p63. *Proc Natl Acad Sci U S A* **110**, 17969-17974, doi:10.1073/pnas.1309362110 (2013).
54 Zhang, Y. *et al.* The tumor suppressor proteins ASPP1 and ASPP2 interact with C-Nap1 and regulate centrosome linker reassembly. *Biochem Biophys Res Commun*, doi:10.1016/j.bbrc.2015.01.136 (2015).

8 XFEL structures of 2OG oxygenases

This chapter presents work leading to the first room-temperature crystal structures of 2OG oxygenases in an active enzyme-substrate complex obtained using X-ray free-electron laser methods. The advantages of XFEL over synchrotron radiation, sample requirements and preparations for XFEL measurements and limitations of the experiments are described and discussed using the examples of IPNS, AlkB, FIH, BcII and VIM-2. Optimisation of micro-crystallisation conditions for XFEL measurements in hanging-drop and batch methods are presented. Further, room- temperature crystal structures of FIH and AlkB in the resting and the oxidised state are presented and compared to the corresponding cryo-cooled crystal structures.

8.1 Introduction

8.1.1 Introduction: time-resolved structural biology

Structural characterization of enzymes can provide valuable insights into their mechanisms and functions. With advances in technology, it has become possible to use time-resolved crystallography to provide information on intermediates in enzyme catalysis[1-9]. Most enzymatic reactions are relatively rapid (>1 ms) and involve unstable intermediates, making it challenging to obtain structural information on them[10-12].

The predominant method for structural determination of protein structures is low temperature X-ray crystallography (also referred to as cryo-crystallography)[13]. Data collection from “cryo-cooled” samples has significant advantages in terms of reducing radiation damage to the crystal[14]. Cryo-cooling may, however, perturb the structure(s) of a protein[15,16]. Solution based structural methods, *e.g.* Nuclear Magnetic Resonance measurements (NMR) or Small Angle X-ray Scattering (SAXS), can capture solution states but in the case of NMR, size limitation and measuring times in the range of minutes to days do not allow the determination of specific dynamic states[17,18]. SAXS measurements can determine dynamic rearrangements of proteins and enzymes but due to resolution limitations only large conformational changes can be observed[19-21]. Stopped flow and rapid quench methods allow the studies of short lived intermediates, but do not provide global structural information. In recent years the X-ray free-electron laser (XFEL) has emerged as a novel technique of potential use in structural enzymology[2,22-28].

8.1.2 X-ray electron-free laser in macromolecular serial crystallography

XFELs provide X-ray pulses with similar a wavelength range compared to synchrotron radiation. The peak brilliance of the XFEL beam is, compared to current 3rd generation synchrotrons, ~10 orders of magnitude higher, which allows the use of very short pulses and enables the measurement of thousands of very small crystals (micro-nanometer range)[13,29]. Current XFEL beams achieve pulses in the range of 10-100 fs, which enables measurements three to four orders of magnitude faster than current 3rd generation synchrotrons[13,29]. With only femto-second pulses, images can be collected from protein crystals before radiation damage occurs. Therefore, XFEL measurements yield, at least in theory, radiation damage free protein structures[30,31]. This is especially of interest for samples sensitive to radiation damage and for collecting data at room-temperature. To get a complete dataset, data from thousands to tens of thousands of crystals need to be acquired and merged, this approach is referred to as serial femtosecond crystallography (SFX)[32]. The short XFEL pulses also allow for time-resolved studies of enzymatic reactions in the crystal state, referred to as time-resolved SFX (trSFX). Taking advantage of short-pulsed diffraction images has the potential for capturing specific “intermediate” states of enzymes during catalysis (see above). This approach is extremely powerful in combination with complementary measurements like spectroscopy[22]. However, the sample requirements for XFEL measurements can differ significantly from traditional X-ray crystallography.

8.1.3 Sample requirements for XFEL experiments

Due to the extreme brightness of a focused XFEL beam, any material interacting with the beam is destroyed, including protein crystals[13,29]. To obtain a complete dataset with random orientated crystals, which is dependent on crystal symmetry, unit cell size and resolution, data from several thousand to tens of thousands, in some cases millions, of crystals are needed[32]. This differs from traditional X-ray crystallography, where data from a single crystal are typically collected.

XFEL measurements do not have, *per se*, an upper limit on crystal size (further discussed below); however, sample availability and the crystal delivery system as well as diffusion times of substrates for trSFX experiments set limits on the size of crystals that can be measured (see below). The small focus of the XFEL beam (*e.g.* ~1 µm at the MFX beamline at the Linac Coherent Light Source, LCLS), allows the measurement of micro- and nano-crystals[13,25,29]. Samples are routinely prepared as crystal slurries containing tens to hundreds of thousands of micro crystals per µL and even higher concentrations for nano-crystals[33].

8.1.4 Sample delivery systems for XFEL

To ensure efficient data collection, continuous sample delivery is needed. Several sample delivery systems have been developed for XFEL. Most commonly liquid jet systems; these have a very low upper crystal size limit due to the small capillaries used and have a high sample consumption due to continuous flow but allow for fast mixing experiments[23,34,35]. Lipid cubic phases (LCPs) are often used for delivery of membrane protein crystals. LCP systems are a highly viscous medium, which is not always compatible with soluble proteins and makes mixing experiments challenging but have a very low sample consumption[36,37]. Fixed target systems have successfully been used for XFEL[38-41], these have low sample consumption, but relatively stringent crystal size limits. They have also not been demonstrated for multidomain protein crystal data collection or mixing experiments[42].

For measurements at the LCLS, an Acoustic Droplet Ejection (ADE) sample delivery system has been developed[42]. ADE allows a discontinuous sample delivery, which reduces sample consumption while still offering the path length to accommodate simultaneous X-ray emission spectroscopy and X-ray diffraction (Figure 2). Additionally, while drops are delivered on a Kapton belt, crystal manipulation *via e.g.* laser irradiation or oxygenation is possible. The size limit for crystals is determined by the size of the capillary transporting the crystal slurry to the dispenser. A larger diameter allows a wider size range of crystals but increases the dead volume. The experiments at the LCLS described below were performed with a capillary diameter of 300 µm with a dead volume of 200 µL allowing crystals sizes of up to ~200 µm.

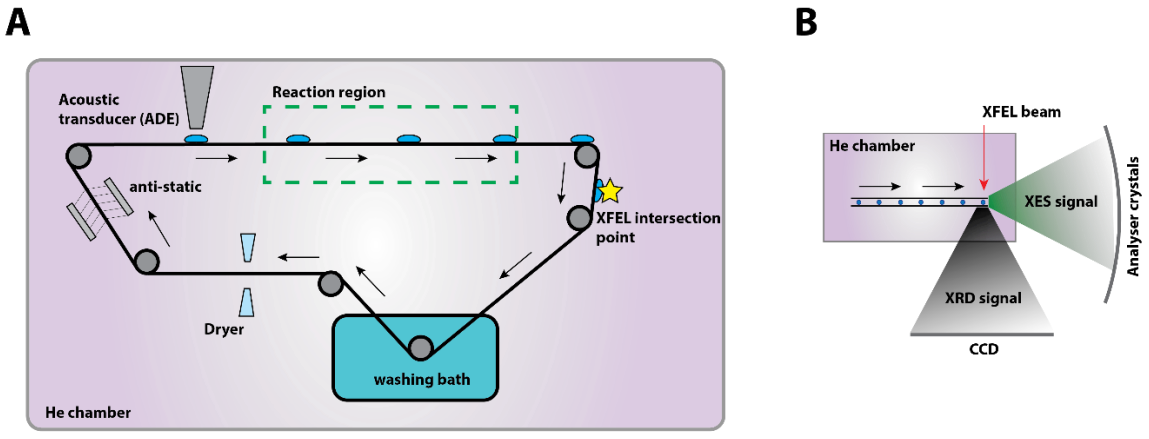


Figure 1 Experimental setup for acoustic droplet ejection sample delivery for XFEL measurements at the LCLS. A Schematic overview of sample delivery mechanism using ADE and a kapton belt to deliver droplets at high rate to the XFEL. **B** Schematic for data collection geometry for simultaneous XRD and XES measurements. Images modified from Fuller and Gul et al[42].

8.1.5 XFEL studies on 2OG oxygenases

Studying the reactions of 2OG oxygenases is a long lasting and ongoing research interest. Using the ADE sample delivery system, 2OG oxygenases were planned to be screened at the LCLS to investigate their potential for time-resolved studies at an XFEL. Follow-up experiments could include the initiation of the reaction of 2OG oxygenases in micro-crystal slurries at defined time points, by diffusion of molecular oxygen or rapid mixing with oxygenated buffer, prior to measurements with the XFEL beam with the aim to create a molecular movie of the reaction of a 2OG oxygenases or related enzyme.

8.2 Sample selection and initial crystallisation optimisation

8.2.1 Sample selection

To follow enzyme reactions in the crystalline state, the proteins have to crystallise readily and diffract to a high resolution with substrate and co-factors in an active complex. Furthermore, catalytic activity of the enzyme must only minimally disrupt the crystal lattice and the reaction must be, at least substantially, uniform across the crystal. In the cases of 2OG oxygenases and related enzymes, reaction can be initiated by oxygen. To avoid substrate and 2OG turnover during crystallisation, crystals must be grown under anaerobic conditions. Only proteins crystallising at ambient temperature were therefore considered. Three 2OG oxygenases and related enzymes that fulfil the above stated criteria are human FIH, the DNA repair enzyme AlkB from *Escherichia coli (E. coli)*, and isopenicillin N synthase (IPNS) from *Aspergillus nidulans (A. nidulans)*. The reactions catalysed by the enzymes are shown Figure 2.

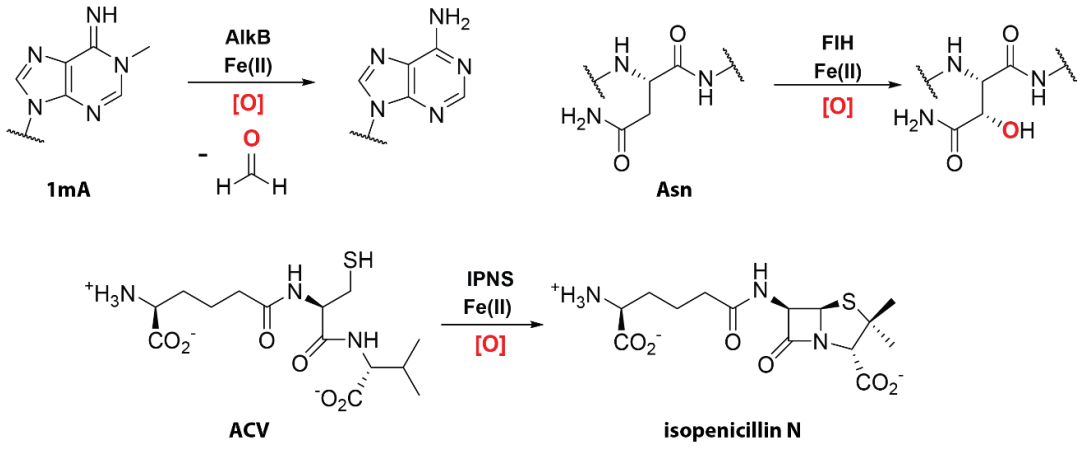


Figure 2 Overview of reactions catalysed by AlkB, FIH and IPNS. 1mA = 1-methyadenine, ACV = delta-(L-alpha-aminoadipoyl)-L-cysteinyl-D-valine.

All three proteins robustly crystallise with substrate and co-factors and diffract to high resolution at synchrotron radiation sources. For IPNS and FIH, detailed spectroscopic analysis of reaction intermediates in solution are available[43,44].

8.2.2 Initial crystallisation trials

8.2.2.1 Initial IPNS and FIH crystallisation trials

IPNS protein was produced by Dr Patrick Rabe (University of Oxford). Delta-(L-alpha-aminoadipoyl)-L-cysteinyl-D-valine (ACV) was synthesized by Dr Patrick Rabe and Mr Jos Kamps (University of Oxford). IPNS crystallisation trials were performed in collaboration with Ms. Anka Lucic (University of Oxford).

In the first set of experiments, IPNS was crystallised under anaerobic conditions in the presence of its substrate ACV (see Figure 2). Crystals were grown according to the published procedures[45]. Two crystal forms have been described, an orthorhombic crystal form crystallising in space group $P2_12_12_1$ diffracting up to ~1.3 Å and a trigonal crystal form crystallising in space group $P3_121$ reported to diffract up to ~2.1 Å using synchrotron radiation sources. IPNS crystals appeared overnight and continued to grow for several days at room temperature. From the two previously observed crystal forms, only the trigonal crystal form was observed under tested conditions (Figure 3). The crystals were harvested under anaerobic conditions and cryo-cooled in liquid nitrogen (Dr Michael McDonough, University of Oxford). Data were collected and the structure solved using molecular replacement (search model PDB code 1QJE[46]). A crystal structure surprisingly revealed that the substrate had been converted to isopenicillin N and no ACV was observed at the active site of IPNS (Figure 3). Iron was found to be present only in low occupancy at the active site, note that the oxidation state of iron could not be determine. Due to the low iron concentration (100 µM Fe(II) for 1.2 mM IPNS), turn-over of ACV to isopenicillin N in the crystal seems unlikely. It appears more likely that ACV was turned over to isopenicillin N by a catalytically active sub-population of IPNS in solution. The isopenicillin N could then bind to IPNS which then crystallised. Solution measurements are needed to confirm this hypothesis. Additionally turn-over of ACV to isopenicillin N strongly suggests that the anaerobic glove box was compromised at some point during the crystallisation trials.

Following from the IPNS.isopenicillin N structure, it was tested if iron(II)/(III) and isopenicillin N could bind to IPNS at the same time. Crystals of the IPNS.isopenicillin N complex were soaked with iron(II) ammonium sulphate solution for various time points under aerobic conditions, data were collected, and the structure solved as previously described (see Figure 3). The iron(II)/(III) occupancy was increased but the substrate could not be correctly fitted into the density in the active site, suggesting that the isopenicillin N may have degraded.

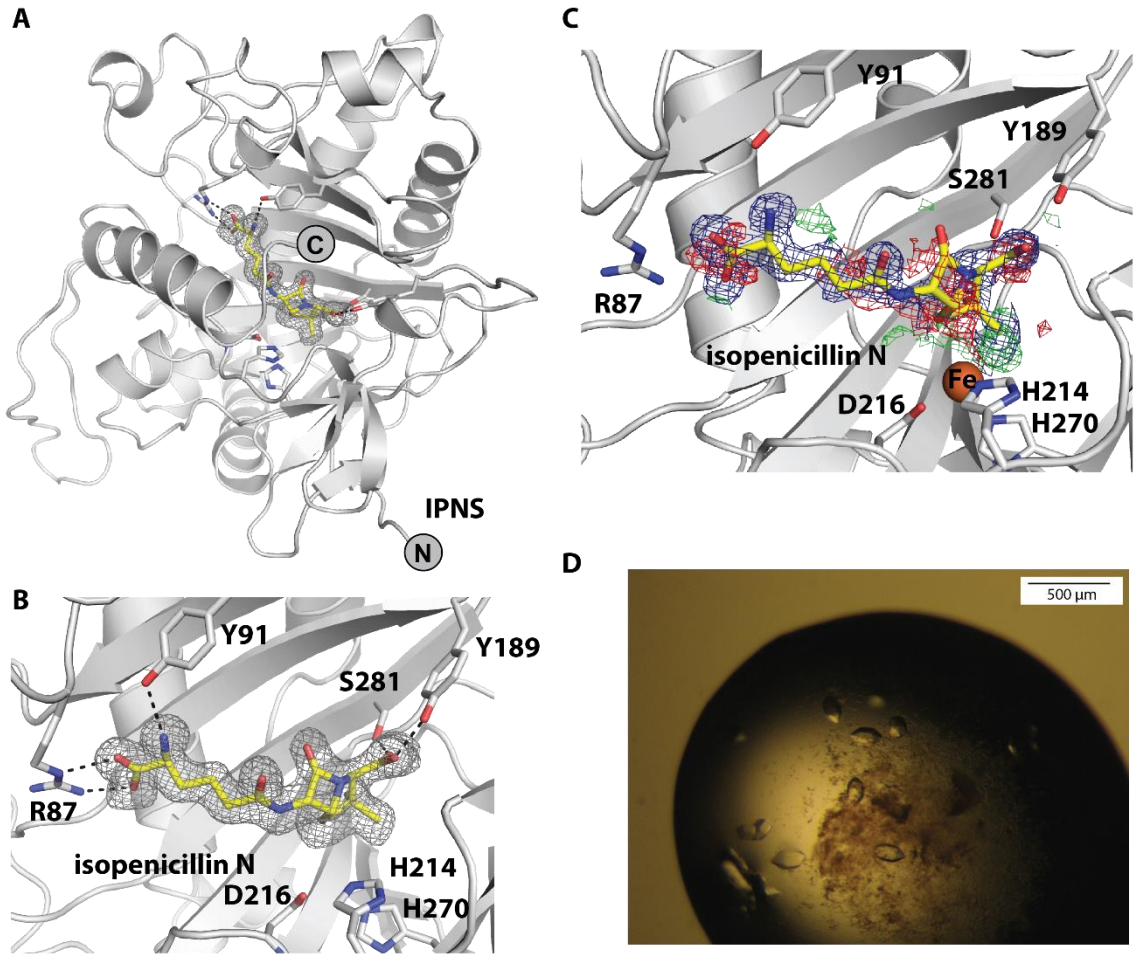


Figure 3 Views from IPNS crystal structures. A,B Views from a crystal structure of IPNS in complex with isopenicillin N; mFo-Fc OMIT maps are contoured to 3σ around the ligand and are shown in grey. **C** View from a crystal structure of IPNS in complex with isopenicillin N after soaking with iron(II) sulphate after one round of refinement with modelled isopenicillin N bound to the active site; 2mFo-DFc maps are contoured to 1σ shown in blue, mFo-Fc contoured to 3σ in green and -3σ in red. **D** Pictures of IPNS trigonal crystals form.

Table 1 Data collection and refinement statistics of IPNS complexes.

Datasets	IPNS.Isopenicillin N	IPNS.Isopenicillin N Fe(II) soaked
Data Collection		
Beamline (Wavelength, Å)	DLS I04 (0.9795)	DLS I04 (0.9795)
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia2-DIALS	Xia2-DIALS
Space group	$P3_121$	$P3_121$
Cell dimensions		
a,*b*,*c* (Å)	101.2, 101.2, 113.3	101.4, 101.4, 115.9
α,β,γ(°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
No. of molecules/ASU	1	1
No. reflections	93925 (4655)*	92916 (4540)*
Resolution (Å)	69.34-1.57 (1.60-1.57)*	87.85-1.59 (1.62-1.59)*
R_{merge}	0.084 (1.748)*	0.068 (1.829)*
I/σI	10.6 (1.1)*	14.4 (1.2)*
Completeness (%)	100 (99.8)*	100 (99.6)*
Multiplicity	9.5 (9.1)*	9.8 (9.1)*
Wilson B value ($Å^2$)	27.5	23.5

* Highest resolution shell in parentheses.

8.2.2.2 Initial crystallisation trials of AlkB

AlkB was produced according to previously reported protocols[47]. Following the IPNS crystallisation trials suggesting that anaerobic glove box had been compromised, initial crystallisation trials with AlkB were performed under aerobic conditions with the inhibitor SS81(kindly provided by Ms. Shifali Shishodia, University of Oxford), and in the presence of manganese(II) chloride to substitute for iron(II) to avoid metal oxidation. Crystals appeared in PEG Rx broad screen (Hampton Research), well D3, and were further optimised (see Figure 4). Crystals from the broad screen and optimised conditions were harvested, cryo-protected with glycerol and cryo-cooled in liquid nitrogen. However, the glycerol appeared to damage the crystals. Crystals from the broad screens crystallised in space group *P*1 with similar unit cell dimensions compared to previously observed AlkB inhibitor structures[48] (Table 2). The structures were solved using molecular replacement (search model PDB code 4JHT[48]). Surprisingly, crystals from optimised conditions crystallised in similar morphology to crystals in the broad screen (see Figure 4), but in space group *P*1 with different unit cell dimensions (Table 2) and two molecules per asymmetric unit. Crystal structures for both crystal forms were solved and showed a conserved binding mode of the inhibitor (Figure 4).

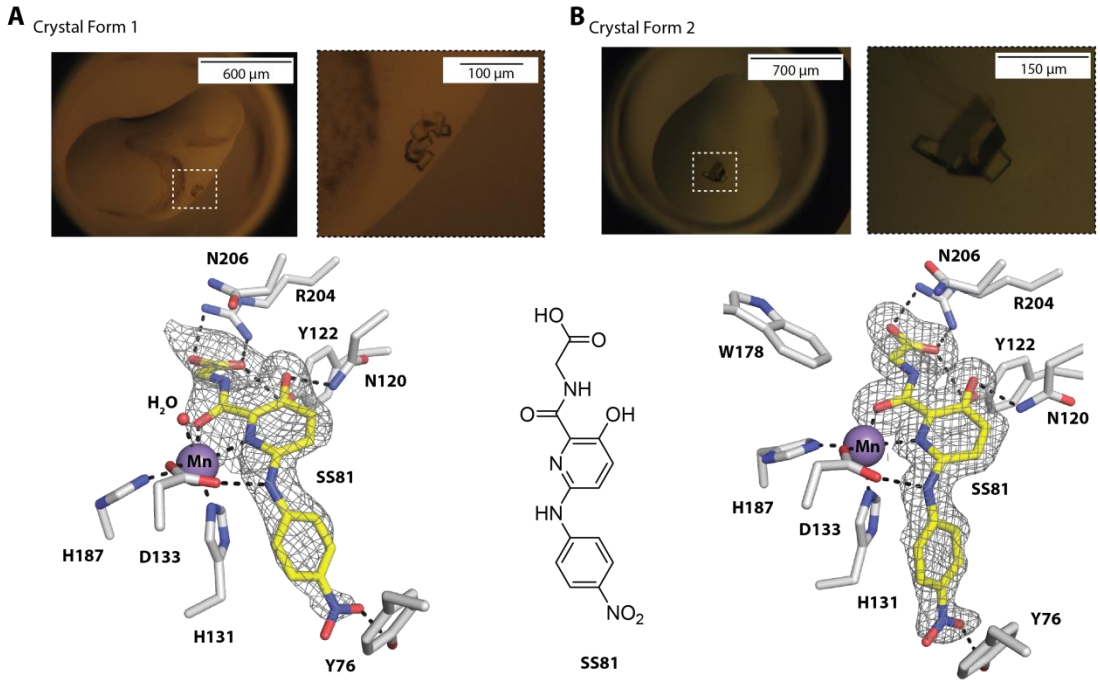


Figure 4 Crystallization of AlkB in complex with Mn(II) and SS81. A Pictures from the AlkB.Mn(II).SS81 complex in crystal form 1 in PEG Rx well D3. The crystals have a cubic like morphology and grew in clusters. View from a crystal structure of the AlkB.Mn(II).SS81 complex illustrates the binding mode of the inhibitor; the mFo-Fc OMIT map are contoured to 3σ around the inhibitor and are shown in grey. **B** Pictures of the AlkB.Mn(II).SS81 complex in crystal form 2. View from an AlkB.Mn(II).SS81 crystal structure illustrates the binding mode of the inhibitor; mFo-Fc OMIT maps are contoured to 3σ around the inhibitor and are shown in grey.

Following purging and regeneration of the glove box to ensure an anoxic environment, AlkB was crystallised with Fe(II), 2OG and a modified DNA trimer substrate T-1mA-T (with the assistance of Ms. Anka Lucic, University of Oxford). Crystallisation conditions were designed according to the reported conditions[47]. Crystals of AlkB appeared overnight under anoxic conditions at room temperature and continued to grow for several days. Surprisingly, the crystals did not grow in a tetragonal space group as it would have been expected from the reported AlkB substrate complexes in the tested conditions[47], but rather a triclinic space group analogous to the previously described AlkB inhibitor structures in space group *P*1 (Figure 10A). AlkB complex crystals were harvested under anoxic conditions (Dr Michael McDonough, University of Oxford) and data were collected under cryo-conditions. To ensure completeness in the high-resolution shells, 3 sweeps of 180° at χ=0°, 20°, 45° were collected. Data were integrated using XDS[49], merged using POINTLESS[50] and scaled with SCALA[51]. Data collection and processing confirmed space group *P*1 (Table 2). The crystal structure was solved by molecular replacement as previously described. Solving the crystal structure revealed bound co-factors and substrate to the active site (see Figure 5). Notably, this structure is currently the highest resolution reported for a crystal structure of a 2OG oxygenase in an active substrate complex.

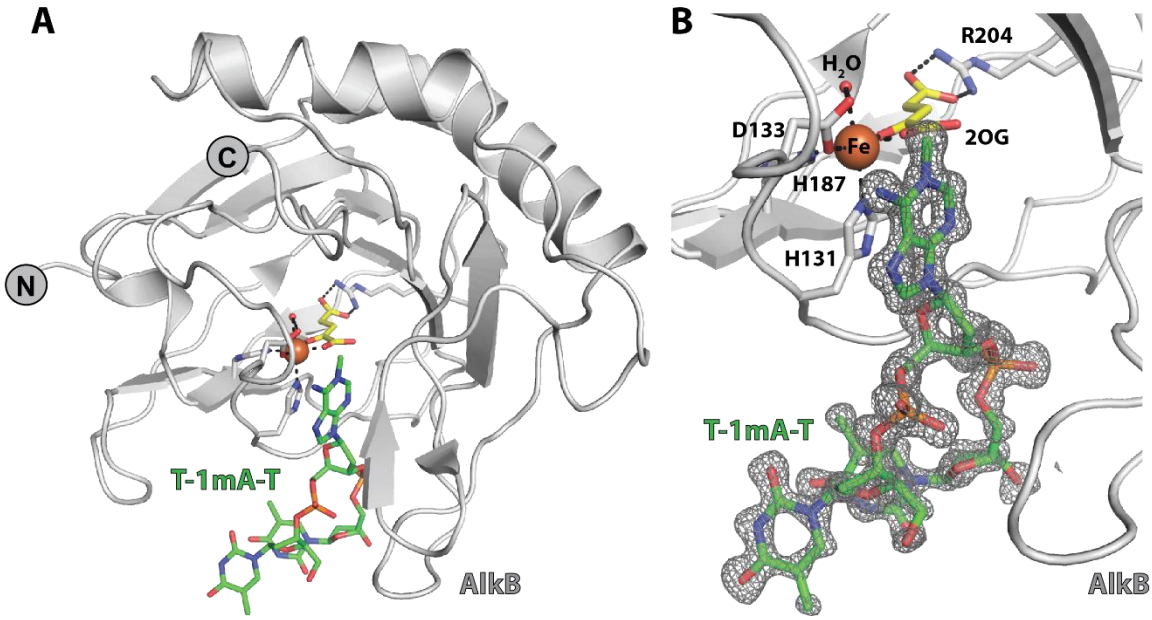


Figure 5 Views from a crystal structure of AlkB in complex with T-1mA-T DNA fragment. **A** View from a crystal structure of the AlkB.Fe(II).2OG.T-1mA-T complex at 1.10 Å illustrates the binding mode of AlkB to the DNA in the crystal state. **B** View from the active site of the AlkB substrate complex; mFo-Fc OMIT maps are contoured to 3σ around the T-1mA-T DNA substrate and shown in grey.

Table 2 Data collection and refinement statistics of AlkB complexes.

Datasets	AlkB.Fe(II).2OG T-mA-T	**AlkB**.Mn(II).SS81 Crystal form 1**	**AlkB**.Mn(II).SS81 Crystal form 2**
Data Collection			
Beamline (Wavelength, Å)	DLS I04 (0.9795)	DLS I04 (0.9795)	DLS I04 (0.9795)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	XDS, POINTLESS, SCALA	XDS, POINTLESS, SCALA	XDS, POINTLESS, SCALA
Space group	*P*1	*P*1	*P*1
Cell dimensions			
a,*b*,*c* (Å)	35.3, 28.5, 40.1	36.4, 38.6, 40.2	36.2, 45.8, 57.7
α,β,γ(°)	78.5, 78.6, 67.6	77.5, 75.9, 66.9	83.1, 78.0, 88.8
No. of molecules/ASU	1	1	2
No. reflections	76422 (42679)*	15183 (2207)*	55892 (7337)*
Resolution (Å)	32.32-1.10 (1.16-1.10)	38.62-1.90 (2.00-1.90)*	56.02-1.50 (1.58-1.50)*
R_{merge}	0.043 (0.200)*	0.077 (1.097)*	0.062 (0.769)*
I/σI	22.2 (5.7)*	9.0 (1.2)*	13.8 (2.0)*
Completeness (%)	99.1 (97.6)*	99.8 (99.9)*	96.9 (86.9)*
Multiplicity	4.6 (3.9)*	5.0 (5.0)*	4.8 (4.7)*
Wilson B value (Å^2)	9.95	34.78	19.26
Refinement	PHENIX	PHENIX	PHENIX
R_{work}/R_{free}	0.1157/0.1394	---	---
No. atoms			
-Enzyme (A/B)	1685	---	---
-Substrate	79	---	---
-Ligand	2OG (10)	---	---
-Water	394	---	---
Avg. B-factors			
Enzyme	13.9	---	---
Substrate	19.5	---	---
Ligand	12.4	---	---
Water	28.3	---	---
R.m.s deviations			
-Bond lengths (Å)	0.010	---	---
-Bond angles (°)	1.274	---	---

* Highest resolution shell in parentheses.

** refinement ongoing.

†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5% of reflections excluded from the refinement.

8.3 FIH sample preparation and data collection

8.3.1 FIH sample preparation and data collection at the LCLS

8.3.1.1 FIH micro-crystal optimisation under anaerobic conditions

FIH crystallises reliably in the presence of cofactors, with and without substrates, under optimised conditions (1.4-1.9 M ammonium sulphate, HEPES pH = 7.5, 4-7% PEG400) at room temperature in sitting, as well as, hanging drop vapour diffusion conditions. The FIH crystals however have a tendency to sick to the bottom in sitting drop plates making harvesting challenging, especially for crystal slurries needed for XFEL. Therefore, hanging drop vapour diffusion methods were pursued.

The crystal size highly depends on the number of nuclei forming in the crystallisation drops. This can lead to an inhomogeneous crystal size distribution (Figure 6A). Introducing additional crystallisation seeds *via* seeding stocks allows controlling of the number of nucleation points present in the drops. Thereby crystal numbers and crystal sizes can be controlled leading to a homogenous size distribution (Figure 6B). Optimisation of batch methods for FIH micro-crystallisation are described below.

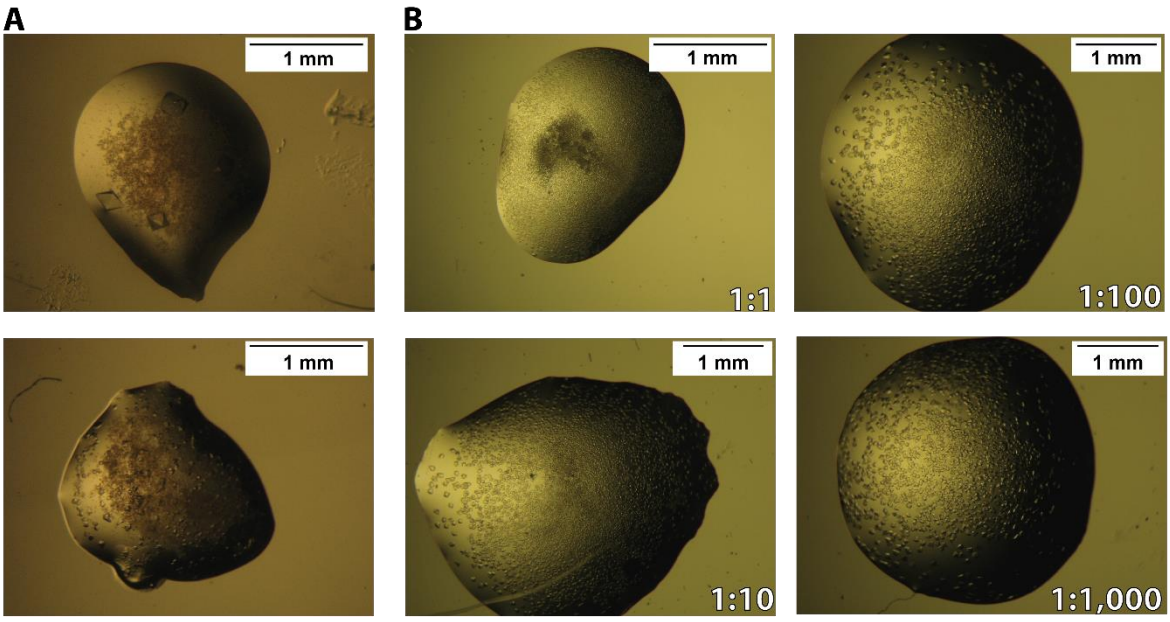


Figure 6 FIH Microcrystal optimisation. Anaerobically grown crystals of the FIH.Fe(II).2OG complex under optimized crystallisation conditions. Hanging drop vapour diffusion, drop size 2 µL well solution + 2 µL protein solution **A** Representative drops for FIH.Fe(II).2OG complex crystals without added seed. **B** Serial dilution of FIH seed stocks for FIH.Fe(II).2OG complex crystallisation.

8.3.1.2 FIH sample preparation

Following from the FIH.Fe(II).2OG complex micro-crystal optimisation, FIH was crystallised in the presence of co-factors and co-substrates with a HIF-1α derived 19mer peptide substrate. A new seed stock for the FIH-HIF complex was prepared from 8 wells containing large crystals of the FIH complex in 50 µL well solution and was diluted 1:10 in well solution. FIH was crystallised in a single condition (1.4 M ammonium sulphate, 5% PEG400, 100 mM HEPES pH = 7.5) in 24 well hanging drop plates, 6 uL drops (3 uL protein solution, 2.5 uL well solution and 0.5 uL seeding stock 1:10 dilution) with up to 4 drops per well. The crystals appeared within 6h. Crystal stocks were harvested by re-suspension of the crystal slurry in 5 uL well solution and combining the re-suspended slurries in a 1.5 mL Eppendorf tube. A representative drop and the running sample are shown in Figure 7.

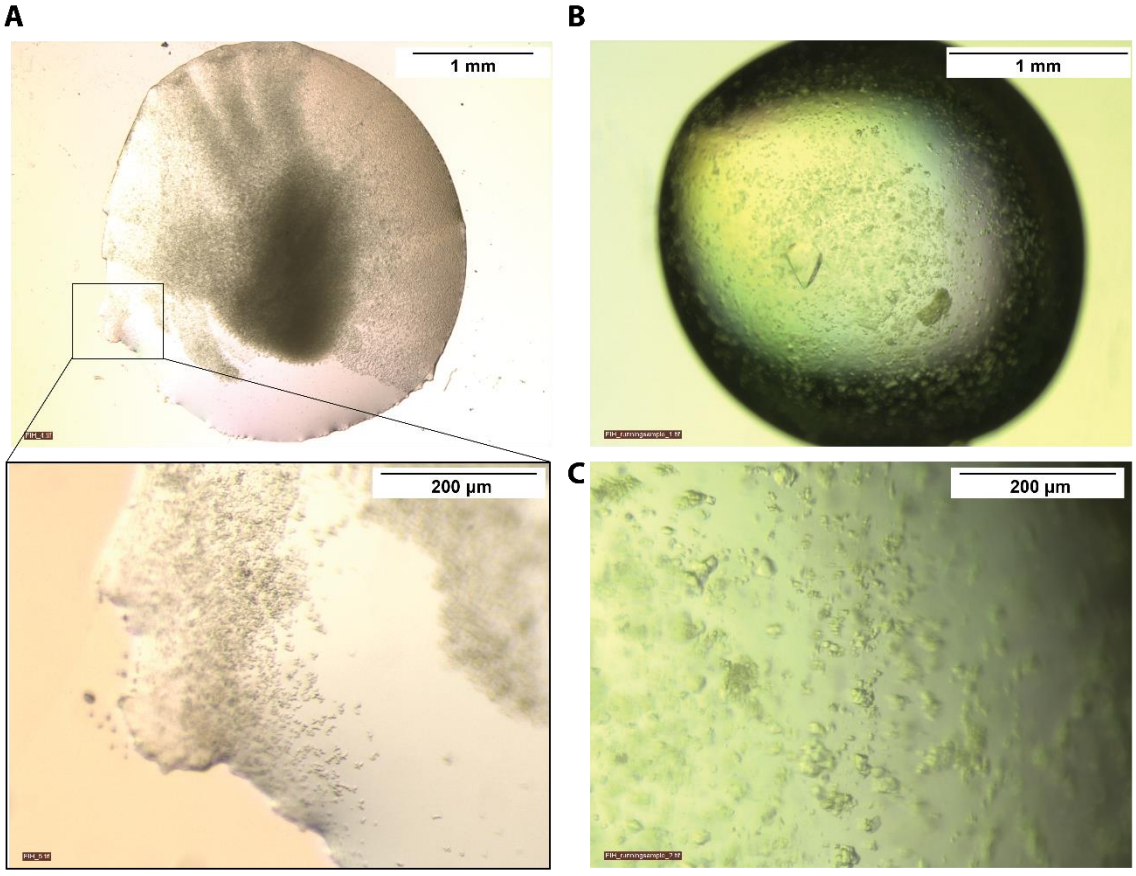


Figure 7 FIH-HIF-1α XFEL sample. Anaerobically grown microcrystals of the FIH.Fe(II).2OG.HIF-1α(19mer) complex used for the XFEL measurements. Note that the samples were exposed to oxygen while pictures were taken. No changes in crystal morphology have been observed for FIH crystal taken from anaerobic to aerobic conditions. **A** Representative drop containing FIH microcrystals, the average crystal size is 5-10 µm. **B,C** Sample from the data collection illustrates crystal size and density in the slurry used for the XFEL measurements. Note, a large FIH crystal observed in the drop potentially continued to grow in the running sample.

6.3.1.3 Data collection

FIH micro-crystals, crystallised as described above, were re-suspended without further concentration and loaded into a gas-tight syringe (Hamilton Co). Due to a small nitrogen bubble in the syringe, introduced during sample loading, the syringe was attached to the sample delivery system with the outlet facing downwards and not shaken during data collection. The flow rate of the crystal slurry during collection was 3-4 µL/min and detector was run in a 2x2 binning mode at 10 Hz. An overview of hit rates and indexing rates is shown in Figure 8. The initial hit and indexing rates for FIH crystals were ~30%. FIH crystals settled during the data collection at the bottom of the syringe leading to occasional blocking of the system indicated by significant drops in the hit rate (Figure 8). During the data collection, the hutch had to be opened on several occasions and the syringe manually shaken to loosen the clogging of the system.

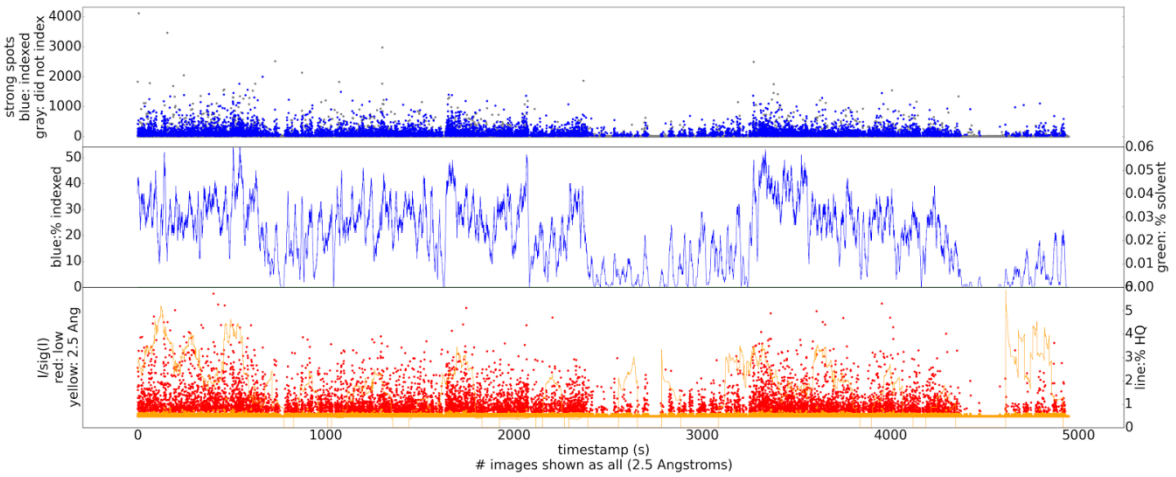


Figure 8 Hit rates and indexing rates of FIH resting state data collection. Top: Strong spots captured by the detector, blue spots were indexed, grey spots were not indexed. **Middle:** percentage of indexed images over time. Bottom: I/σ(I) of spots in low and high-resolution shells.

Following from the initial success of the data collection of FIH in complex with its HIF-1α substrate, Fe(II) and 2OG, it was tested if the crystal slurry is still catalytically active. FIH crystals in the resting state were taken out of the glove box and the drops were allowed to equilibrate with ambient air. Analogous to previously described FIH micro-crystal harvesting (see above), the oxygenated FIH crystals were harvested outside the glove box and combined in two Eppendorf tubes and incubated for a total of 12-14 h. Crystals settled over the course of 6-8 h in the tubes yielding a white crystal pellet (Figure 9A). The pellet was suspended in ~500 µL of supernatant and used for XFEL measurements. The hit rates for the FIH-oxygenated sample were overall lower than the hit rates observed for the FIH complex in the resting state. For future application, the natural settling of the crystals could be used for concentrating the crystal slurry prior to the measurements increasing crystal density in the crystal slurries and thus increasing hit rates and thereby reducing measuring times. A representative diffraction pattern of an FIH crystal measured at the LCLS is shown in Figure 9B.

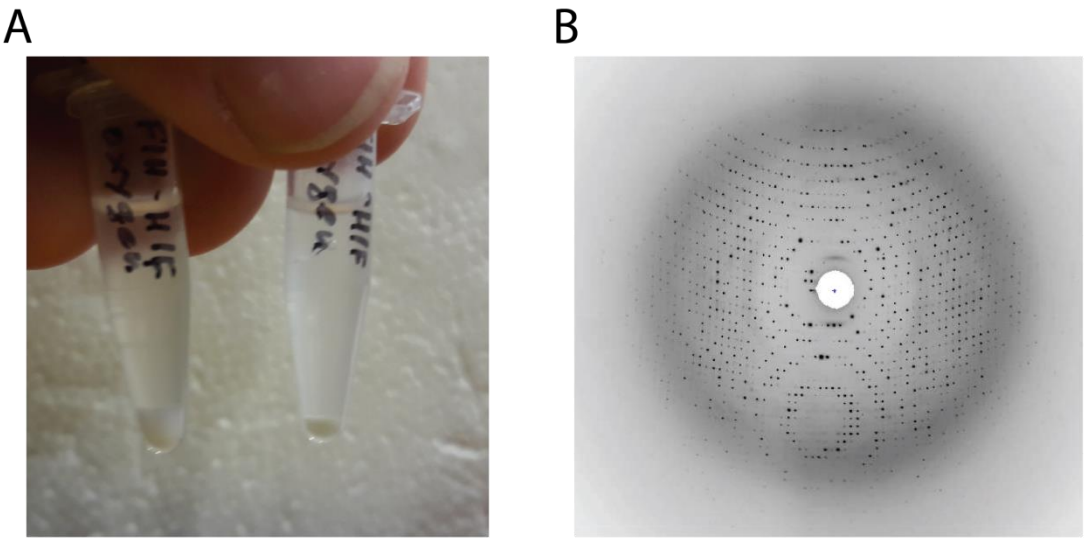


Figure 9 XFEL samples of FIH. A Harvested FIH crystals settled over the course of several hours in the tubes forming a white pellet. **B** Representative diffraction pattern of FIH crystals at the LCLS MFX XFEL Beamline showing diffraction >2.5 Å.

8.4 AlkB sample preparation and data collection

8.4.1 AlkB sample preparation

AlkB was crystallised according to previously described procedures (see above, with assistance of Ms. Anka Lucic, University of Oxford). AlkB micro-crystal growth was enhanced by addition of concentrated seed stocks. The AlkB crystals were grown with and without the addition of seeding stocks (Figure 10A). Addition of seeding stock, as previously observed for FIH, led to a high level of control in the number and size of crystals in each drop. Crystals used for the data collection were harvested as previously described for FIH. In brief, 4 µL of well solution was added to each drop, micro-crystals re-suspended and combined in Eppendorf tubes under anoxic conditions.

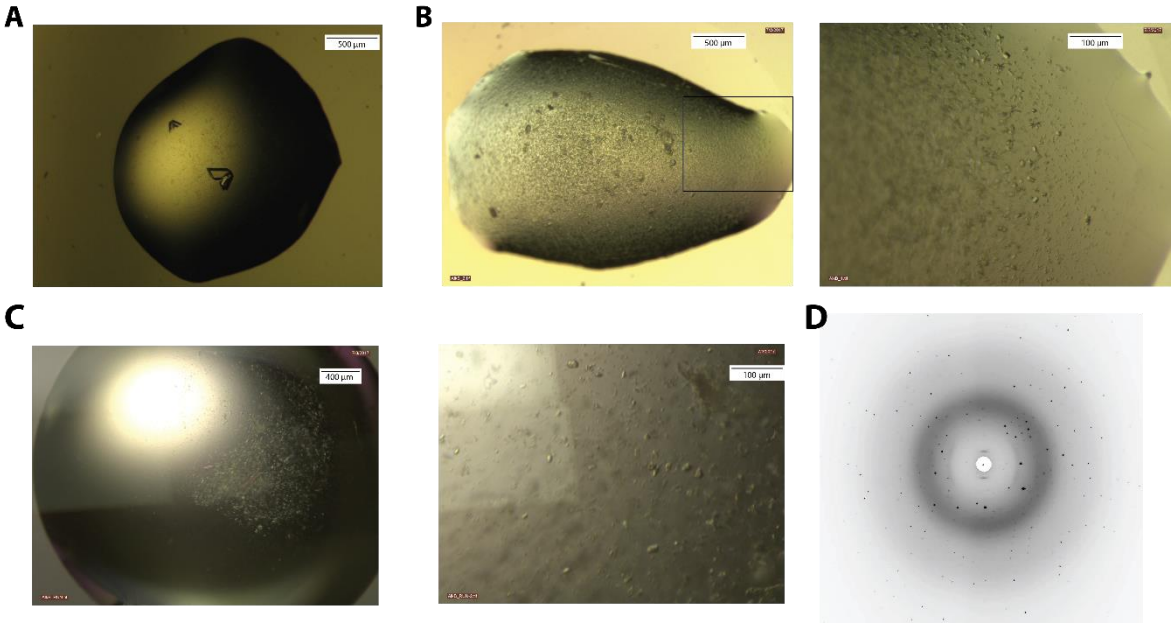


Figure 10 Micro-crystal optimization of AlkB.Fe(II).2OG complex with a T-1mA-T DNA fragment. A Crystal of AlkB substrate complex, hanging drop 4 µL, 2 µL protein and 2 µL well solution. **B** Representative image of optimized AlkB micro-crystals. **C** Images from the sample used for AlkB data collection of the first day shows low crystal density. **D** Representative diffraction pattern of AlkB shows diffraction to the edge of the detector (~1.65 Å).

8.4.1 AlkB data collection at the LCLS

Data collection for AlkB in the resting state was performed at a flow rate of 4 µL/min with a detector frequency of 10 Hz. AlkB samples were split into two syringes containing ~800 µL crystal slurry each. Initial data collection showed only a low hit and indexing rate (Figure 11). Data collection was stopped after ~90 min with ~1,500 images collected as it was unlikely to get a full dataset to high resolution with this crystal concentration. Crystals in the 2nd syringe were allowed to settle for ~24 h and the supernatant removed. A total of ~300 µL concentrated crystal slurry remained and was used for a second data collection with ~5,500 images in ~60 min. Variations in the cell-constants, potentially due to variations in the hydration state of the crystals, were observed in the sample. Settling and equilibrating the crystals overnight in the syringe led to homogenous unit cell constant distributions (data not shown). This increased sample quality and together with higher concentration of the sample the data collection efficiency and overall data quality was improved.

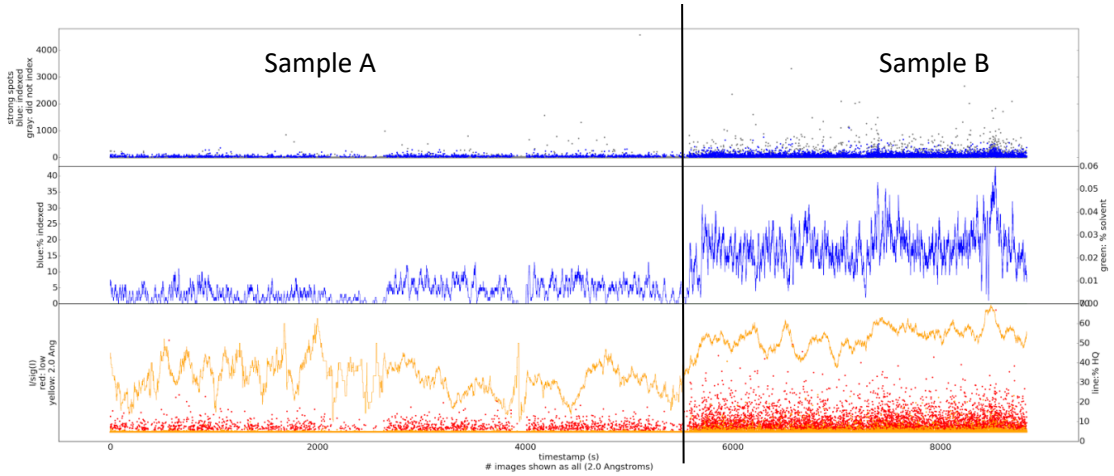


Figure 11: Hit rates and indexing rates for AlkB resting state data collection. Sample was changed after ~5800 s. **Top:** Strong spots captured by the detector, blue spots were indexed, grey spots were not indexed. **Middle:** percentage of indexed images over time. Bottom: I/σ(I) of spots in low and high-resolution shells.

To investigate whether AlkB could turn over its substrate in the crystalline state and if the reaction could be followed on a structural basis (analogous to FIH, see above), the AlkB crystal slurry was allowed to equilibrate with ambient oxygen for 6 h. A high concentration of PEG 3350, with the addition of glycerol, makes the AlkB crystallisation conditions very viscous and the settling of the crystals very slow. To concentrate the crystals, samples were spun down using a bench-top centrifuge. The crystal pellet was re-suspended in ~300 µL supernatant (~5-fold concentration). The overall crystal density was lower than observed for the overnight settled sample but within 40 min ~2,500 images could be collected. Due to beam time limitations, the data collection could not be continued after 40 min.

8.5 XFEL data processing from the LCLS

During data collection, an arbitrary number of images were grouped as runs for data analysis. For each run, data collection parameters, including beam position, detector position and kapton belt shadow (discussed below) were assumed to be constant. Data were processed using the cctbx-xfel pipeline (Dr Nicholas Sauter and coworkers, University of California Berkeley, discussed on the example of FIH XFEL resting state data collection). Data processing was performed in collaboration with Dr Pierre Aller (Diamond Light Source) and Dr Ian Clifton (University of Oxford).

Data collection of FIH in the anoxic resting state was split into 16 runs. Runs 1-2 were collected with a detector distance of 99 mm, run 3 with a detector distance of 94 mm and runs 4-16 were collected with a detector distance of 84 mm. Detector distances were reduced to capture high resolution spots.

For each run group, the mask of the beam centre was adjusted manually. Due to the nature of the experiment no beam stop is in place. Images were indexed using dials.stills. Importantly, the significance filter needs to be enabled which applies an I/σI filter of 1 for each image to avoid completeness problems in the downstream data processing and refinement. Initial cell constants for indexing were used from a corresponding FIH.Fe(II).HIF-1α cryo-structure (PDB code 1H2K[52]). To account for variation in the unit cell size at room temperature compared to cryo-structures, cell parameters for the crystals measured at the XFEL were refined (Figure 12) leading to smaller variation in cell constants observed which enabled the processing of the different run groups together.

The tape drive used for ADE sample delivery is unique as a sample delivery system for XFEL but the kapton belt casts a shadow over the diffraction pattern, reducing the intensity of only some reflections (Figure 13A). To account for the reduced intensity observed in the reflection patter in the kapton belt shadow, the shadow is manually assigned for each run-group and reflection intensities, affected by the shadow, normalised to the reduction of background intensity.

Finally, the detector geometry and distance are refined to account for inaccuracies in detector geometry and distance (Figure 13B). The average detector distance is off by ~1% with a distance difference of ~0.4 mm between the edges of the detector (Figure 13B).

Data was merged using PRIME and structures solved using PHASER[53] (search model PDB code: 1H2K[52]). Altering cycles of fitting in COOT[54] and refining using PHENIX[55] were performed until R_{work} and R_{free} converged (Table 3).

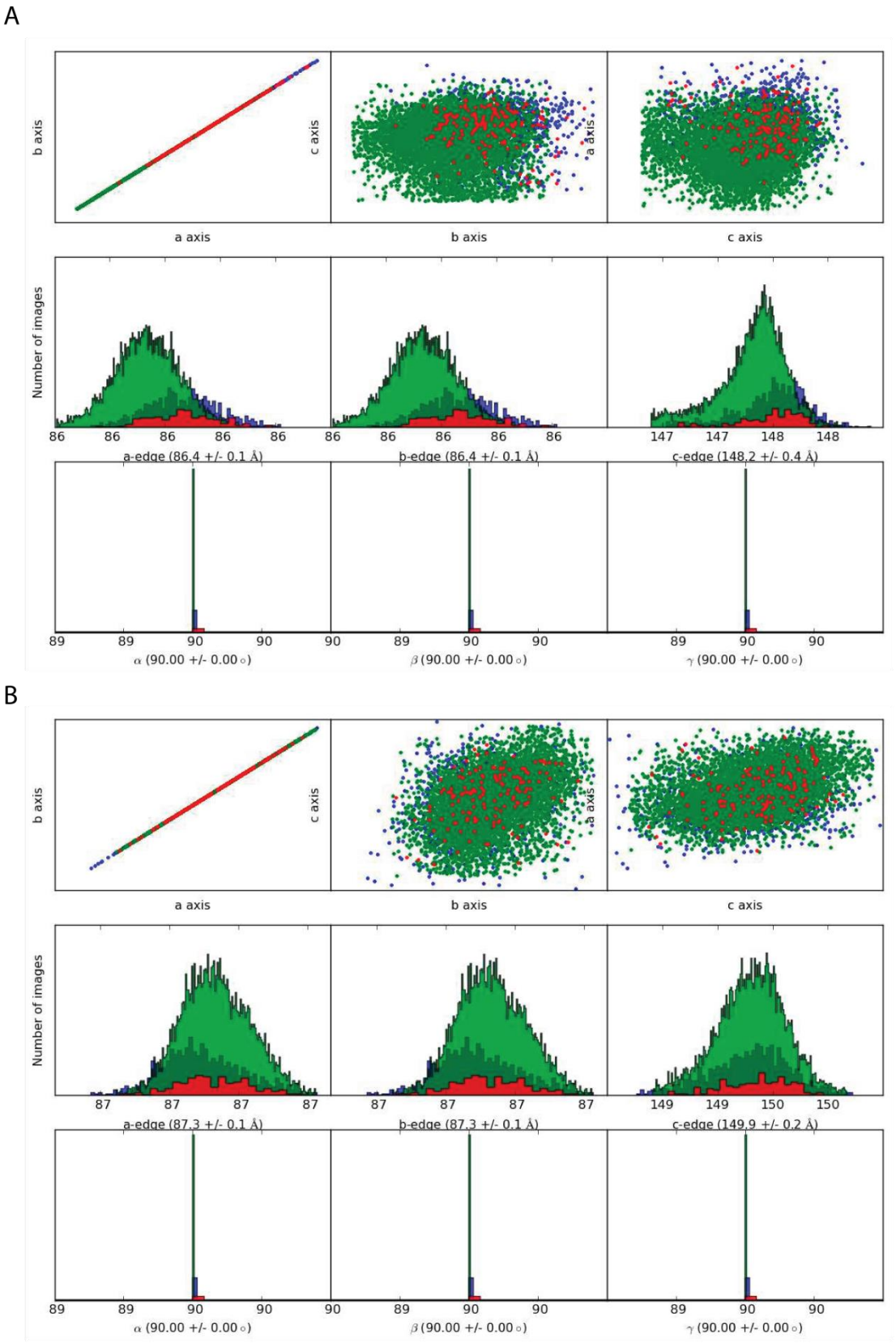


Figure 12 Ensemble refinement of FIH XFEL data. A Unit cell dimensions and angles of FIH.Fe(II).2OG.HIF-1α crystals measured at the LCLS before and **B** after ensemble refinement. Each dot represents a single image. Images are colour-coded according to run-groups. Green detector distance 84 mm, red detector distance 94 mm and blue detector distance 99 mm.

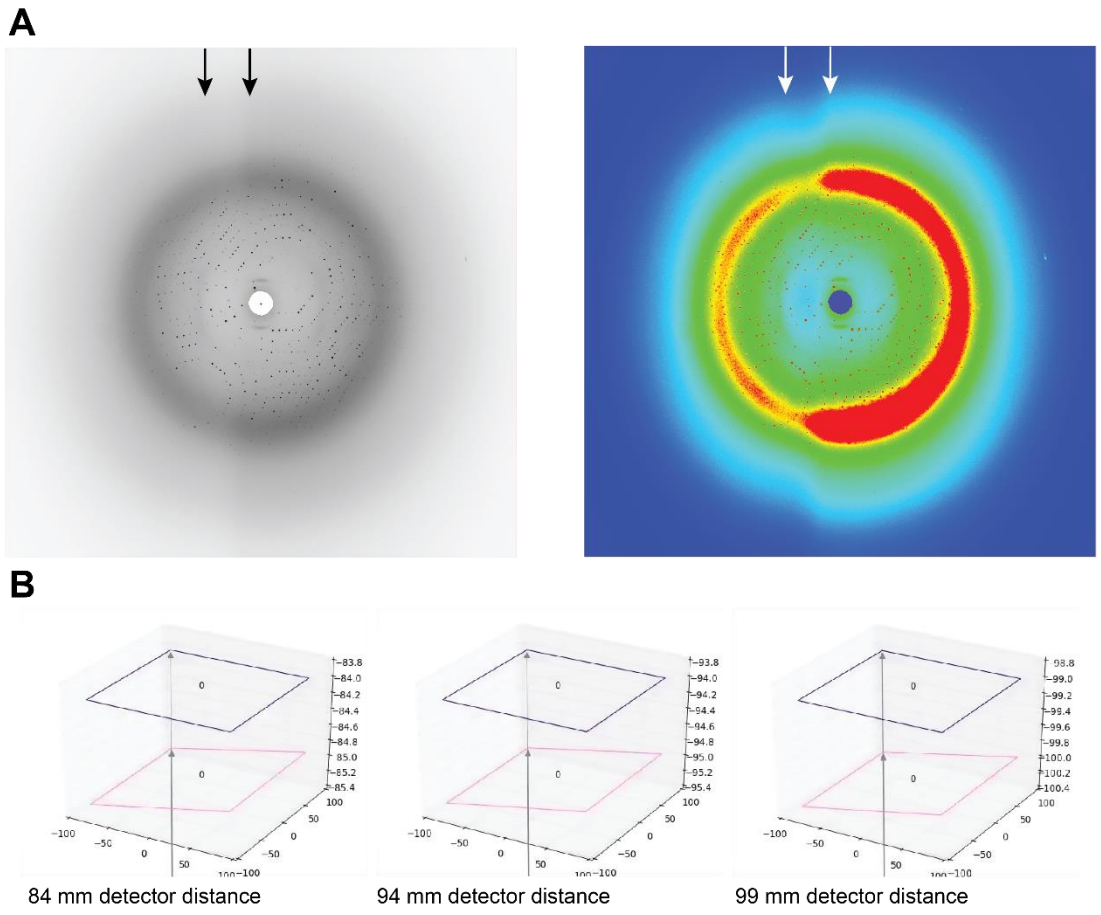


Figure 13 Kapton belt correction and detector distance and geometry refinement of FIH XFEL data. A Representative diffraction pattern collected of an FIH.Fe(II).2OG.HIF-1α crystals at the LCLS in grey-scale and rainbow displays the shadow cast by the Kapton belt from the tape drive. Edge of the Kapton belt shadow and maximum of the shadow are marked by the arrows **B** Pre-and post-refinement of detector distances and geometry for different detector distances.

AlkB data processing was performed according to previously described methodology (see above). Data for the AlkB resting state were collected on two consecutive days. Variations in cell constants were observed between different runs and days, challenging data processing. Batch variations between protein samples have previously been observed, possibly resulting from different hydration status of the crystal slurries (Dr Aaron Brewster, University of California, Berkeley, unpublished data).

Equilibrating the crystal slurries overnight, variations of the cell constants were reduced yielding a homogenous sample. (Data processing is ongoing, preliminary data collection and refinement statistics are shown in Table 4).

Table 3 Data collection and refinement statistics of FIH.Fe(II).2OG.HIF complexes at the LCLS.

Datasets	**FIH**.Fe(II).2OG.HIF-1α anoxic resting state	**FIH**.Fe(II).2OG.HIF-1α oxidised state
Data Collection		
Beamline (Wavelength, Å)	LCLS MFX	LCLS MFX
Detector	Rayonix MX170-HS (2x2 binning)	Rayonix MX170-HS (2x2 binning)
Data Processing	cctbx.xfel/ PRIME	cctbx.xfel/ PRIME
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	87.3, 87.3, 149.3	87.4, 87.4, 150.1
α,β,γ(°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
No. images collected	49490	55060
No. images indexed	8810	8169
No. images integrated	8318	7887
No. reflections	26432 (1272)*	26257 (1111)*
Resolution (Å)	21.84-2.30 (2.34-2.30)*	21.84-2.30 (2.34-2.30)*
R_{merge}	64.49 (87.58)*	65.04 (67.49)*
R_{split}	16.35 (73.70)*	17.61 (65.08)*
I/σI	1.56 (0.72)*	1.65 (0.67)*
CC1/2	0.978 (0.506)*	0.973 (0.709)*
Completeness (%)	99.9 (99.3)*	99.1 (87.3)*
Wilson B value (Å^2)	31.78	30.37
Refinement		
R_{work}/R_{free}†	0.2202/0.2490	0.2212/0.2548
No. atoms		
-Enzyme (A/B)	2686	2633
-Substrate	92	33
-Ligand	2OG (10)	2OG (10)
-Water	93	105
Avg. B-factors		
Enzyme	46.5	47.1
Substrate	89.1	96.7
Ligand	27.6	53.2
Water	43.4	35.9
R.m.s deviations		
-Bond lengths (Å)	0.002	0.002
-Bond angles (°)	0.510	0.486

* Highest resolution shell in parentheses.

†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

Table 4 Data collection and refinement statistics of AlkB.Fe(II).2OG.T-mA-T complexes at the LCLS.

Datasets	**AlkB**.Fe(II).2OG.T-1mA-T anoxic resting state	**AlkB**.Fe(II).2OG.T-1mA-T oxygenated for 6 h
Data Collection		
Beamline (Wavelength, Å)	LCLS MFX	LCLS MFX
Detector	Rayonix MX170-HS (2x2 binning)	Rayonix MX170-HS (2x2 binning)
Data Processing	cctbx.xfel/ cxi.merge	cctbx.xfel/ cxi.merge
Space group	*P*1	*P*1
Cell dimensions		
a,*b*,*c* (Å)	36.3, 39.2, 40.8	37.2. 39.4, 40.0
α,β,γ(°)	79.0, 78.3, 66.9	78.1, 76.0, 66.5
No. images collected	32525	25062
No. images indexed	5956	2230
No. images integrated	5949	2230
No. reflections	10202(511)*	4093 (194)*
Resolution (Å)	28.33-2.20 (2.31-2.20)*	21.62-3.00 (3.43-3.00)*
R_{split}	42.7 (50.9)*	41.9 (52.4)*
I/σI	122.9 (51.6)*	98.4 (33.4)*
CC1/2	71.7 (52.2)	68.5 (42.8)*
Completeness (%)	100 (100)*	99.8 (100)*
Multiplicity	19.8 (10.0)*	13.0 (6.9)*
Wilson B value (Å^2)	12.1	25.4
Refinement		
R_{work}/R_{free}†	0.2464/0.2828	0.2305/0.2644
No. atoms		
-Enzyme (A/B)	1572	1542
-Substrate	35	---
-Ligand	2OG (10)	2OG (10)
-Water	106	4
Avg. B-factors		
Enzyme	13.0	22.0
Substrate	30.5	---
Ligand	12.1	45.8
Water	20.2	8.0
R.m.s deviations		
-Bond lengths (Å)	0.002	0.003
-Bond angles (°)	0.622	0.580

* Highest resolution shell in parenthesis.

†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

8.6 XFEL crystal structures of FIH and AlkB

8.6.1 XFEL crystal structures of FIH in the resting and oxidised states

The first XFEL structure of a 2OG oxygenase to be solved was the structure of FIH in complex with Fe(II), 2OG and HIF-1α (19mer) to 2.30 Å (Figure 14A). The structure was solved using molecular replacement (search model PDB code 1H2K[52]). Solving the structure reveals that that 2OG, Fe(II) and the HIF-1α peptide are bound to the active site (Figure 14A), confirming that structural information can be obtained from FIH micro-crystals at a high resolution under anaerobic conditions.

Solving the crystal structure of the oxygen exposed FIH.Fe(II).2OG.HIF-1α complex using molecular replacement as previously described (see above) revealed density for the active site metal, and some density in the 2OG binding pocket (Figure 14B). No density for the HIF-1α substrate, however, was observed in the active site. Refinement of the structure with 2OG and succinate suggests that, at least predominantly, 2OG, and not succinate, is bound (Figure 14B). It was noted that the average B-factor for 2OG is significantly higher (~2 fold) in the oxygen exposed crystals compared to the resting state at similar resolution (Table 5).

Some, relatively low level, density was observed in the substrate binding site further away from the active site (Figure 14B) close to the -8 leucine binding site (for modes of FIH substrate binding see Chapter 1). This can be explained by substrate turnover with resulting loss of affinity of the hydroxylated asparagine in the active site.

The significant changes observed between the structures of the resting state versus the structure of the end point of the reaction, after 12 h exposure to ambient oxygen, demonstrate that the FIH reaction can take place in crystals. Loss of the substrate density after oxygen exposure shows that structural changes in the active site can be observed.

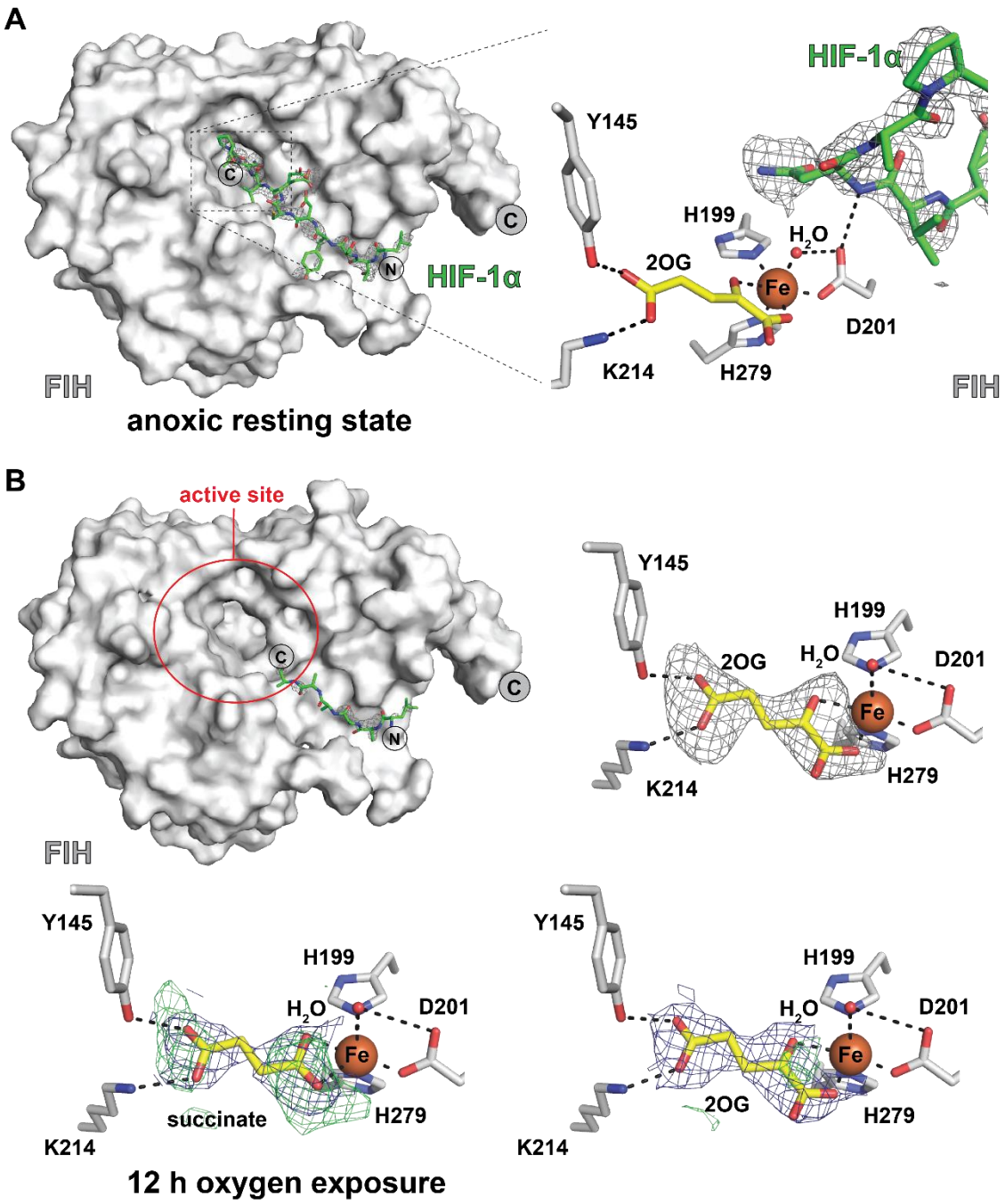


Figure 14 Views from XFEL crystal structures of FIH resting state and after ~12 h O_2 exposure. A Views from a crystal structure of anaerobically collected data from an FIH.Fe(II).2OG.HIF-1α(19mer) complex; the mFo-Fc OMIT map is contoured to 3σ around HIF-1α and shown in grey. **B** Views from a crystal structure of the FIH.Fe(II).2OG.HIF-1α(19mer) complex after 12 h oxygen exposure. **Top:** Views from the FIH proteins show little density for the HIF-1α substrate. **Bottom:** Models of succinate and 2OG bound to the active site of FIH after one round of refinement. 2mFo-Fc maps are contoured to 1σ (blue) and mFo-Fc maps contoured to ±3σ (green positive density and red for negative density). Note that the oxidation state of the active site iron could not be determined.

8.6.2 XFEL crystal structures of AlkB in the resting and oxidised states

Following from FIH XFEL crystal structures obtained in the resting state and after oxygen exposure, AlkB samples were prepared analogously. AlkB crystals were harvested under anaerobic conditions see above) and crystals divided into two batches. One set was measured anaerobically, the other was allowed to equilibrate with ambient oxygen levels for 6 h prior to XFEL measurements. The low symmetry of the AlkB crystals and batch to batch variations for the AlkB crystals made data processing challenging. Note, to date, no XFEL crystal structure of any protein in space group *P*1 has been reported.

Images from the data collection show diffraction to the edges of the detector (~1.65 Å) for both samples but the data could, to date, only be processed to 2.20 Å and 3.00 Å for the anoxic resting state and the oxygen exposed sample respectively (the preliminary data is shown in Figure 15).

The overall fold of the AlkB protein in the anoxic state compared to the oxygen exposed sample, did not show any significant changes (Figure 15). After exposure to ambient air for 6 h, little to no density for the 1-methyladenine could however be observed in the active site compared in the oxygen exposed AlkB.Fe(II).2OG.T-1mA-T complex structure. Notably, in the resting state the active site Fe(II) is coordinated by the facial triad and in a bidentate manner by 2OG with the last coordination site occupied by water. After exposure to oxygen, the coordination of 2OG and the sidechains of the facial triad are comparable (Figure 15). Low resolution of the oxygen exposed sample challenges the interpretation of the data.

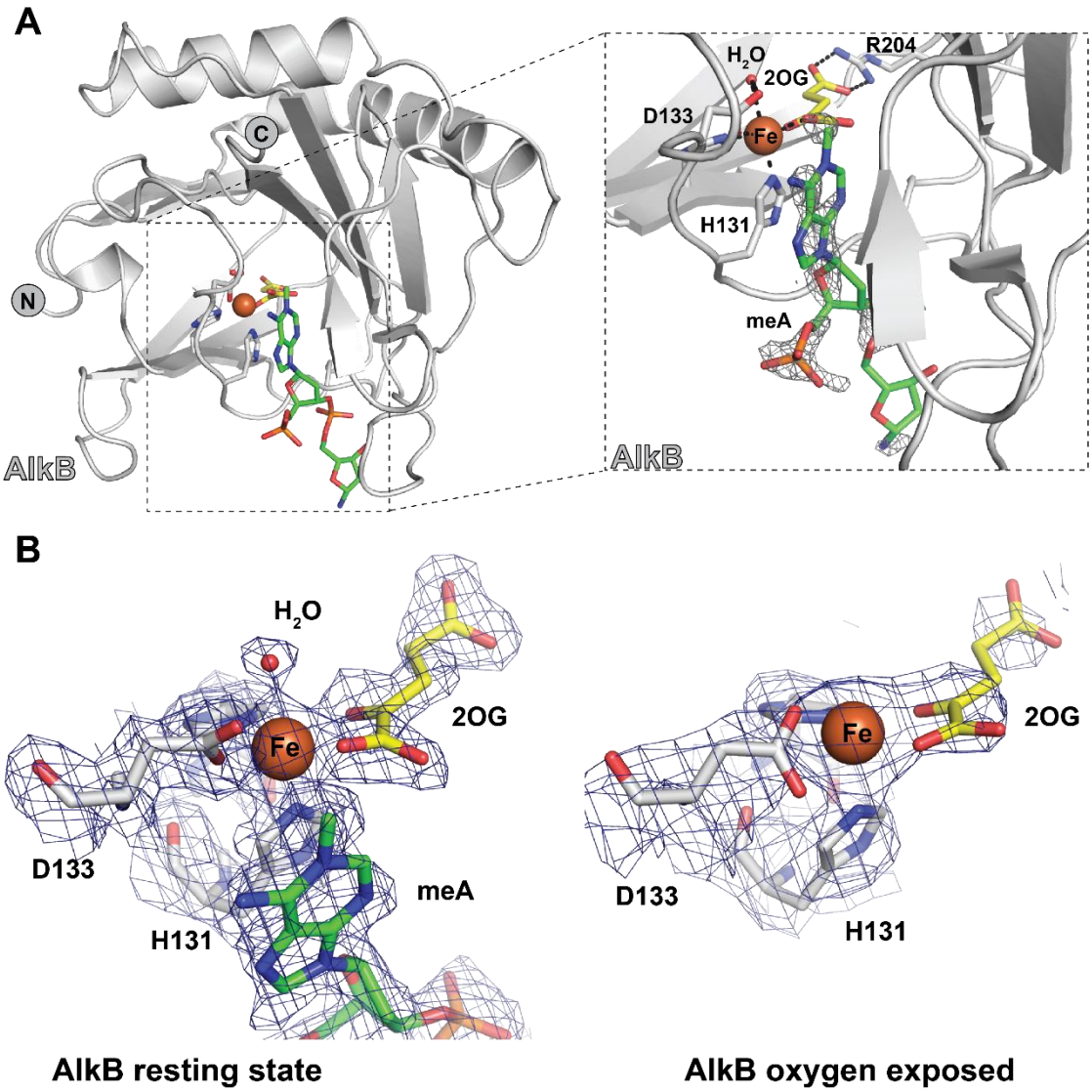


Figure 15 AlkB XFEL crystal structures in resting and oxidised state. A Views from a crystal structure of AlkB in the resting state at the room temperature, mFo-Fc OMIT maps contoured to 3σ around the T-1mA-T DNA ligand and shown in grey. **B** Views from the active site of AlkB in the resting state (left) and after oxygen exposure for 6 h (right). 2mFo-Fc maps are contoured to 1 σ.

8.6.3 Cryo-crystal structures of FIH and AlkB in the resting state and after oxygen exposure

To investigate the impact of cryo-cooling compared to room-temperature data collection of FIH and AlkB crystals, data from corresponding the FIH.Fe(II).2OG.HIF-1α complex and the AlkB.Fe(II).2OG.T-1mA-T complexes were collected under cryo-conditions in single crystal sweep data collections at the Diamond Light Source. FIH and AlkB macro-crystals were grown under nearly anaerobic conditions (<1 ppm oxygen, see above) and crystals were harvested under anaerobic conditions. The same crystal plate was then taken out of the anaerobic chamber and allowed to equilibrate with ambient oxygen levels prior to cryo-cooling (see Figure 16 and Figure 17). Data from single crystals were collected (Table 5) and the structures solved using molecular replacement (see above).

A crystal structure of an anaerobically frozen crystal of FIH in complex with Fe(II), 2OG and its HIF-1α substrate shows clear density corresponding to the substrate in the active site and the substrate binding site (Figure 16A) in an overall conserved manner compared to other known FIH substrates complexes (see also chapter 1 and chapter 4). 2OG is bound to the active site as previously observed and the Fe(II) is coordinated hexagonally by the facial tirade, 2OG and one water. Data from FIH.Fe(II).2OG.HIF-1α crystals exposed to ambient oxygen levels for 12 h showed no detectable density corresponding to bound substrate in the active site or in the substrate binding sites further away for the active site. Density corresponding to bound 2OG can however be observed in the active site of FIH of the oxygen expose substrate complex (Figure 16B). Note, the oxidation state of the iron could not be determined.

Several datasets for AlkB in complex with a modified T-1mA-T DNA fragment were collected after different exposure times to ambient oxygen levels (Figure 17). The structure of anoxic harvested and cyro-cooled AlkB.DNA complexes is described above. After approximately 6 min exposure to ambient oxygen levels, little to no density corresponding to bound substrate could be observed and succinate was found to bind to the active site of AlkB and not 2OG indicating, at least nearly complete, 2OG turnover. Crystals frozen after 6 h, corresponding to the XFEL structure, show density in the active site corresponding to the demethylated product and 2OG, indicating substrate turnover and replacement of the resulting succinate by 2OG. After 22 h exposure to ambient oxygen levels, no density corresponding to substrate or product could be observed and clear density for 2OG bound to the active site of AlkB could be observed.

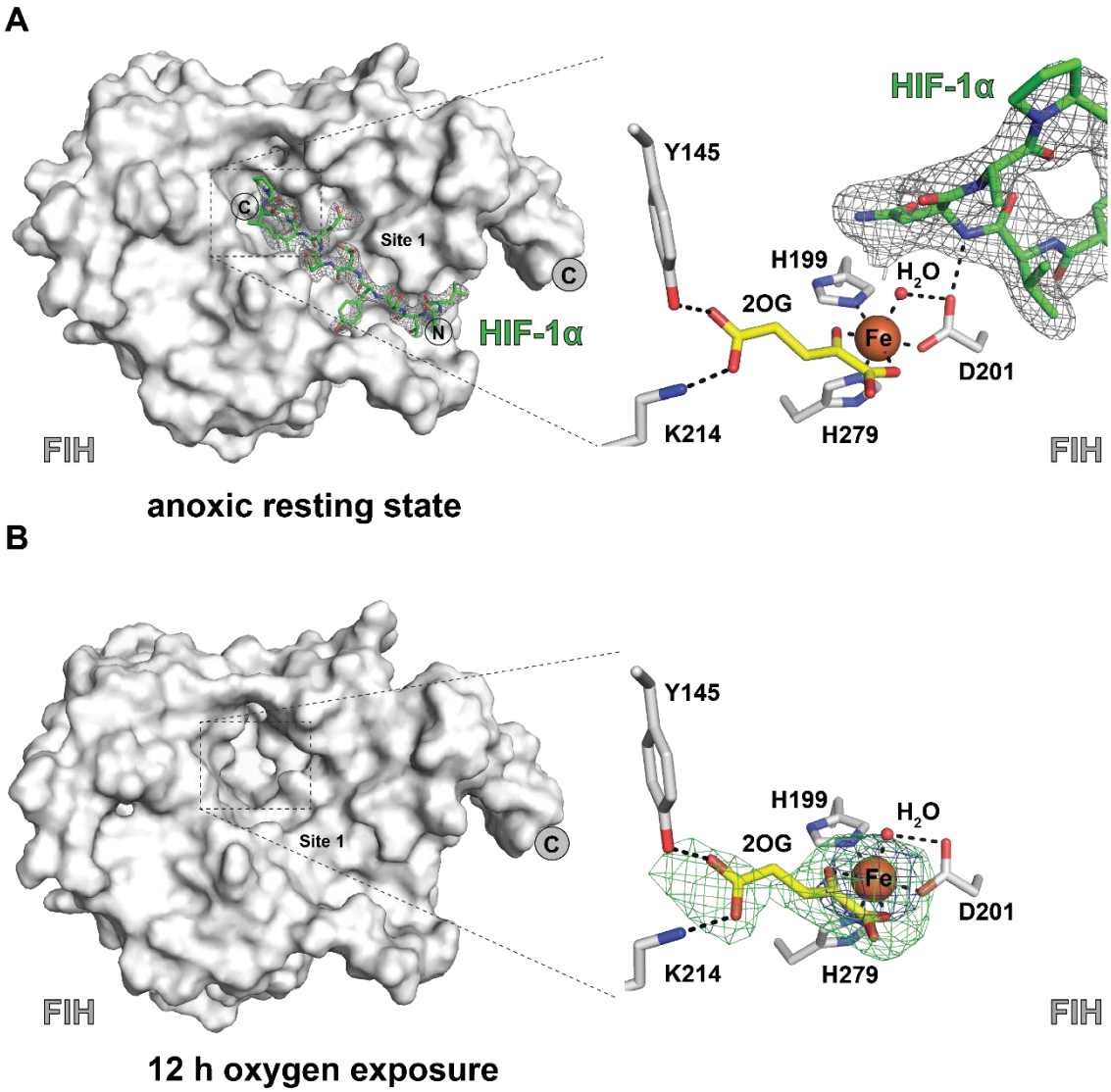


Figure 16 Views from cryo-crystal structures of FIH in the anoxic resting state and after 12 h O_2 exposure. A Views from a crystal structure of an anaerobically harvested FIH.Fe(II).2OG.HIF-1α(19mer) crystal; the mFo-Fc OMIT maps are contoured to 3σ around the HIF-1α substrate and coloured in grey. **B** Views from a cryo-crystal structure of the FIH.Fe(II).2OG.HIF-1α(19mer) complex after exposed to oxygen for 12 h. Views from the active site after refinement without 2OG and Fe(II)/(III) in the active site: mFo-Fc maps are contoured to ±3σ (coloured in green for positive values) and 2mFo-Fc maps are contoured to 1σ (coloured in blue) around 2OG and Fe(II)/(III); note, the oxidation state of the iron was not determined.

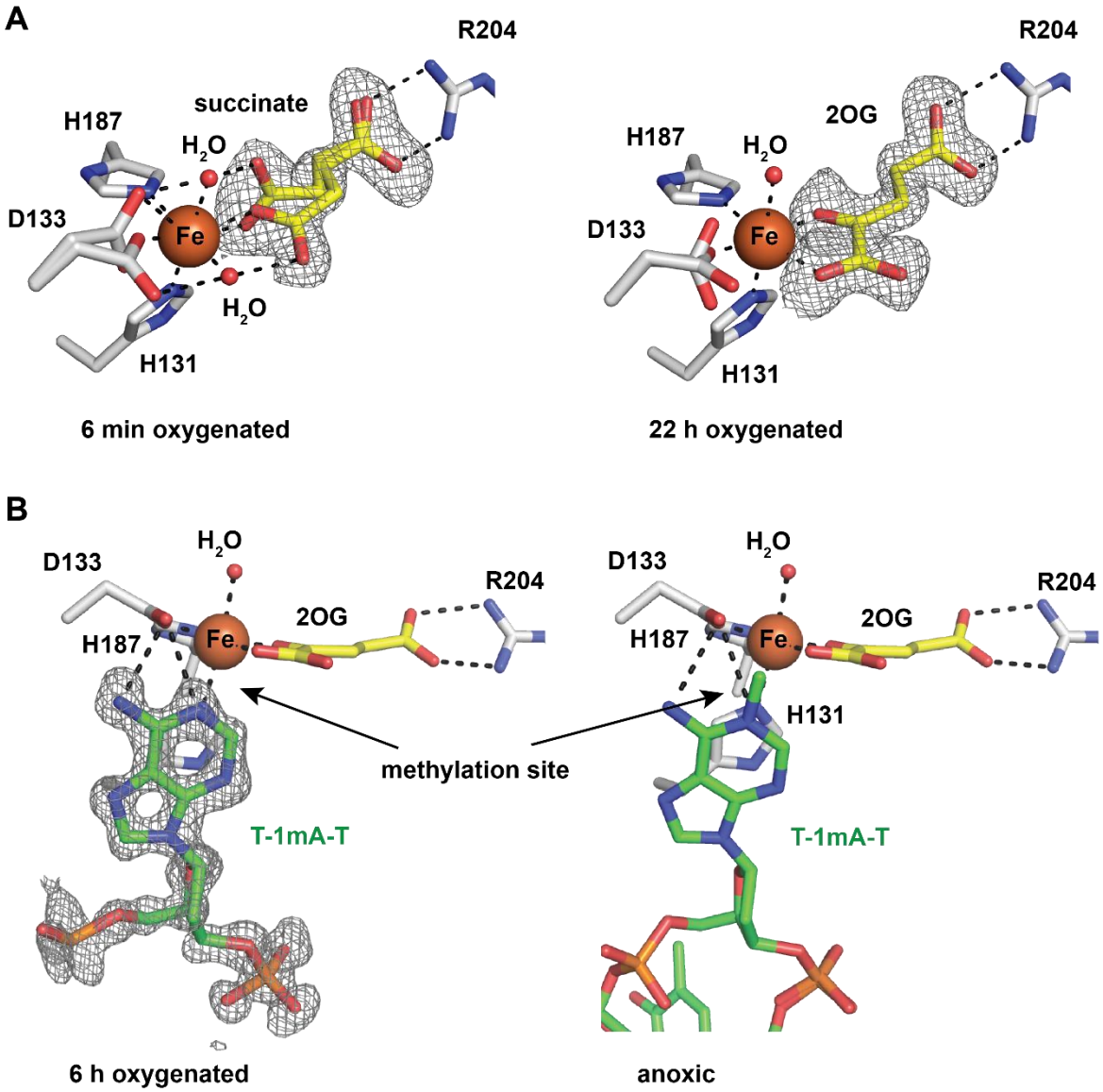


Figure 17 Views from a crystal structures of AlkB in complex with T-1mA-T substrate, Fe(II) and 2OG. A Views from a cryo-crystal structures of the AlkB.Fe(II).2OG.T-1mA-T complex oxygenated for 6 min and 22 h. Note, after 6 min oxygen exposure, succinate was found to be binding in the active site of AlkB instead of 2OG. No density resulting from bound substrate or product was observed. **B** Left: View from a crystals structure of AlkB.Fe.2OG.T-1mA-T after 6 h oxygen exposure shows the demethylated product bound to the active site next to 2OG. Right: View from a crystal structure of AlkB.Fe(II).2OG.T-1mA-T in the resting state. Note that the oxidation state of the active site iron could not be determined.

Table 5 Data collection and refinement statistics of FIH.Fe(II).2OG.HIF-1α complexes under anaerobic and aerobic conditions.

Datasets	**FIH**.Fe(II).2OG.HIF-1α resting state anoxic	**FIH**.Fe(II).2OG.HIF-1α oxygenated 12 h
Data Collection		
Beamline (Wavelength, Å)	DLS I03 (0.9763)	DLS I04-1 (0.9159)
Detector	PILATUS 6M-F	PILATUS 6M-F
Data Processing	MOSFLM/SCALA	Xia2 3d (XDS)
Space group	$P4_12_12$	$P4_12_12$
Cell dimensions		
a,*b*,*c* (Å)	86.7, 86.7, 146.4	86.7, 86.7, 147.1
α,β,γ(°)	90, 90, 90	90, 90, 90
No. reflections	44694 (6407)*	24188 (1372)*
Resolution (Å)	86.59-1.90 (2.00-1.90)*	43.43-2.35 (2.39-2.35)*
R_{merge}	0.118 (1.429)*	0.079 (1.364)*
I/σI	9.8 (1.8)*	15.5 (1.9)*
CC1/2	0.998 (0.564)*	1.0 (0.7)*
Completeness (%)	100 (100)*	100 (100)*
Multiplicity	12.8 (12.4)*	10.5 (11.4)*
Wilson B value (Å^2)	38.3	60.01
Refinement		
R_{work}/R_{free}†	0.1765/0.2114	0.2073/0.2260
No. atoms		
-Enzyme (A/B)	2686	2798
-Substrate	92	---
-Ligand	2OG (10)	2OG (10)
-Water	93	74
Avg. B-factors		
Enzyme	46.5	76.2
Substrate	89.1	---
Ligand	27.6	71.4
Water	43.6	72.2
R.m.s deviations		
-Bond lengths (Å)	0.019	0.003
-Bond angles (°)	1.423	0.589

* Highest resolution shell in parentheses.
†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.

Table 6 Data collection and refinement statistics of AlkB.Fe(II).2OG.T-1mA-T complexes under anaerobic and aerobic conditions.

Datasets	**AlkB**.Fe(II).2OG T-1mA-T 6min oxygenated	**AlkB**.Fe(II).2OG T-1mA-T 6h oxygenated	**AlkB**.Fe(II).2OG T-1mA-T 22h oxygenated**
Data Collection			
Beamline (Wavelength, Å)	DLS I03 (0.9796)	DLS I03 (0.9796)	DLS I03 (0.9796)
Detector	PILATUS 6M-F	PILATUS 6M-F	PILATUS 6M-F
Data Processing	Xia 3dii	MOSFLM/POINTLESS/ AIMLESS/CTRUNC	Xia 3d
Space group	*P*1	*P*1	*P*1
Cell dimensions			
a,*b*,*c* (Å)	36.9, 39.1, 40.6	35.2, 38.7, 40.6	35.1, 38.6, 40.5
α,β,γ(°)	77.1, 74.1, 65.8	80.0, 78.3, 67.8	79.9, 78.1, 68.3
No. of molecules/ASU	1	1	1
No. reflections	34109 (1687)*	61628 (2946)*	51001 (2485)*
Resolution (Å)	38.72-1.44 (1.44-1.46)*	39.51-1.14 (1.16-1.14)*	26.81-1.24 (1.26-1.24)*
R_{merge}	0.054 (1.363)*	0.056 (0.121)*	0.065 (1.256)*
I/σI	11.5 (1.1)*	9.1 (3.8)*	9.3 (1.1)*
Completeness (%)	95.4 (93.6)*	88.2 (85.8)*	93.5 (90.7)*
Multiplicity	3.5 (3.6)*	2.9 (2.5)*	3.4 (3.4)*
Wilson B value (Å^2)	20.3	8.5	14.7
Refinement			
R_{work}/R_{free}	0.1643/0.1838	0.1487/0.1674	---
No. atoms			
-Enzyme (A/B)	1609	1684	---
-Substrate	---	26	---
-Ligand	8 (succinate)	2OG (10)	---
-Water	187	362	---
Avg. B-factors			
Enzyme	28.6	10.5	
-Substrate	---	13.8	---
-Ligand	29.6	8.5	---
-Water	39.8	24.1	---
R.m.s deviations			
-Bond lengths (Å)	0.009	0.0078	---
-Bond angles (°)	1.030	1.130	---

* Highest resolution shell in parentheses.
†$R_{factor}= \Sigma_{hkl}||F_{obs}(hkl)|-k|F_{calc}(hkl)||/\Sigma_{hkl}|F_{obs}(hkl)|$, R_{free} is the R_{factor} for ~5%of reflections excluded from the refinement.
** Refinement is ongoing.

8.6.4 Comparison of room temperature crystal structures with cryo- crystal structures

Comparing the structures of the FIH.Fe(II).2OG.HIF-1α complex collected under cryo-conditions at Diamond Light Source and at room temperature at the LCLS XFEL shows a conserved substrate binding mode of HIF-1α under anaerobic conditions (Figure 18A). The overall binding is nearly identical including active site interactions (Figure 18B). Alternative sidechain orientations can be observed for glutamate-801 and asparate-799 (Figure 18C), which could also result from crystal to crystal variations. The distances of the Cα-atoms of the main chain vary <1 Å between the two structures (Figure 18D).

Comparison of the two structures of oxygen exposed FIH.Fe(II).2OG.HIF-1α complex collected under cryo-conditions at Diamond Light Source and at room temperature at the LCLS XFEL show in both cases no density corresponding to the substrate in the active site (Figure 18E). In both cases, 2OG was found to bind, at least predominantly, in the active site and not succinate. Low occupancy for the HIF-1α substrate can be observed in the XFEL structure but not in the cryo-structure. Notably is also that the water-iron bond appears to be elongated in the case of the XFEL structure (~2.7 Å compared to 2.2 Å typical for iron(II)-water coordination, higher resolution data are needed to confirm this observation).

The AlkB dataset from the XFEL could, to date, only be processed to low resolution (2.20 Å for the resting state and 3.00 Å for the oxygen exposed substrate complex), although diffraction was observed to higher resolution (see above). The cryo-datasets were processed to a resolution of 1.10-1.46 Å. Significant differences in resolution challenges the direct comparison of the different structures. The overall fold, co-factor and substrate binding does not appear to be changed by cryo-protection compared to room-temperature data collection (see above). For detailed comparison, data processing for AlkB data from the LCLS has to be further optimised.

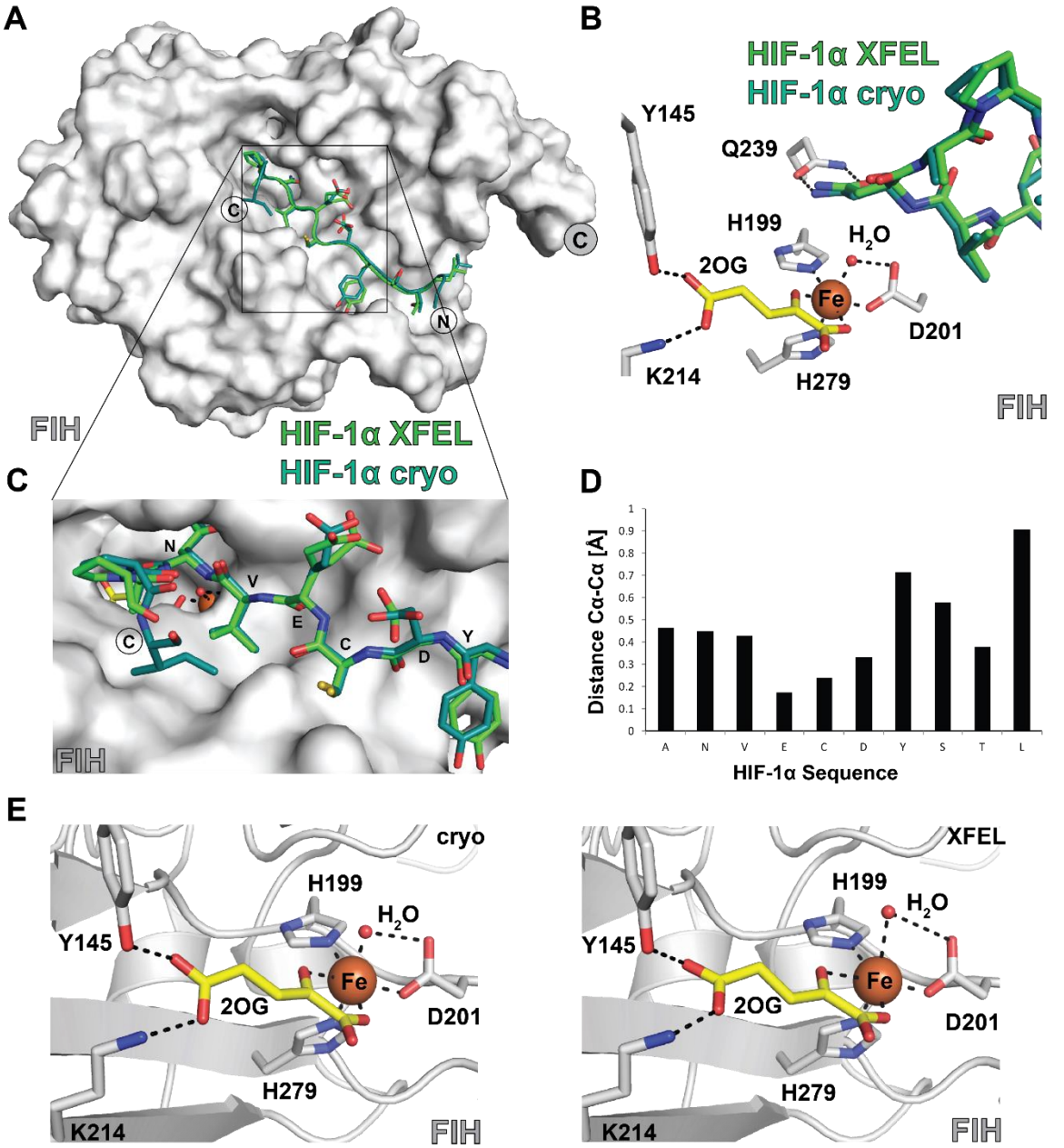


Figure 18 Comparison of cryo-cooled and room-temperature XFEL structures of FIH in complex with HIF-1α. A Overview of the HIF-1α binding in the FIH.Fe(II).2OG.HIF-1α complex in its resting state in a cryo-cooled crystal structure and a room temperature crystal structure. **B** Views from the active site of FIH in complex with HIF-1α binding shows conserved HIF-1α binding under cryo-conditions compared to a room-temperature structure. **C** Views from the FIH site one binding shows alternative side chain conformations of the HIF-1α substrate. **D** Distance between Cα of corresponding amino acids from HIF-1α in the synchrotron versus XFEL structures. **E** Comparison of the active sites of oxygen exposed FIH.HIF-1α complexes.

8.7 Preliminary experiments at the European XFEL

8.7.1 Sample selection for time-resolved crystallography at the European XFEL

Following from the initial experiments at the LCLS, leading to the first 2OG oxygenase structures in the resting state at room temperature, further studies were carried out at the European XFEL in Hamburg. In addition to FIH, which was measured at the LCLS, IPNS and the metallo-β-lactamases (MBLs) Verona-Integron-Encoded MBL-2 (VIM-2) and MBL-II from *Bacillus cereus* (BcII) were targeted. MBLs play a key role in antibiotic resistance of multiple resistance strains (MRS) and are of great clinical importance[56]. IPNS and VIM-2 and BcII share the common biological pathway of antibiotic biosynthesis and resistance but are unrelated in terms of their structures and reactions catalysed (Figure 19).

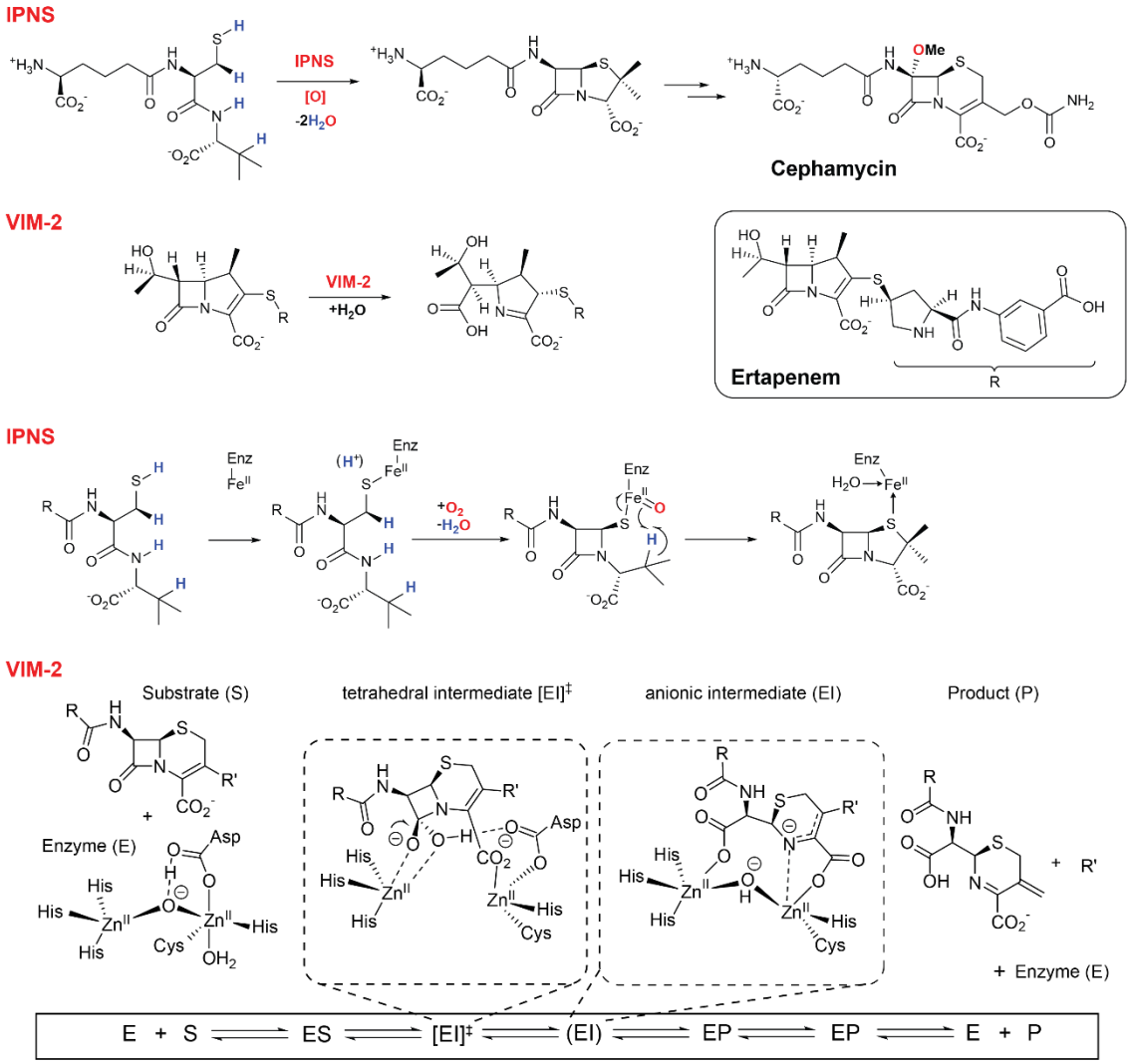


Figure 19 Simplified mechanisms of reactions catalysed by IPNS and VIM-2.

8.7.2 VIM-2 and BcII micro-crystal optimisation

Both VIM-2 and BcII are readily crystallisable proteins (Dr Jurgen Brem, University of Oxford, personal communication). Characteristic crystal forms of VIM-2 and BcII are shown in in Figure 20. As previously observed for FIH and AlkB (see above) addition of concentrated seed stocks to the crystallisation conditions allows the control of micro-crystal number and size. BcII appears to crystallise preferably in small clusters upon seed addition.

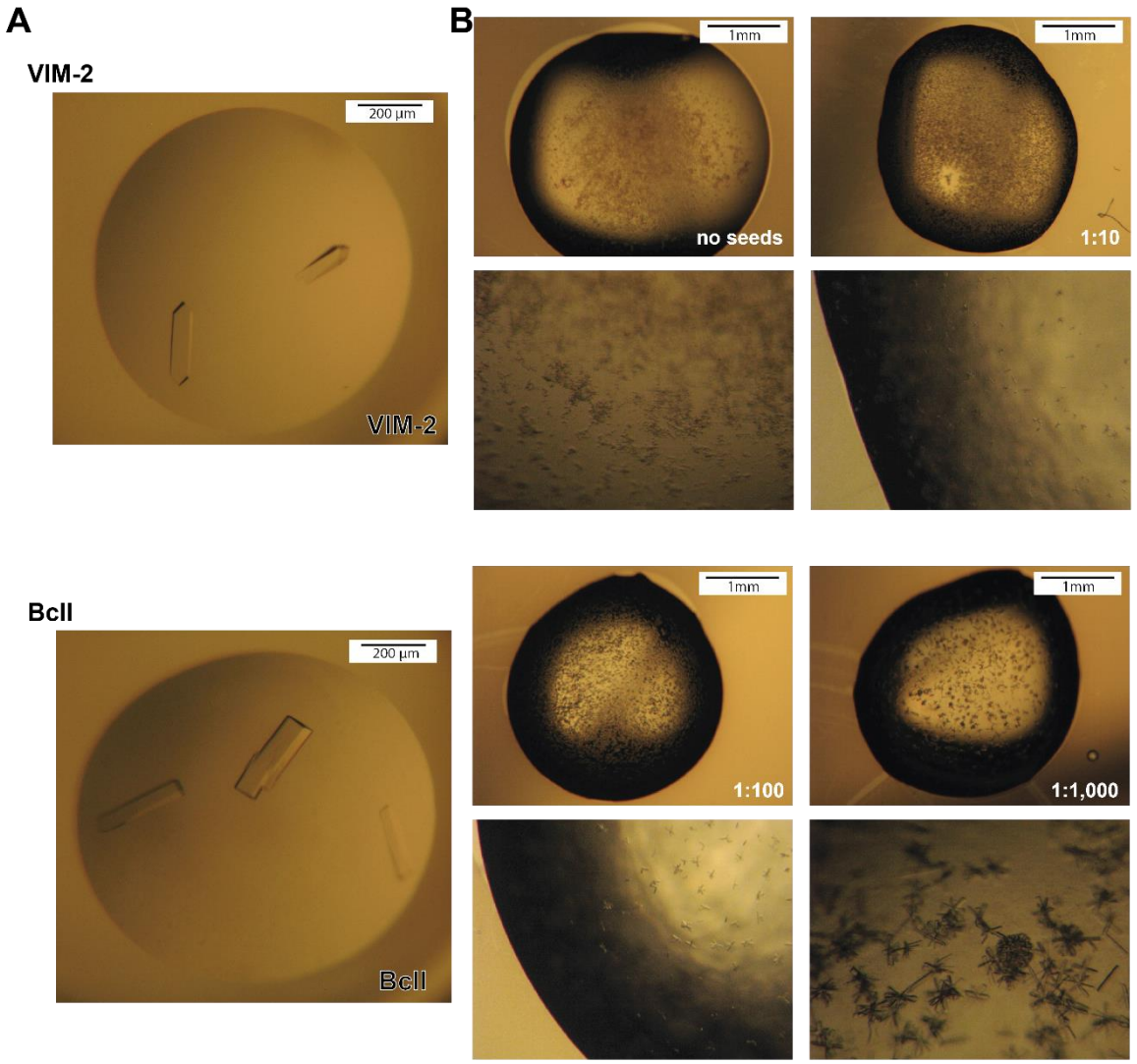


Figure 20 VIM-2 and BcII micro-crystal optimization. A Macro-crystals of VIM-2 (conditions: 22% PEG 3350, 0.25 M magnesium formate) and BcII (conditions: 25% PEG 3350, 0.1 bis-tris pH = 5.5, 0.1 M ammonium sulphate) grown using sitting drop vapour diffusion techniques. **B** Serial dilution of seeds stock of VIM-2 and BcII. Crystals grown in hanging drops with indicated dilutions of seed stocks, VIM-2 conditions 26% PEG 3350, 0.1 M magnesium formate, BcII conditions: 29% PEG 3350, 0.1 Bis-Tris pH = 5.5, 0.1 M ammonium sulphate).

8.7.3 Batch optimisation of XFEL proteins

Initial trials of batch crystallisation of AlkB and FIH did not yield microcrystals (see above). Large scale vapour diffusion techniques were tested but did not yield the desired microcrystals. Hanging drops with added seed stock gave good control over crystal size and density, but were very laborious, especially when large quantities of micro-crystal slurries are needed, for example with the liquid jet injectors, used at the European XFEL (see above). Additionally, batch variations were observed between different plates of AlkB crystals (see above). Following from Mahon et al[57] describing agitating batch crystallisation conditions using plate shakers, it was tested if shaking of small batches would indeed facilitate micro-crystal formation of FIH, VIM-2 or BcII.

In initial trials, small batches of FIH (2 µL) were added to equal volume of precipitant solution in a 24 sitting well plate using the standard FIH optimisation conditions yielding under gentle agitation (using a plate shaker at 700 rpm) micro-crystals in all conditions tested within 24 h. To ensure homogeneity of the FIH crystal slurry, a single condition was chosen (1.9 M ammonium sulphate, AMS, 6% PEG400, 50 mM Tris-HCl pH = 7.5) for batch optimisation based on the appearance of micro crystals in a 24-well plate hanging drop optimisation plate.

With optimised parameters including protein stock concentration, agitation of the crystallisation trials and optimised precipitant solution, in 96 well plates containing 100 µL batches the optimised ratio of precipitant to protein was determined to be a protein to precipitant ratio of 1:5 (v/v). Finally, it was tested what effect seeding the batch with FIH seeds has. A dilution series of FIH under the above described conditions was tested but no effect of the seed stock was observed. The crystal density, size and morphology were comparable or all seed stock dilutions, judging by visual inspection.

In parallel with the optimisation of FIH micro-crystal batches, analogously, small batch optimisation for IPNS and VIM-2 were pursued. For both proteins, conditions from hanging drop and sitting drop plates respectively were chosen for batch optimisation.

IPNS crystallises in at least two different crystal forms under given optimisation conditions (see above). The first crystal from is orthorhombic ($P2_12_12_1$ space group) with a plate morphology which was reported to diffract to ~1.2 Å under cryo-conditions with substrate, using synchrotron radiation[45]. The second crystal form is trigonal ($P3_121$ space group) diffracting to ~1.6 Å under cryo-conditions using synchrotron radiation (see above). The orthorhombic crystal form is favoured for a higher resolution, whereas higher symmetry and the crystal morphology are favoured in the trigonal crystal form. To distinguish the crystal forms, macro-crystals from both crystal forms were harvested separately and seed

stocks from both prepared. Initial batch optimisation exclusively yielded IPNS crystals in the form of small needles (see Figure 20) which were identified to be the orthorhombic crystal form. Due to limited ACV substrate and protein, as well as limited seed stock availability of the trigonal crystal form, further optimisation was only pursued for the orthorhombic form (Dr Patrick Rabe, University of Oxford). The use of highly concentrated seed stocks appears to be beneficial for high crystal density in the batches. The needle size could, to some degree, be influenced by the speed of the plate shaker. Higher shaking intensities appear to favour smaller crystals but in the case of IPNS, the needles got shorter and thinner leading to overall worse quality. Additionally, at higher shaking speed resulted in higher levels of protein precipitation (judged by visual inspection). Attempts to break the needles into smaller fragments using seed beads lead to inhomogeneous samples of poor quality (Dr Patrick Rabe, University of Oxford).

Microcrystals with needle morphology can challenge the sample delivery systems of serial crystallographic instruments as they have a high tendency to clog the capillaries and mixers. For future studies optimisation of the trigonal crystal form should also be considered.

Initial experiments demonstrated that the size and quantity of VIM-2 crystals can be regulated by the amount of seed stock present in the crystallisation conditions under hanging drop optimisation (see above). Following from these observations, the same trend was observed for batch optimisation. Higher crystal concentrations were observed in wells with higher seed supplementation.

VIM-2 crystals grew mostly as single rod-like crystals, with a similar morphology compared to the macro-crystals (Figure 20). A large population, however, also grew in the form of small clusters (comparable to the observed crystals in hanging drops shown in Figure 20B). Small micro-crystal clusters don't interfere with XFEL measurements, as in many cases the beam is very focused and the chance of hitting more than one crystal-lattice out of a cluster is very low. A larger beam size, as used at the European XFEL (~10 µm due to technical problems), can be challenging as multiple crystal lattices are hit simultaneously. Nonetheless, several crystal lattices can be indexed by dials.stills in case of multiple lattices measured at once.

BcII started to crystallise during the setup of the crystallisation plates, leading to a cluster of crystals and possible precipitate (data not shown), thus challenging the batch optimisation of homogenous samples which is required for measurements. To avoid premature crystallisation, BcII was added to the precipitant solution in a 1:20 (v/v) ratio under mild-vortexing. The crystallisation mixture turned immediately cloudy, presumably

from micro-crystal formation, but the sample was incubated for further 2 h to yield best results (an overview of batch crystallization conditions is shown in Table 7).

Table 7 Batch crystallization conditions for FIH, VIM-2, BcII and IPNS.

	FIH	VIM-2	BcII	IPNS
concentration	0.3 mM	0.35 mM	2.0 mM	1.4 mM
substrate	HIF1a (1.5 mM)	n/a	n/a	ACV (2.7mM)
cofactor	Fe (0.9mM)	n/a	n/a	FE (5mM)
cosubstrate	2OG (1.8 mM)	n/a	n/a	n/a
seeding required	no	yes	yes*	yes
shaking speed	700 rpm	500-600 rpm	500-600 rpm	700 rpm
protein/precipitant	1:5	1:2	1:10	1:13
time	24-48 h	48 h	2 h**	48 h
precipitant solution	0.1 M HEPES	0.1 M Mg formate	0.1 M BisTris pH = 5.5	115 mM Tris HCl pH = 8.5
	1.9M AMS	26% PEG 3350	29% PEG 3350	1.8 M LiSO4
	6% PEG400		0.1 M AMS	

* BcII crystals appeared without seed but overall crystal quality was worse.
** BcII crystal appeared within 2 h but continued to grow for 48 h.

8.7.4 Pre-screening of micro-crystals from FIH, IPNS, VIM-2 and BcII at DLS I-24

Evaluation of the quality of micro-crystal slurries can be challenging using light microscopy due to the size of the crystals. The recognition of distinct crystal morphologies is limited by the microscope's resolution which can make the distinguishing between crystals and precipitate challenging.

To evaluate the quality of crystals grown in batch methods, samples of FIH, IPNS, VIM-2, and BcII were screened at room temperature using thin film sandwich sample mounts at Diamond I-24 beamline. FIH and IPNS crystals batches grew at crystal densities of ~10^7-10^8 crystals/mL. Snapshots from the data collection are shown in Figure 21.

IPNS crystallised exclusively in $P2_12_12_1$ space group with a needle morphology at a size of up to 3x3x30 µm (Figure 21B). Diffraction was observed to >3 Å with 100% transmission, 50 ms exposure time and a 9x6 µm beam. Data processing using DIALS gave the cell constants: of a= 42.0 Å, b= 76.6 Å, c= 103.0 Å, close to previously reported cell constants for IPNS under cryo-conditions[45,58].

VIM-2 crystals grew in small clusters of ~10 µm (Figure 21C). The crystal density was much lower than observed for FIH and IPNS. Diffraction was observed to ~3.4 Å with 100% transmission and 100 ms exposure. The crystals decayed quickly under the exposure and the spots could not be indexed using DIALS, XDS or MOSFLM.

BcII grew in very low crystal density. Most of the crystals grew in small clusters (~5 µm) and did not diffract well. A small fraction of the crystal population grew to small rod like crystals of ~5x15 µm (Figure 21D) which diffracted to ~2.6 Å using 100% transmission and 50 ms exposure. Spots were indexed using MOSFLM in a *C*2 space group with the cell constants of a= 54.1 Å, b= 62.3 Å, c= 70.4 Å, and α= 90 °, β= 93.4 °, γ= 90 °.

FIH crystals grew to an average size of ~10 µm diameter (Figure 21A) in irregular shape, although some bipyramidal crystals could be observed. Using 70% transmission and a 9x6 µm beam with 10 ms exposure, diffraction to ~3 Å was observed.

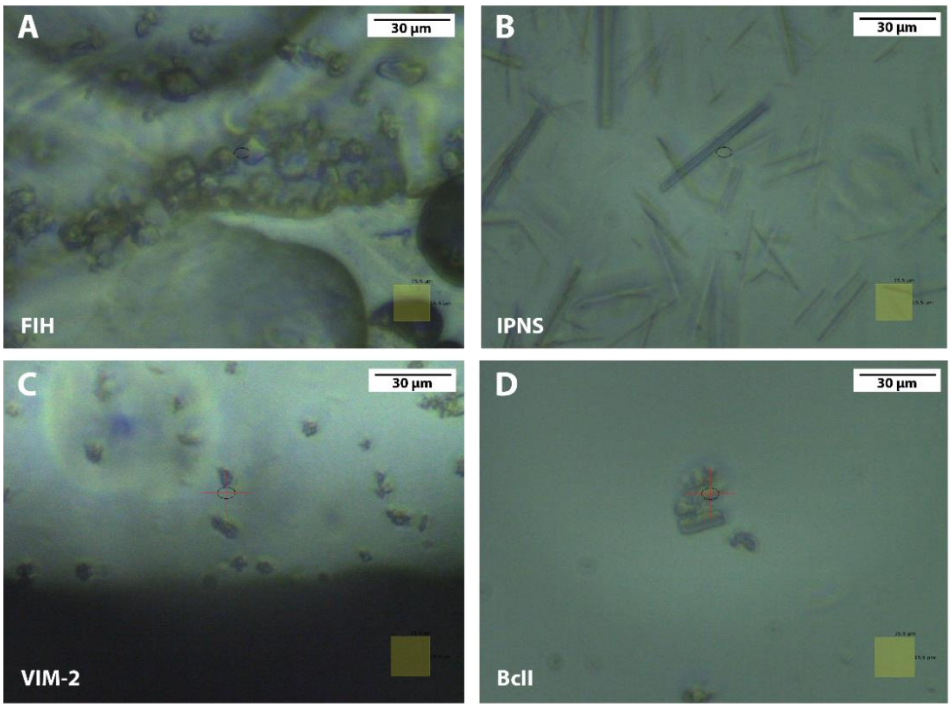


Figure 21 Images from micro crystals screened at Diamond I-24 beamline. Snapshots from data collection of **A** FIH, **B** IPNS, **C** VIM-2 and **D** BcII at DLS I-24. Pictures are on courtesy of Dr Pierre Aller and Dr Agata Butryn (Diamond Light Source).

8.7.5 Data collection of VIM-2 at the European XFEL

Data collection at the European XFEL was hampered by technical problems. Micro-crystal batches for VIM-2 and BcII were prepared in collaboration with Dr Jurgen Brem (University of Oxford) but no FIH micro-crystal batches were prepared due to beam-time limitations. During the run, only data for VIM-2 in the ground state could be collected. Using a Cornell liquid jet system (Lois Pollack and co-workers, Cornell University) at a flow rate of 20 µL/min mixed with a sheath liquid flow rate of 20 µL/min. In total ~3,500 images were collected over ~90 min of data collection. Diffraction was observed >2 Å upon visual inspection. Data analysis and processing is on-going.

8.8 Conclusions and outlook

Structural characterisations of reactive intermediates of 2OG oxygenases and related proteins remain a major research goal in the field. Using time-resolved crystallography at an XFEL source offers a unique opportunity to follow enzymatic reactions. As work towards the application of XFEL for studying 2OG oxygenase and MBL catalysis, reliable and robust micro-crystallisation protocols were developed for FIH, AlkB and IPNS as well as the metallo-β-lactamases VIM-2 and BcII suitable for time-resolved crystallography at an XFEL.

Using optimised micro-crystallisation conditions, the first structure of a 2OG oxygenase in an active substrate complex at room-temperature, was solved, namely FIH, at the LCLS. This was followed by a room temperature resting state XFEL crystal structure of AlkB. Further it was shown that both enzymes, FIH and AkB, can turn over their substrates in the crystal state. Comparison of crystal structures collected from serial crystallography at room-temperature with corresponding structures from cryo-cooled sweep-data collection, at a synchrotron, manifest conserved substrate binding modes for FIH and AlkB between cryo-crystallography and room-temperature data collection. Only local differences in side-chain orientations of HIF1-α, bound to FIH, away from the active site were observed.

At the European XFEL, using a rapid-mixing liquid jet injector system, data for VIM-2 was collected, the first XFEL data for a metallo-β-lactamase. VIM-2 is also one of the first proteins ever to be measured at the European XFEL and was shown to diffract beyond 2 Å resolution. Due to the novelty of the system, no data processing pipeline is developed to date and the VIM-2 data processing is ongoing.

Progress in sample preparation, especially reliable batch crystallisation, should enable the production of large amounts of micro-crystal slurries for further XFEL experiments of 2OG oxygenases and related enzymes, as well as MBLs. Anoxic ground state and oxygen exposed crystal structures of FIH and AlkB will lay the foundation for future time-resolved XFEL studies on 2OG oxygenases.

8.9 References

1 Ihee, H. *et al.* Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proc Natl Acad Sci U S A* **102**, 7145-7150, doi:10.1073/pnas.0409035102 (2005).

2 Pande, K. *et al.* Femtosecond structural dynamics drives the trans/cis isomerization in photoactive yellow protein. *Science* **352**, 725-729, doi:10.1126/science.aad5081 (2016).

3 Schmidt, M., Šrajer, V., Purwar, N. & Tripathi, S. The kinetic dose limit in room-temperature time-resolved macromolecular crystallography. *J Synchrotron Radiat* **19**, 264-273, doi:10.1107/S090904951105549X (2012).

4 Key, J., Srajer, V., Pahl, R. & Moffat, K. Time-resolved crystallographic studies of the heme domain of the oxygen sensor FixL: structural dynamics of ligand rebinding and their relation to signal transduction. *Biochemistry* **46**, 4706-4715, doi:10.1021/bi700043c (2007).

5 Schotte, F. *et al.* Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science* **300**, 1944-1947, doi:10.1126/science.1078797 (2003).

6 Srajer, V. *et al.* Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science* **274**, 1726-1729 (1996).

7 Miller, R. J. Femtosecond crystallography with ultrabright electrons and x-rays: capturing chemistry in action. *Science* **343**, 1108-1116, doi:10.1126/science.1248488 (2014).

8 Neutze, R. & Moffat, K. Time-resolved structural studies at synchrotrons and X-ray free electron lasers: opportunities and challenges. *Curr Opin Struct Biol* **22**, 651-659, doi:10.1016/j.sbi.2012.08.006 (2012).

9 Cammarata, M. *et al.* Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering. *Nat Methods* **5**, 881-886, doi:10.1038/nmeth.1255 (2008).

10 Martin, J. L. & Vos, M. H. Femtosecond biology. *Annu Rev Biophys Biomol Struct* **21**, 199-222, doi:10.1146/annurev.bb.21.060192.001215 (1992).

11 Gai, F., Hasson, K. C., McDonald, J. C. & Anfinrud, P. A. Chemical dynamics in proteins: the photoisomerization of retinal in bacteriorhodopsin. *Science* **279**, 1886-1891 (1998).

12 Wang, H. *et al.* Protein dynamics control the kinetics of initial electron transfer in photosynthesis. *Science* **316**, 747-750, doi:10.1126/science.1140030 (2007).

13 Johansson, L. C., Stauch, B., Ishchenko, A. & Cherezov, V. A Bright Future for Serial Femtosecond Crystallography with XFELs. *Trends Biochem Sci* **42**, 749-762, doi:10.1016/j.tibs.2017.06.007 (2017).

14 Garman, E. Cool data: quantity AND quality. *Acta Crystallogr D Biol Crystallogr* **55**, 1641-1653 (1999).

15 Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc Natl Acad Sci U S A* **108**, 16247-16252, doi:10.1073/pnas.1111325108 (2011).

16 Keedy, D. A. *et al.* Crystal cryocooling distorts conformational heterogeneity in a model Michaelis complex of DHFR. *Structure* **22**, 899-910, doi:10.1016/j.str.2014.04.016 (2014).

17 Mertens, H. D. T. & Svergun, D. I. Combining NMR and small angle X-ray scattering for the study of biomolecular structure and dynamics. *Arch Biochem Biophys* **628**, 33-41, doi:10.1016/j.abb.2017.05.005 (2017).

18 Lisi, G. P. & Loria, J. P. Solution NMR Spectroscopy for the Study of Enzyme Allostery. *Chem Rev* **116**, 6323-6369, doi:10.1021/acs.chemrev.5b00541 (2016).

19 Kirby, N. M. & Cowieson, N. P. Time-resolved studies of dynamic biomolecules using small angle X-ray scattering. *Curr Opin Struct Biol* **28**, 41-46, doi:10.1016/j.sbi.2014.07.007 (2014).

20 Akiyama, S. *et al.* Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. *Proc Natl Acad Sci U S A* **99**, 1329-1334, doi:10.1073/pnas.012458999 (2002).

21 Bujalowski, W. M. & Jezewska, M. J. Fluorescence intensity, anisotropy, and transient dynamic quenching stopped-flow kinetics. *Methods Mol Biol* **875**, 105-133, doi:10.1007/978-1-61779-806-1_6 (2012).

22 Kern, J. *et al.* Simultaneous femtosecond X-ray spectroscopy and diffraction of photosystem II at room temperature. *Science* **340**, 491-495, doi:10.1126/science.1234273 (2013).

23 Stagno, J. R. *et al.* Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242-246, doi:10.1038/nature20599 (2017).

24 Kupitz, C. *et al.* Structural enzymology using X-ray free electron lasers. *Struct Dyn* **4**, 044003, doi:10.1063/1.4972069 (2017).

25 Spence, J. C. H. XFELs for structure and dynamics in biology. *IUCrJ* **4**, 322-339, doi:10.1107/S2052252517005760 (2017).

26 Aquila, A. *et al.* Time-resolved protein nanocrystallography using an X-ray free-electron laser. *Opt Express* **20**, 2706-2716 (2012).

27 Nango, E. *et al.* A three-dimensional movie of structural changes in bacteriorhodopsin. *Science* **354**, 1552-1557, doi:10.1126/science.aah3497 (2016).

28 Schlichting, I. Serial femtosecond crystallography: the first five years. *IUCrJ* **2**, 246-255, doi:10.1107/S205225251402702X (2015).

29 Cheng, R. K. Y., Abela, R. & Hennig, M. X-ray free electron laser: opportunities for drug discovery. *Essays Biochem* **61**, 529-542, doi:10.1042/EBC20170031 (2017).

30 Barty, A. *et al.* Self-terminating diffraction gates femtosecond X-ray nanocrystallography measurements. *Nat Photonics* **6**, 35-40, doi:10.1038/nphoton.2011.297 (2012).

31 Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752-757, doi:10.1038/35021099 (2000).

32 Chapman, H. N. *et al.* Femtosecond X-ray protein nanocrystallography. *Nature* **470**, 73-77, doi:10.1038/nature09750 (2011).

33 Kupitz, C. *et al.* Microcrystallization techniques for serial femtosecond crystallography using photosystem II from Thermosynechococcus elongatus as a model system. *Philos Trans R Soc Lond B Biol Sci* **369**, 20130316, doi:10.1098/rstb.2013.0316 (2014).

34 Sierra, R. G. *et al.* Nanoflow electrospinning serial femtosecond crystallography. *Acta Crystallogr D Biol Crystallogr* **68**, 1584-1587, doi:10.1107/S0907444912038152 (2012).

35 Oberthuer, D. *et al.* Double-flow focused liquid injector for efficient serial femtosecond crystallography. *Sci Rep* **7**, 44628, doi:10.1038/srep44628 (2017).

36 Weierstall, U. *et al.* Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nat Commun* **5**, 3309, doi:10.1038/ncomms4309 (2014).

37 Sugahara, M. *et al.* Grease matrix as a versatile carrier of proteins for serial crystallography. *Nat Methods* **12**, 61-63, doi:10.1038/nmeth.3172 (2015).

38 Hunter, M. S. *et al.* Fixed-target protein serial microcrystallography with an x-ray free electron laser. *Sci Rep* **4**, 6026, doi:10.1038/srep06026 (2014).

39 Baxter, E. L. *et al.* High-density grids for efficient data collection from multiple crystals. *Acta Crystallogr D Struct Biol* **72**, 2-11, doi:10.1107/S2059798315020847 (2016).

40 Roedig, P. *et al.* Room-temperature macromolecular crystallography using a micro-patterned silicon chip with minimal background scattering. *J Appl Crystallogr* **49**, 968-975, doi:10.1107/S1600576716006348 (2016).

41 Oghbaey, S. *et al.* Fixed target combined with spectral mapping: approaching 100% hit rates for serial crystallography. *Acta Crystallogr D Struct Biol* **72**, 944-955, doi:10.1107/S2059798316010834 (2016).

42 Fuller, F. D. *et al.* Drop-on-demand sample delivery for studying biocatalysts in action at X-ray free-electron lasers. *Nat Methods* **14**, 443-449, doi:10.1038/nmeth.4195 (2017).

43 Tarhonskaya, H. *et al.* Kinetic Investigations of the Role of Factor Inhibiting Hypoxia-Inducible Factor (FIH) as an Oxygen Sensor. *J Biol Chem*, doi:10.1074/jbc.M115.653014 (2015).

44 Tamanaha, E. *et al.* Spectroscopic Evidence for the Two C-H-Cleaving Intermediates of Aspergillus nidulans Isopenicillin N Synthase. *J Am Chem Soc* **138**, 8862-8874, doi:10.1021/jacs.6b04065 (2016).

45 Roach, P. L. *et al.* Anaerobic crystallisation of an isopenicillin N synthase.Fe(II).substrate complex demonstrated by X-ray studies. *Eur J Biochem* **242**, 736-740 (1996).

46 Burzlaff, N. I. *et al.* The reaction cycle of isopenicillin N synthase observed by X-ray diffraction. *Nature* **401**, 721-724, doi:10.1038/44400 (1999).

47 Yu, B. *et al.* Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. *Nature* **439**, 879-884, doi:10.1038/nature04561 (2006).

48 Hopkinson, R. J. *et al.* 5-Carboxy-8-hydroxyquinoline is a Broad Spectrum 2-Oxoglutarate Oxygenase Inhibitor which Causes Iron Translocation. *Chem Sci* **4**, 3110-3117, doi:10.1039/C3SC51122G (2013).

49 Kabsch, W. XDS. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132, doi:10.1107/S0907444909047337 (2010).

50 Evans, P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**, 72-82, doi:10.1107/S0907444905036693 (2006).

51 Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**, 235-242, doi:10.1107/S0907444910045749 (2011).

52 Elkins, J. M. *et al.* Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem* **278**, 1802-1806, doi:10.1074/jbc.C200644200 (2003).

53 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674, doi:10.1107/S0021889807021206 (2007).

54 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).

55 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, doi:10.1107/S0907444909052925 (2010).

56 Bonomo, R. A. *et al.* Carbapenemase-Producing Organisms: A Global Scourge! *Clin Infect Dis*, doi:10.1093/cid/cix893 (2017).

57 Mahon, B. P. *et al.* Microbatch Mixing: “Shaken not Stirred”, a Method for Macromolecular Microcrystal Production for Serial Crystallography. *Crystal Growth & Design* **16**, 6214-6221, doi:10.1021/acs.cgd.6b00643 (2016).

58 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, doi:10.1038/375700a0 (1995).

9 Summary

Structural studies on FIH

The crystallographic results presented in this thesis demonstrate that FIH is a tractable and useful model 2OG oxygenase for structural studies involving inhibitors. Crystal structures of FIH in complex with Molidustat, GSK1278863 and Vadadustat, PHD inhibitors that are currently in clinical trials for anaemia treatment, were solved. Together with inhibition and NMR binding studies in solution, the crystallographic data on the binding modes of the “clinical” PHD inhibitors helps to define their mechanisms of action. The results reveal that the clinical candidates inhibit PHDs *via* 2OG displacement with varying degrees of substrate competition.

For the crystallographic analysis of a crystal structure of FIH in complex with Vadadustat, Pan-Dataset Density Analysis (PANDDA) was used to investigate the multiple conformations of a single bound ligand. The use of PANDDA will potentially help in elucidating the binding mode of small molecule ligands binding in multiple conformations, for both FIH and other enzymes.

In addition to the “clinical” PHD inhibitors, several structures of FIH in complex with non-clinical PHD inhibitors were solved; these revealed novel binding modes of new types of FIH inhibitors. The identified binding modes of small molecules interacting with FIH could be of use in the development of selective FIH inhibitors in ongoing work.

For the identification of small molecules binding in potential allosteric binding sites of FIH, a workflow for a high-throughput crystallographic fragment screen was optimised and several libraries were screened for binding to FIH in the crystalline state. The results of a crystallographic fragment screening with FIH revealed several compounds that bind to the active site and substrate binding site of FIH. One fragment, identified in the fragment screen was found to bind to the active site of FIH and inhibited its activity with a low µM IC_{50}. This fragment is currently being optimised with the goal of producing a potent, FIH inhibitor which is active in cells. Further, data from the fragment screen could lay the foundation for the development of substrate specific inhibitors of FIH.

In a proof-of-principle approach, in-crystal reactions of boronic acids with hydrazines and thiol-amino compounds were tested for FIH, and a novel class of compounds interacting with FIH based on cyclic boronates was found. Investigations on the binding modes of cyclic boronates to FIH and potential implication of boronate binding on FIH on its activity are currently ongoing.

New substrates for FIH

FIH can accept, in addition the HIF-1/2α transcription factors, a large range of ARD proteins as substrates; these include a subset of the TRP ion channels. TRPA1 is a member of the TRP ion channels and was identified as new FIH substrate. Aspargine-336 of TRPA1 was found to be hydroxylated in an FIH dependent manner.

In contrast to the considerable number of FIH substrates identified, only two ankyrin substrates have been structurally characterised in complex with FIH.

To investigate potential differences in the binding mode of peptides derived from ankyrin peptides and HIF-1/2α, 6 crystal structures of FIH in complex with peptides derived from TRPA1, TRPV3, and one peptide derived from HIF-2α were solved. Crystal structures of FIH in complex with TRPV3 and TRPA1 manifest differences in their binding mode to FIH compared to HIF-1/2α. Differences in HIF-1/2α and ankyrin substrate binding to FIH could potentially be exploited in the development of substrate specific FIH inhibitors.

ARD proteins have to unfold, at least partially, in order to bind to FIH. Very little is known about the mechanism of unfolding and potential “long-range” interactions between (partially) unfolded ARD proteins and FIH. To investigate potential long-range interactions of FIH-ARD proteins, crystallisation attempts of FIH in complex with 4 ARD proteins, 3CA, mNotch1, iASPP, and ASPP2, were pursued. Crystallisation trials with co-purified FIH-ASPP2 or FIH-iASPP complexes were unsuccessful, possibly due to the necessary unfolding of ARDs leading to disorder in the ARD fold upon FIH binding. *In situ* complex formation FIH complexes with 3CA, iASPP, and ASPP2 lead to crystallisation of 3CA. Attempts to crystallise the FIH-ASPP2 complex by reducing potential flexible parts by limited proteolysis did not yield crystals suitable for X-ray crystallography.

For stabilisation of an FIH-ARD complex, a disulfide cross-linking approach for FIH complexes with 3CA, mNotch1, iASPP, and ASPP2, was investigated. Crystal structures of FIH in complex with a 3CA fragment cross-linked to FIH, *via* an engineered disulphide bond, and an mNotch1 fragment cross-linked to FIH, *via* an engineered non covalent heavy metal mediated cross-link, were solved and showed evidence for proteolytic cleavage of the ARD substrate. The potential mechanisms of the proteolytic cleavage of 3CA and mNotch1 however, remain unclear.

FIH shows remarkable substrate promiscuity. *In vivo*, FIH can potentially accept 100s of ARD proteins as substrates. FIH can also hydroxylate at least three different amino acid residues. For all FIH substrates identified to date, the modified amino acid is located in a conserved location in the ankyrin.

Studies on FIH-mediated hydroxylation of ASPP1, ASPP2 and iASPP proteins showed evidence for an FIH mediated hydroxylation of two asparagines within the same ankyrin repeat. FIH appears to show a preference for a “VNVN” motif for double hydroxylation. In the addition to the ASPP proteins, evidence for FIH mediated double hydroxylaiton was found for the “VNVN” ARD proteins p18INK4C and ASB11. Double hydroxylation was also not observed at the peptide level for ASPP derived peptides and consensus peptides bearing a “VNVN” or related motif, suggesting the requirement of non-active site enzyme-substrate interactions for double hydroxylation.

To investigate the potential effects of FIH mediated ASPP2 (di)hydroxylation, approaches for studying the interactome of ASPP2 using label free quantitative mass-spectrometry were established for overexpressed and endogenous ASPP2 proteins. Significant differences in the interactomes of overexpressed ASPP2 WT and a protein phosphatase 1 binding deficient variant were observed in U2OS cells. Changes were observed for the abundance of proteins interacting with ASPP2 WT and a PP1 binding deficient variant as well as in the levels of ASPP2 phosphorylation between the two samples: Two proteins, KIF1C and PRRC2A were found to bind exclusively to a PP1 binding deficient variant, whereas ZBTB2 and CCDC85C appear to be enriched in the ASPP2 WT sample. Phosphorylation of serine-414, serine-714, and serine-737 of ASPP2 appeared to be enriched in the PP1 binding deficient variant compared to ASPP2 WT, potentially resulting from an auto-dephosphorylation mechanism.

Investigations of ASPP2 interactions in U2OS cells in the presence and absence of FIH showed no evidence for significant changes of the interactome of ASPP2 using overexpressed ASPP2 protein and at endogenous ASPP2 protein levels. This may suggest that, at least under tested conditions, FIH-mediated hydroxylation does not influence the protein-protein interactions of ASPP2 and thereby its function.

XFEL crystallography

Methodologies for the growth of homogenous micro-crystal slurries of FIH, AlkB, VIM-2, BcII and IPNS, suitable for XFEL measurements and rapid mixing experiments at an XFEL were presented.

Data for an FIH.Fe(II).2OG.HIF-1α complex were collected at the LCLS XFEL source. The data were processed and the first XFEL derived structure of a 2OG oxygenase, with bound substrate and cofactors in its anaerobic resting state at room temperature, was solved. Comparison of a room-temperature crystal structure of an FIH.Fe(II).2OG.HIF-1α complex collected at the LCLS XFEL source with a crystal structure solved from cryo-crystallography data from Diamond Light Source of the same complex, manifests a

conserved binding mode of the substrate and cofactors. Notably, some alternative sidechain conformations were observed for the HIF substrate at room-temperature compared to the cryo-crystallography data.

Exposing crystals from the FIH.Fe(II).2OG.HIF-1α complex to oxygen lead to apparent substrate turnover and subsequent ejection of the substrate. Low occupancy of the HIF-α substrate could only be observed in the case of oxygen exposed micro-crystal slurries of the FIH complex measured at the LCLS and not in the corresponding structures from cryo-crystallography of macro-crystals collected at Diamond Light Source.

High-resolution crystal structures of an AlkB.Fe(II).2OG.T-mA-T complex, using cryo-crystallography, exposed to varying length of ambient oxygen levels, showed distinct states of the catalytic cycle. Preliminary room-temperature data from XFEL measurements of AlkB.Fe(II).2OG.T-mA-T complexes showed a conserved substrate binding between the cryo-cooled and room-temperature data collection.

Using optimised VIM-2 crystallisation conditions, large quantities of micro-crystals of the metallo-β-lactamase VIM-2 were produced and tested at the European XFEL; VIM-2 is one of the first enzymes ever to be measured at the European XFEL. Using a liquid-jet rapid mixing nozzle, VIM-2 showed X-ray diffracted to a resolution beyond 2 Å showing that, a least in principal, VIM-2 microcrystals are compatible with the rapid mixing for serial femtosecond crystallography (SFX).

Overall the results inform on the active site substrate binding modes of FIH.

Despite considerable effort, it was not possible to obtain a structure of an intact FIH-ARD complex. It seems likely that: due to the fact that the ARD must substantially unfold in order to bind productively at the FIH active site.

XFEL mediated crystallography has also been shown to be a promising technique for studying FIH mediated catalysis. Future studies could include using substrate/intermediate trapping techniques, *e.g.* the disulfide cross-linking methodology developed, to enable the use of XFEL crystallography to trap transient intermediates in FIH catalysis.

10 Methods and materials

10.1 General

All chemicals and reagents were from Sigma and used without further purification unless otherwise stated. Oligonucleotides were from Sigma and Eurofins. Molecular biology enzymes and buffers were from New England Biolabs. All DNA plasmids were sequenced prior to use (a list of sequencing primers used is shown in Table 1). Size exclusion chromatography (SEC) was performed using Superdex 75 and Superdex 200 columns (GE Healthcare). Affinity chromatography was carried out using His-Trap (GE Healthcare) and Glutathione Sepharose Fast Flow resin (Amersham Bioscience). Ion exchange chromatography was carried out using SP Sepharose fast flow column, 50 mL (GE Healthcare) on Äkta FPLC systems (GE Healthcare).

Table 1 Sequencing primers used.

Gene/Vector	Name	Sequence
iASPP	iASPPseq1F	GAT GGC GCA GGC AGC TC
	iASPPseq2F	CTC CTC TGG CCA GCG AC
	iASPPseq3F	CTG TAG CAA GGC CTC TC
	iASPPseq4F	GAG ATG AAC GAC CCG AG
ASPP1	ASPP1_seq2	GCT CCT GTT GCT TTA GAA AAT G
	ASPP1_seq3	CAA TAC TGA GAG ACT GGG G
	ASPP1_seq4	GTA CAT GGA GTA TAT GGA ACT TG
	ASPP1_seq4	CAT TGG CAT CTG ATG ACG G
	ASPP1_seq6	TGT AGC CTT CCT CCA TCT C
diverse	T7_f*	TAA TAC GAC TCA CTA TAG GG
	T7_R*	GCT AGT TAT TGC TCA GCG G
	pGEX5'*	GGG CTG GCA AGC CAC GTT TGG TG
	pGE3'*	GGT CGT CAG ACT GTC GAT GAA GCC
	BGH_R*	TAG AAG GCA CAG TCG AGG
	pCMV_F*	GAG CTC GTT TAG TGA ACC GTC
	M13R*	TGT AAA ACG ACG GCC AGT
	pBADR*	GAT TTA ATC TGT ATC AGG
	T3*	ATT AAC CCT CAC TAA AG
	SP6*	ATT TAG GTG ACA CTA TAG
	pEF_f	TTC AGG TGT CGT GAG GAA TTA GC
	pEF_r	GGA ACC TTT AAT AGA AAT TGG ACA G

10.2 Microbiological techniques

10.2.1 Plasmid DNA isolation

Plasmid DNA purification was accomplished using GeneJET Miniprep Kit (Thermo Scientific), QIAprep Spin Miniprep Kit (Qiagen) or GeneElute HP Plasmid Maxiprep Kit (Sigma-Aldrich) according to the manufacturers' protocols. Purified plasmids were eluted using sterile water.

10.2.2 Site directed mutagenesis

Site directed mutagenesis was performed according to an adapted QuickChange protocol (Agilent Technologies) using Q5 and *Pfu* Turbo DNA Polymerase according the manufacturers' protocol. The primers used are listed in Table 3. A general PCR protocol is shown in Table 2.

Table 2 PCR Protocol for Site directed mutagenesis.

Step	Temp	time
Initial denaturing	95-98°	3 min
Denaturing	95-98°	10 sec
Annealing	62-65°	30 sec
Extension	68-72°	6-10 min
Final extension	68-72°	10 min
Final hold	4°	∞

Table 3 Primers used for site-directed mutagenesis.

mutation	sequence
FIH_G299C_f	GAA CTT CTG GTA TAA GTG CGC TCC CAC CCC TAA G
FIH_G299C_r	CTT AGG GGT GGG AGC GCA CTT ATA CCA GAA GTT C
mNotch_A1940C_f	AAG CGC TTG CTG GAG TGC AGT GCA GAT GCC AAC
mNotch_A1940C_r	GTT GGC ATC TGC ACT GCA CTC CAG CAA GCG CTT
3CA_A79C_f	AAA CTG CTG CTG GAA TGC GGT GCG GAT GTG AAT G
3CA_A79C_r	CAT TCA CAT CCG CAC CGC ATT CCA GCA GCA GTT T
pET28a_Thrombin 2 TEV_f	CAT CAT CAT CAC AGC AGC GAG AAC CTG TAC TTC CAG AGC CAT ATG GCT AGC ATG
pET28a_Thrombin 2 TEV_r	CAT GCT AGC CAT ATG GCT CTG GAA GTA CAG GTT CTC GCT GCT GTG ATG ATG ATG
ASPP2_F981C_f	AAG TTC CTG GTA CAG TGT GGT GTA AAT GTA AAT G
ASPP2_F981C_r	CAT TTA CAT TTA CAC CAC ACT GTA CCA GGA ACT T
ASPP2_N984A_f	TAC AGT TTG GTG TAG CTG TAA ATG CTG CTG ATA GTG ATG G
ASPP2_N984A_r	CCA TCA CTA TCA GCA GCA TTT ACA GCT ACA CCA AAC TGT A
ASPP2_N984A_N986A_f	GTT TGG TGT AGC TGT AGC TGC TGC TGA TAG TGA TG
ASPP2_N984A_N986A_r	CAT CAC TAT CAG CAG CAG CTA CAG CTA CAC CAA AC
ASPP2_N986A_f	GTG TAA ATG TAG CTG CTG CTG ATA GTG ATG GAT GGA C
ASPP2_N986A_r	GTC CAT CCA TCA CTA TCA GCA GCA GCT ACA TTT ACA C
iASPP_A682C_Fne	GAT TTC CTC ATC ACC TGT GGT GCC AAT GTC AAC TC
iASPP_A682C_Rne	GAG TTG ACA TTG GCA CCA CAG GTG ATG AGG AAA TC

10.2.3 Agarose gel electrophoresis

Nucleic acids were purified using agarose gel electrophoresis. 0.8-1% agarose (molecular grade, Bioline) was heated in TAE buffer (40 mM Tris-Acetate pH = 8.0, 1 mM EDTA) containing ethidium bromide (1:50,000 dilution). DNA sample were mixed in 4x loading buffer (Fermentas) prior to loading on the gel. Electrophoresis was performed at 80-100 V for 45-60 min in TAE buffer. Gene Ruler 1 kbp (Fermentas) was used as a marker and DNA bands were visualised with UV light (254 nm).

10.2.4 Molecular sub-cloning

Constructs were amplified using PCR analogously to those previously described (section 10.2.2) using Phusion polymerase (NEB) with adjusted extension times (30 sec/1kb), primers used for sub-cloning are listed in Table 4. Amplified inserts were purified using agarose gel electrophoresis; bands were excised and DNA extracted using the gel extraction kit (QIAGEN) according to the manufacturers' protocol. Purified inserts and empty target vector were digested using appropriate restriction enzymes (NEB) and purified via agarose gel extraction (vector backbone) or using the PCR purification kit for the insert (QIAGEN) according to manufacturers' protocol.

Table 4 Primers used for amplification of inserts for molecular sub-cloning.

construct	Name	sequence
ASPP1	ASPP1_883_BamHI_f	CGA CGA GGA TCC ATG AGA GTC CGG TTT AAC CC
	ASPP1_1090_NotI_r	GTT GCG GCC GCT CAG GCG AGT GTT CGC TG
FIH	FIH_d25_Nhe_fwd2	GAG TCA GCT AGC GCC TGG GAT GAA TCC CAG
	FIH_d36_Nhe_fwd2	CAA TCA GCT AGC TTC CCG ACT AGG CCC ATT C
	FIH_BAM_rev2	GCA CGA GGA TCC CTA GTT GTA TCG GCC CTT G

10.2.5 Bacterial growth medium

Bacterial cultures were grown in lysogen broth medium (LB, 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) and 2-tryptone yeast extract medium (2TY, 16 g/L tryptone, 10 g/L yeast extract, 5 g/L NaCl), autoclaved at 121 °C before use. Antibiotic stock solutions (1000x) of ampicillin (50 mg/mL in water), chloramphenicol (30 mg/mL in ethanol) and kanamycin (30 mg/mL in water) were filter-sterilized and stored at -20 °C until use.

10.2.6 Transformations

The plasmid DNA of interest was transformed into α-select electro-competent cells, bronze or silver efficiency (Bioline) or BL21(DE3) cells (Agilent Technologies) according to the manufacturers' protocol and plated on LB agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, 20 g/L agar) supplemented with appropriate antibiotics. Bacterial colonies were cultured overnight at 37 °C and stored at 4 °C.

10.2.7 Glycerol stocks

For long term storage, bacterial cultures were grown to stationary phase and mixed with 1 volume of 2x glycerol buffer (65% glycerol, 0.1 M $MgSO_4$, 45 mM Tris-HCl pH = 7.5) and immediately cooled to -80 °C and stored at that temperature.

Table 5 Bacterial expressions vectors.

Construct	Vector	Source	Additional information
FIH WT	pET28a	Prof. Chris Schofield[1]	thrombin cleavable N-terminal His
FIH WT	pNIC28-BSA4	Dr Adam Hardy, University of Oxford	TEV cleavable N-terminal his tag
FIH G299C	pET28a	*site directed mutagenesis*	TEV cleavable N-terminal his tag
FIH G299C	pNIC28-BSA4	*site directed mutagenesis*	TEV cleavable N-terminal his tag
$FIH_{dN26-349}$	pET28a_TEV	*sub-cloned*	TEV cleavable N-terminal his tag
$FIH_{dN39-349}$	pET28a_TEV	*sub-cloned*	TEV cleavable N-terminal his tag
$ASPP2_{889-1128}$	pET22b(+)	Prof Xin Lu, University of Oxford[2]	non-cleavable C-terminal His tag with and without AVI tag
$ASPP2_{889-1128}$ F981C	pET22b(+)	site directed mutagenesis	non-cleavable C-terminal His tag with and without AVI tag
$ASPP2_{920-1057}$	pNH-TrxT	Dr Adam Hardy, University of Oxford	N-terminal His-Trx tag with 3C cleavage site
$ASPP1_{883-1090}$	PGEX-6P2	*sub-cloned*	N-terminal GST tag with 3C cleavage site
$iASPP_{625-828}$	pET22b(+)	Prof. Xin Lu, University of Oxford[2]	non-cleavable C-terminal His tag with and without AVI tag
$iASPP_{625-828}$ A682C	pET22b(+)	*site directed mutagenesis*	non-cleavable C-terminal His tag with and without AVI tag
$mNotch1_{1899-2106}$	pGEX4-T1	Prof. Chris Schofield[3]	thrombin cleavable GST
$mNotch1_{1899-2106}$ A1940C	pGEX4-T1	*site directed mutagenesis*	thrombin cleavable GST
3 Consensus Ankyrin	pET28a	Dr Adam Hardy, University of Oxford[4]	TEV cleavable N-terminal his tag
3 Consensus Ankyrin 80C	pET28a	*site directed mutagenesis*	TEV cleavable N-terminal his tag
$AlkB_{dN11}$	pET24+	Prof. Chris Schofield[5]	un-tagged
$ASB11_{64-287}$	pNIC-BSA1-ElonginBC	Dr Alex Bullock	TEV cleavable N-terminal his tag
P18-INK4C	PGEX-6P1	Prof. Jane Endicott	N-terminal GST tag with 3C cleavage site

10.3 Recombinant protein production

10.3.1 Large scale recombinant protein production and purification

All proteins were produced in *E. coli* BL21 (DE3) cells. Bacterial cells were transformed with the plasmid of interest (see chapter 10.2.6). A single colony was picked and grown in LB or 2TY medium (100 mL) supplemented with the appropriate antibiotics overnight. The stationary starter culture was diluted 1:100 in fresh medium supplemented with the appropriate antibiotics and grown to an OD_{600}~0.8 at 37 °C prior to induction with isopropyl β-D-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation (4 °C, 10,000 g, 8 min, JA-10 rotor, Beckman Coulter), re-suspended in stated buffers (~5x of the pellet mass), lysed by sonication with added DNaseI from bovine pancreas (Roche) and, unless otherwise stated, cell debris pelleted after lysis (4 °C, 34,000 g, 30 min, JLA16.25 rotor, Beckman Coulter) and the supernatant filtered (0.45 µm filter). Purifications were carried out at 4 °C. Purity and identity of all proteins were verified by SDS PAGE (see section 10.3.2) and denaturing mass spectrometry (see section 10.6).

Human FIH WT, FIH G299C, $FIH_{dN26-349}$, $FIH_{dN39-349}$, mouse GST-$Notch1_{1899-2106}$ and GST-$Notch1_{1899-2106}$ A1940C protein production was induced by the addition of IPTG (final concentration 0.5 mM) for 4 h at 37 °C prior to harvesting *via* centrifugation (10,000 g, 8 min, JA-10 rotor, Beckman Coulter). The cell pellet was stored at -80 °C.

FIH containing pellets were re-suspended in 50 mM HEPES buffer (pH = 7.5, 500 mM NaCl, 5 mM imidazole) with added DNaseI from bovine pancreas (Roche) and lysed by sonication. The lysate was centrifuged (4 °C, 34,000 g, 30 min, JLA1625 rotor, Beckman Coulter) and the supernatant filtered to remove cell debris. His-tagged proteins were immobilized on a His-Trap (5 mL) column, washed with low imidazole buffer (50 mM HEPES pH = 7.5 buffer supplemented with 500 mM NaCl and 20-40 mM imidazole) and eluted with a linear gradient of imidazole (max. 50 mM HEPES pH = 7.5, 500 mM NaCl, 500 mM imidazole). The N-terminal His_6-tag was cleaved by the addition of TEV protease (1:50 ration for FIH from pNIC28-BSA4 vectors) or Thrombin (4 units per mg for FIH from pET28a vectors) overnight at 4 °C. Proteases and cleaved tag were removed *via* SEC using a Superdex S200 column equilibrated with 50 mM Tris-HCl pH = 7.5. Completeness of the cleavage was verified by Mass Spectrometry.

$ASB11_{64-287}$, P18-INK4C and $ASPP1_{883-1090}$ protein production was induced by the addition of IPTG (final concentration 0.5 mM) for 16 h at 18 °C prior to harvesting. The cell pellet was stored at -80 °C.

ASB11 containing pellets were re-suspended in 15 mM HEPES buffer (pH = 7.5, 15 mM imidazole, 500 mM NaCl). Purification was performed analogously to FIH protein purification with the exception of the final purification step over an S75 SEC column equilibrated with 50 mM HEPES pH = 7.5 supplemented with 250 mM NaCl and 0.5 mM tris(2-carboxyethyl)phosphine (TCEP).

P18-INK4C containing pellets were re-suspended in 50 mM (HEPES pH = 7.5, 150 mM NaCl, 2 mM DTT), lysed by sonication and cleared of cell debris by centrifugation and filtration prior to immobilizing the GST-tagged proteins on a Sepharose column. The column was washed with the same buffer (100 mL) prior to elution with 40 mM HEPES (pH = 7.4, 200 mM NaCl, 10 mM reduced glutathione). Fractions containing proteins were pooled, 3C protease (1:50 ratio) added and the GST-tag cleaved over night at 4 °C. The cleaved protein was further purified using a Superdex S75 column (40 mM HEPES pH = 7.4, 200 mM NaCl).

ASPP1 and Notch1 containing pellets were re-suspended in 40 mM HEPES (pH = 7.4, 200 mM NaCl), lysed by sonication, and cleared of cell debris by centrifugation and filtration prior to immobilizing the GST-tagged proteins on a Sepharose column. The column was washed with the same buffer (100 mL) prior to elution with 40 mM HEPES (pH = 7.4, 200 mM NaCl, 10 mM reduced glutathione). Protein containing fractions were combined and buffer exchanged into 40 mM HEPES (pH = 7.4, 200 mM NaCl) using PD-10 desalting columns (GE Healthcare) according to manufacturers' protocol. 3C protease (1:50 ratio) was added for the ASPP1 construct and thrombin (4 units per mg of protein) was added to the mNotch protein and incubated overnight at 4 °C. The cleaved GST tag was immobilized on a Sepharose column, the flow-through was concentrated and purified *via* SEC using a Superdex S75 column (40 mM HEPES, pH = 7.4, 200 mM NaCl).

$AlkB_{dN11}$ protein production was induced by the addition of IPTG (final concentration 0.2 mM) for 16 h at 15 °C prior to harvesting *via* centrifugation. The cell pellet was stored at -80 °C. Cell pellets were re-suspended in S Sepharose Buffer A (0.1 M MES, pH = 5.8), lysed by sonication, and cleared of cell debris by centrifugation (4 °C, 58,000 g, 30 min, JLA25.50 rotor, Beckman Coulter) and filtration prior to immobilizing the protein on a SP Sepharose column. The column was washed with S Sepharose Buffer A until the UV trace returned to initial level. AlkB was eluted with a linear gradient of sodium chloride (max. 0.1 M MES, pH = 5.8, 1 M NaCl), concentrated and further purified using a Superdex S75 column (50 mM Tris-HCl pH = 7.5, 100 mM NaCl, 1 mM DTT).

Human $ASPP2_{889-1128}$-Avi-His_6, human $ASPP2_{889-1128}$-Avi-His_6 F981C, human $ASPP2_{889-1128}$-His_6, human $ASPP2_{889-1128}$-His_6 F981C, human $iASPP_{625-828}$-Avi-His_6 A682C, human $iASPP_{625-828}$-His_6, and human $iASPP_{625-828}$-His_6 A682C protein production was induced by the addition of IPTG (final concentration 0.4 mM) overnight at 18 °C prior to harvesting by centrifugation (8 min, 10,000 g). The cell pellet was stored at -80 °C. Pellets were re-suspended in 50 mM HEPES (pH = 7.5, 500 mM NaCl, 5 mM imidazole) with additional DNaseI from bovine pancreas (Roche) and lysed by sonication. The lysate was cleared of cell debris by centrifugation and filtration. His-tagged proteins were immobilized on a His-Trap (5 mL) column, washed with low imidazole (50 mM HEPES pH = 7.5, 500 mM NaCl, 40 mM imidazole) and eluted with a linear gradient of imidazole (max. 50 mM HEPES pH = 7.5, 500 mM NaCl and 500 mM imidazole) and proteins were immediately subject to SEC (50 mM HEPES pH = 7.5, 150 mM NaCl, 0-5% glycerol).

10.3.2 SDS-PAGE

Protein samples (5-10 µg) were heated in Laemmli SDS loading buffer (2x) for 5 min at 100 °C prior to loading on SDS PAGE. The gel was run in 1x running buffer (25 mM Tris-HCl pH = 8.8, 0.04% SDS) with constant 180 V for 45 min in a Mini-Protean Tetra Cell (Bio-RAD). The gel was rinsed with water and stained using coommaissie stain (40% methanol, 10% acetic acid, 0.05% Coommassie Blue R) for 5-10 min under gentle agitation. Excess stain was rinsed off with water and the gel destained using destaining solution (10% methanol, 10% acetic acid) for 4-18 h.

Table 6 SDS gel composition.

Solution components	Separating 6% (10mL)	8% (10mL)	12% (10mL)	Stacking (3mL)
H_2O	5.3	4.6	3.3	2.1
30% acryamide mix	2.0	2.7	4.0	0.5
1.5 M Tris-HCl pH= 8.8	2.5	2.5	2.5	
10% SDS	0.1	0.1	0.1	0.03
10% ammonium persulfate	0.1	0.1	0.1	0.03
TEMED	0.008	0.006	0.004	0.003
1.5 M Tris-HCl pH= 6.8				0.38

10.4 Cell culture

10.4.1 General

All experiments were carried out using standard sterile techniques for human tissue culture. Human U2OS and AGS cells were cultured in DMEM (Lonza) medium supplemented with 10% FBS and Glutamax (Gibco) with the addition of penicillin and streptomycin at 37 °C in a humidified 5% CO_2 atmosphere. U2OS FIH KO cells were a kind gift from Dr Ya-Min Tian (University of Oxford), U2OS ASPP2 KO and U2OS FIH / KO ASPP2 KO cells were kindly provided by Mr James Chan (University of Oxford) and AGS ASPP2 KO cells were kindly provided by Dr Ludovico Buti (University of Oxford).

10.4.2 Transfections

All transfections were performed using Fugene 6 (Promega) according to an optimised manufacturers' protocol. Plasmids used for transfection are stated in Table 7.

10 cm dishes:

Cells were seeded in 10 cm dishes and grown to ~50-80% confluency. For transfection of a single plasmid, to 570 µL of DMEM, without any supplements, 18 µL of Fugene 6 were added, vortexed and incubated for 5 min prior to the addition of 6 µg DNA. For the transfection of two plasmids simultaneously, to 570 µL of pure DMEM, 24 µL of Fugene 6 was added, the mixture vortexed and incubated for 5 min at room temperature prior to the addition of 4 µg of each plasmid. The mixture was vortexed again and incubated for further 15-20 min at ambient temperature. After incubation, the transfection mixture was added directly to the cells and the cells harvested after 24 h.

15 cm dishes:

Cells were seeded in a 15 cm dish and grown to ~50-80% confluency. For transfection of a single plasmid, the DNA (7.5 µg) was added to Opti-MEM (2 mL) and gently mixed. To the mixture, Fugene 6 was added (30 µL), mixed and incubated for 20 min at ambient temperature. The cell media was replaced with fresh medium prior to dropwise addition of the transfection mixture. Cells were incubated for 24 h before harvesting.

10.4.3 Antibody cross-linking

Antibodies were covalently conjugated to the beads using the Dynabeads Antibody cross-coupling kit (Life Technologies) according to the manufacturers' protocol. Per 1 mg of beads 5 µL of anti ASPP2 antibody (Dx54.7) was used. The coupling step was performed rotating for 18-20 h at 37 °C. To reduce the leakage of antibody into the IP, all additional washing steps were performed.

10.4.4 Small-scale immunoprecipitation

Cells were washed with cold PBS (3x) and scraped off the plate in NETN-lysis buffer (0.75 mL, 50 mM Tris-HCl pH = 7.4, 0.5% NP40, 150 mM sodium chloride, 20 mM magnesium chloride supplemented with protease and phosphatase inhibitors) on ice. Cells were lysed for 30-45 min under gentle agitation at 4 °C, the cell debris was pelleted (20,000 g, 15 min, 4 °C) and the protein concentration of the supernatant determined using the Bradford assay (Bio-RAD) according to the manufacturers' protocol. 20 µL of lysate were kept as input control for the IP. Total amounts of 0.5-2 mg of protein were used per IP.

Immunoprecipitations using protein G Sepharose beads: The lysate was pre-cleaned by addition of 40 µL PBS-washed protein G Sepharose beads and incubated for 1 h at 4 °C. The beads were pelleted and the supernatant transferred into a new tube. The beads used for the IP were blocked by incubation with FBS containing DMEM medium for 1 h at 4 °C. After incubation with cell culture medium, the beads were washed with lysis buffer (3x). 40 µL of blocked and washed beads were added per IP along with the antibody for IP and incubated for 4 h (overexpressed proteins) or 16 h (endogenous) at 4 °C. Beads were spun down and the supernatant transferred into a new tube as a control. The beads were washed with NET buffer (3x, 50 mM Tris-HCl pH = 7.4, 0.5% NP40, 150 mM sodium chloride, 20 mM magnesium chlorides, 5 mM EDTA).

Immunoprecipitations using Dynabeads: The antibody used for IP was added in appropriate concentration to the lysate and incubated for 3 h at 4 °C prior to the addition of 20 µL Dynabeads Protein G per 1 mg of cell lysate. The beads were incubated for 1 h at 4 °C under gentle agitation before pelleting. Magnetic beads were isolated from the sample using a DynaMag-2 magnet (Invitrogen), the supernatant removed, and the pellet washed with NET buffer (3x, see above).

Proteins were eluted from the beads by adding 2x SDS buffer (40 µL) straight to the beads and heating to 95 °C for 5 min. These samples were directly used for SDS PAGE. Antibodies used are stated in Table 8.

Immunoprecipitations using antibodies conjugated to the beads: Antibodies were coupled to the beads as previously described. 1.5 mg of conjugated beads were used per IP (~1 mg of cell lysate). The beads were added directly to the lysate and incubated for 16-18 h at 4 °C. The magnetic beads were isolated from the sample using a DynaMag-2 magnet (Invitrogen) and the supernatant removed. The beads were washed with chilled NET buffer (4x 1 mL) and eluted using elution buffer (100 µL, 100 mM glycine pH = 2.5) under gentle agitation for 30 min at 4 °C. The beads were separated from the elution

buffer and the supernatant transferred into a new tube. The supernatant was neutralized using Tris-HCl buffer (1 M, pH = 8.0), diluted to 200 µL and proteins were precipitated using methanol/chloroform precipitation (see below). Precipitated proteins were re-dissolved in 2x SDS buffer (40 µL), heated to 95 °C for 5 min and directly used for SDS PAGE.

10.4.5 Large scale immunoprecipitation

Cells were grown in 15 cm dishes to near confluency. The cell-monolayer was washed with cold PBS (2x10 mL) and cells were scraped off on ice in chilled NETN-lysis buffer (1.5 mL, see above). Cells from multiple plates were collected in falcon tubes, vortexed and incubated at 4 °C for 30-60 min under gentle agitation. Cell debris was pelleted at 3.000 g for 10-15 min at 4 °C and supernatant transferred into a new pre-cooled falcon tube. Protein concentration was determined using Bradford assay (Bio-RAD) according to the manufacturers' protocol and ~20 µL of the supernatant was taken for input analysis.

In case of pre-coupled beads: Per large scale IP 7.5-14 mg of conjugated beads were added directly to the lysate and incubated under gentle agitation for 16-20 h at 4°C. Samples were washed as previously described.

In case of protein G agarose beads, the lysate was pre-cleared and the beads pre-blocked as previously described prior to the addition of antibody and beads. 200 µL of protein G Sepharose beads (50% in lysis buffer) were added to the lysate and incubated for 16-20 h at 4°C. Samples were washed as previously described.

Samples taken for mass spectrometric analysis: After incubation overnight at 4°C, conjugated beads were pelleted (3,000 g for 10-15 min at 4 °C) and the supernatant removed. The beads were re-suspended in NET buffer and transferred into a pre-cooled 2 mL Falcon tube. The beads were separated from washing buffer using a DynaMag-2 magnet (Invitrogen) and washed 3x with NET buffer. Elution buffer (360 µL, see above) was added and the sample with gentle agitation for 30 min at 4 °C. The beads were separated from the elution buffer and the supernatant transferred into a new tube. The supernatant was neutralized using Tris-HCl buffer (1 M, pH = 8.0) and split into two tubes (~200µL each). About 5% of the IP were taken for WB analysis of the IP efficiency. DTT (dithiothreitol, 5 µL, 200 µM in 100 mM Tris-HCl pH = 7.5) was added and the sample and incubated at ambient temperature for 30-60 min prior to the addition of iodoacetamide (20 µL, 200 mM in 100 mM Tris-HCl pH = 7.5). The samples were incubated for another 30-60 min at ambient temperature before the proteins were precipitated using methanol/chloroform precipitation (see below). For the in-solution digest, protein pellet from methanol/chloroform precipitation was re-suspended in urea buffer (100 µL, 6 M

urea, 100 mM Tris-HCl pH = 7.5) by vortexing and sonication. The urea concentration was reduced by dilution with water (775 uL) before the corresponding protease was added. The digestion mixture was then incubated at 37 °C for 15-18 h. The resulting digested peptides were further processed for LC-MS/MS analysis as described in chapter 10.5.5.

10.4.6 Immunoblotting

Proteins were separated using SDS-PAGE**.** Blotting on nitrocellulose membrane (Schleicher and Schull, Germany) was performed for 16-18 h at constant 20 V or for 3 h at constant 80 V at 4 °C in a Hofer Transphor electrophoresis unit in transfer buffer (0.3% Tris-HCl, 1.44% glycine and 20% methanol). Membranes were blocked with fat-free milk (Marvel, UK) in TBST (5%, 20 mM Tris-HCl pH = 7.4, 150 mM NaCl, 0.05% Tween-20) for 60 min before incubation with primary antibody (1 h at ambient temperature or 16-18 h at 4 °C) at recommended dilution (see Table 8). The membrane was washed with TBST (3 x 10 min) and the membranes were then incubated with horseradish peroxidase (HRP) labelled secondary antibody (30 min, ambient temperature), secondary antibodies used are stated in Table 9. The membrane was washed with TBST (3 x 10 min) and bands visualized using enhanced chemiluminescent detection (ECL, Amersham Bioscience) using X-ray film (Fujifilm).

Table 7 List of plasmids used.

Construct	Vector	Source	Additional information
empty vector	pEF6	Dr Mathew Coleman, University of Birmingham	HA-tag
FIH WT	pEF6	Dr Mathew Coleman, University of Birmingham[3]	HA-tag
FIH D201A	pEF6	Dr Mathew Coleman, University of Birmingham[3]	HA-tag
FIH L340R	pEF6	Dr Mathew Coleman, University of Birmingham	HA-tag
empty vector	pcDNA 3.1	Prof. Xin Lu, University of Oxford	V5-tag
ASPP2 WT	pcDNA 3.1	Prof. Xin Lu, University of Oxford[6]	V5-tag
ASPP2 N984A	pcDNA 3.1	Dr Jaroslav Zak, University of Oxford	V5-tag
ASPP2 N986A	pcDNA 3.1	Dr Jaroslav Zak, University of Oxford	V5-tag
ASPP2 N984A/N986A	pcDNA 3.1	Dr Jaroslav Zak, University of Oxford	V5-tag
ASPP2 V922A/F924A	pcDNA 3.1	Prof. Xin Lu, University of Oxford	V5-tag
ASPP1	pcDNA 3.1	Prof. Xin Lu, University of Oxford[6]	V5-tag
iASPP	pcDNA 3.1	Prof. Xin Lu, University of Oxford[7]	V5-tag

Table 8 List of primary antibodies used.

Name	Antigen	Source	Species	Application
DX54.7	ASPP2	Ascites	mouse	IP
DX54.10	ASPP2	Ascites/ purified IgG[6]	mouse	WB/IP
Bp77	ASPP2	Serum[8]	rabbit	IP
Bp78	ASPP2	Serum	rabbit	IP
S32	ASPP2	Serum[9]	rabbit	IP
19	ASPP2	Santa Cruz	mouse	WB
	ASPP1	Atlas Antibodies	rabbit	WB
LX0 49.3	iASPP	Ascites/ purified IgG[10]	mouse	WB
E-9	PP1α	Santa Cruz Sc-7482	mouse	WB
C-19	PP1α	Santa Cruz sc-6104	goat polyclonal	WB
C-5	PP1β	Santa Cruz sc-373782	mouse	WB
C-19	PP1γ	Santa Cruz sc-6108	goat	WB
D8H1X	YAP1	Cell signalling	rabbit	WB
	E6AP	Santa Cruz	rabbit	WB
4B1	RASSF8	Santa Cruz sc-81934	mouse	WB
MCA1360	V5	Serotec	mouse	WB/IP
Ab15828	V5	Abcam	rabbit	IP

Table 9 List of secondary antibodies used.

Antibody name	Source	Species	Application
Anti-mouse Igs/HRP	Dako	Rabbit	WB
Anti-mouse κ-light/HRP	Abcam	Rabbit	WB
Anti-rabbit Igs/HRP	Dako	Swine	WB
Anti-goat Igs/ HRP	Dako	Rabbit	WB

10.4.7 In-gel digestions

Samples were taking directly from the IP (see above). Protein samples were run on 4-12% Bis-Tris Protein Gels, (1.5 mm, Life Technologies) according to the manufacturers' protocol. Proteins bands were visualised using Intant Blue (Expedion Inc.) and the bands of interest were cut out using a sterile razor blade. Isolated bands were further cut into 1-2 mm^3 pieces and rinsed with wash solution (200 µL, 50% methanol, 5% acetic acid) at ambient temperature under gentle agitation for 16-18 h, centrifuged and the supernatant discarded. The gel pieces were rinsed in wash solution for another 2-3 h at room temperature before being dehydrated with acetonitrile (2x 200 µL, 5 min). The gel pieces were subsequently treated for 30 min with DTT (10 mM in 100 mM NH_4HCO_3) and iodoacetamide (50 mM in H_2O), dehydrated by addition of acetonitrile (200 µL, 5 min) and rehydrated in ammonium bicarbonate buffer (100 mM). The gel pieces were then dehydrated by addition of acetonitrile (200 µL, 5 min) and Elastase was added (30 µL, 25 ng/µL elastase in bicarbonate buffer, 50 mM) and the mixture incubated at 37 °C for 16-18 h.

50 µL of bicarbonate buffer was added to the tube containing the gel pieces and incubated for 10 min at ambient temperature with occasional vortex mixing. The supernatant was collected in a separate tube. Subsequently, extraction buffer I (50% acetonitrile, 5% formic acid) and extraction buffer II (85% acetonitrile 5% formic acid) were added (10 min incubation at rt with occasional vortex mixing) and the supernatants unified. The combined supernatants were concentrated to dryness *in vacuo* and the resulting residue was dissolved in 20 µL Solution A (98% H_2O, 2% CH_3CN, 0.1% HCOOH) and the peptides analysed by mass spectrometry.

10.5 Hydroxylation assays

10.5.1 Peptide hydroxylation assays

FIH (4-50 µM final concentration) was added to 50 mM Tris-HCl (pH = 7.5) supplemented with peptide substrate (50-100 µM), 2OG (1 mM), ascorbate (2 mM) and iron (II) ammonium sulphate (200 µM) and incubated at 37 °C for the given time points. The reaction was quenched by the addition of formic acid (10%, 1 volume). All time points were performed in triplicates. Hydroxylation levels were determined using MALDI-MS along with no-protein controls.

10.5.2 Protein hydroxylation assays

FIH (5-10 µg, 0.5-1 µM) was added to 50 µL of 50 mM Tris-HCl (pH = 7.5) supplemented with protein substrate (10 µg, 1-2 µM), 2OG disodium salt (1 mM), sodium ascorbate (2 mM) and iron (II) ammonium sulphate (200 µM) and incubated at 37 °C for 1.5 – 3 h. The reaction was quenched by methanol/ chloroform precipitation.

10.5.3 Methanol/chloroform precipitation

Samples from protein hydroxylation assays were diluted in reaction buffer to 200 µL and MeOH (600 µL) and $CHCl_3$ (150 µL) were added. The samples were vortexed, water (450 µL) added and the sample vortexed again. After centrifugation (1 min, 16,000 g, rt), precipitated proteins were located at the inter-phase between aqueous and organic phase. The upper phase was removed without disturbing the protein precipitate, further methanol was added (450 µL) and the sample was vortexed again prior to centrifugation (2 min, 16,000 g, rt). The supernatant was removed yielding the protein pellet ready for in-solution digest.

10.5.4 In-solution digestions

Protein pellet from methanol/chloroform precipitation was resuspended in urea buffer (6 M urea, 100 µL, 100 mM Tris-HCl pH = 7.5,) by vortexing and sonication. DTT (5 µL, 200 µM in 100 mM Tris-HCl pH = 7.5) was added and the sample incubated at ambient temperature for 30-60 min prior to the addition of iodoacetamide (20 µL, 200 mM in 100 mM Tris-HCl pH = 7.5). The samples were incubated at rt for further 30-60 min. The urea concentration was reduced by dilution with water (775 uL) before the corresponding protease was added and the digestion mixture incubated at 37 °C for 15-18 h. The resulting digested peptides were de-salted using the SEP-PAK C18 purification protocol.

10.5.5 SEP-PAK C18 purification

SEP-PAK C18 columns (Waters) were equilibrated with 5 mL of solution B (65% CH_3CN, 35% H_2O, 0.1% HCOOH) and subsequently 10 mL solution A (98% H_2O, 2% CH_3CN, 0.1% HCOOH). The samples from the in-solution digest were diluted in 10 mL solution A and slowly applied on the columns. The columns were washed with additional solution A (10 mL) before the peptides were eluted with solution B (2x 1 mL). Desalted peptides were dried under reduced pressure and the resulting pellet re-suspended in solution A (20 µL) by vortexing and sonication and analysed using LC MS/MS.

10.6 Mass spectrometry

10.6.1 MALDI TOF-MS

Matrix-assisted laser desorption ionization (MALDI) mass spectrometry was performed using a Waters MALDI Micro Mx mass spectrometer with a time-of-flight (TOF) analyser. Samples were mixed and crystallised with either α-cyano-4-hydroxycinnamic (CHCA, 10 mg/mL in 50% acetonitrile, 0.1% formic acid, 49.9% water) for short peptides (<30 amino acids) or sinapinic acid (SA, 10 mg/mL in 50% acetonitrile 0.1% formic acid) matrix for longer peptides (>30 amino acids) in substrate matrix ratios 1:1 directly on a 96-spot MALDI plate (Waters).

10.6.2 MALDI TOF-MS/MS

MALDI-TOF-MS/MS analysis were performed in collaboration with Dr. Holger Kramer (Centre for Gene Function, *OX*ION, University of Oxford) using a Bruker Ultraflex TOF/TOF mass spectrometer. Samples were crystallized with CHCA matrix in a matrix sample ratio 4:1. Prior to data acquisition the mass spectrometer was calibrated using Peptide Calibration Standard II (Bruker Daltonics, Coventry, UK). MS/MS spectra were acquired using laser-induced fragmentation (LIFT) methods. Data was analysed using Flex Analysis software (Bruker Daltonics).

10.6.3 Intact protein LC-MS

The protein of interest was diluted to 10-50 µg/mL in appropriate buffer and intact protein LC-MS under denaturing conditions for evaluation of protein purifications were performed on a LCT Premier XE (Waters) using electron spray ionisation (ESI), coupled to a Merck Chromolith C18 guard column, with a time-of-flight detector. The data was analysed using MassLynx V4.0 (Waters).

10.6.4 Non-denaturing MassSpec

Non-denaturing mass spectrometry was performed in collaboration with Mr Joseph Donlan (Prof. Carol Robinson Group, University of Oxford) using a modified Q-Exactive Orbitrap (Waters) with electrospray ionisation. The data was analysed using MassLynx V4.0 (Waters).

10.6.5 LC-MS/MS analysis

LC MS/MS analysis was performed in collaboration with Prof. Benedikt Kessler (University of Oxford). Digested samples were analysed using Oribitrap Elite (Thermo Scientific) with UPLC (Nao-Acquity, Waters) or Q-Exactive with Ultimate3000 (Thermo Scientific). The data was analysed using Mascot (Matrix Science) or PEAKS7/8 (Bioinformatics Solutions INC.).

10.6.6 Data analysis

Post translational modifications were assigned using PEAKS and MASCOT software packages. For MASCOT analysis: MS/MS spectra were converted to .mgf using the 200 most abundant peaks. Label-free quantification was achieved using Progenesis software packages. Aligned and normalized peptides were further analysed using Perseus.

10.7 Crystallography

10.7.1 Crystal preparation and handling

Crystals were grown, unless otherwise stated, using vapor diffusion technique at 20 °C and 4 °C in Intelli-Plate 96-3 with low reservoir (Art Robbins Instruments). Crystallization plates were set up using a Rigaku Phoenix RE Drop setter instrument (Art Robbins Instruments). For aerobic crystallization purposes, Fe (II) was substituted with Zn(II) and Mn(II) to avoid metal oxidation. Crystals were cryo-protected by transfer into mother liquor supplemented with (unless otherwise stated) 20% glycerol and freeze-cooled in liquid nitrogen.

10.7.2 Sample preparation

10.7.2.1 FIH-peptide substrate complexes

FIH WT (10 mg/mL, 0.25 mM, 50 mM Tris-HCl pH = 7.5) was incubated with zinc acetate (0.5 mM, 2 eq), *N*-oxalylglycine (NOG, 1.5 mM, 6 eq) and the peptide substrate (3 mM, 10 eq) on ice for 1-3 h prior to crystallization trials.

10.7.2.2 FIH-peptide cross-linking

FIH G299C (10 mg/mL, 0.25 mM, 50 mM Tris-HCl pH = 7.5) was incubated with zinc acetate (0.5 mM, 2 eq), NOG (1.5 mM, 5 eq) and peptide substrate (2.5 mM, 10 eq) for 48-72 h at 4 °C prior to crystallization trials.

10.7.2.3 FIH-protein substrate cross-linking

FIH G299C in 50 mM Tris-HCl (pH = 7.5) was incubated with zinc acetate (2 eq), NOG (5 eq) prior to the addition of protein substrate. The final concentration of cross-linked FIH protein substrate complexes for crystallization trials was set to 10 mg/mL final concentration assuming additive extinction coefficients for FIH (ε = 70820), 3CA (ε = 5500), iASPP (ε = 44920), ASPP2 (ε = 39420) and mNotch1(ε = 11460). The complexes were incubated for 24-48 h at 4 °C prior to crystallization trials.

10.7.2.4 FIH-small molecule complexes

FIH WT (10 mg/mL, 0.25 mM, 50 mM Tris-HCl pH = 7.5) was incubated with zinc acetate (0.5 mM, 2 eq) in the presence of a small molecule inhibitor in DMSO stocks (0.7 mM-2.0 mM inhibitor, >4% final DMSO concentration) for 30-60 min on ice prior to crystallization trials.

10.7.2.5 Soaking of small molecules

FIH WT (10 mg/mL, 0.25 mM, 50 mM Tris-HCl pH = 7.5) was incubated with zinc acetate (0.5 mM, 2 eq) and crystallized as previously described. Crystals were transferred into a well containing crystallization solution supplemented with 20% glycerol and a small molecule at stated concentrations, and incubated for stated times. After incubation, crystals were cryo-cooled by plunging into liquid nitrogen. Data were collected as previously described.

10.7.2.6 Soaking of peptide substrates

FIH WT (10 mg/mL, 0.25 mM, 50 mM Tris-HCl pH = 7.5) was incubated with zinc acetate (0.5 mM, 2 eq), NOG (1.5 mM, 6 eq) and crystallized as previously described. Crystals were transferred into a well containing crystallization solution supplemented with 20% glycerol and peptide substrate (~5 eq to total amount of FIH) and soaked for 16 h. After incubation, crystals were cryo-cooled directly by plunging into liquid nitrogen. Data were collected as previously described.

10.7.2.7 Soaking of fragment libraries

FIH WT (10 mg/mL, 0.25 mM final concentration, 50 mM Tris-HCl pH = 7.5) was incubated with zinc acetate (0.5 mM, 2 eq) and crystallized as previously described in SwissCi/MRC 3-drop plates. The drops containing crystals and crystal position were identified by manual inspection of imaged plates using TeXRank. Compounds for soaking were added directly to the crystallisation drop using an ECHO acoustic liquid handler. Crystals were fished using a Shifter. Data were collected on Diamond I04-1 beamline, and downstream processing and hit identification was performed using XChemExplorer and PANDDA.

10.7.3 Data collection and analysis

Data sets were collected by using in-house X-radiation CuKα d±1.58 Å coupled with CCD Saturn 944+ detector or alternative using the MX beamlines at the Diamond Light Source (Didcot, UK) and ID30A-1/MASSIF-1 at the European Synchrotron Radiation Facility (ESRF, Grenoble, France). Phases were solved by molecular replacement using Phaser/CCP4 (FIH: PDB accession code 1H2K, 3CA PDB accession code 2ZGG, AlkB accession code 4JHT, IPNS accession codes 1QJE or 1W05 and mNotch1 PDB accession code 2QC9 as search model). Data was processed using MOSFLM or HKL2000. Refinement was done using alternative cycles of Refmac or Phenix and iterative fitting using Coot software packages.

10.7.4 PANDDA analysis

For pan-dataset density analysis, FIH WT (10 mg/mL, 0.25 mM, 50 mM Tris-HCl pH = 7.5) was incubated with zinc acetate (0.5 mM, 2 eq) and crystallized as previously described. Crystals were transferred into a well containing crystallization solution supplemented with 20% glycerol and 2.5% DMSO and incubated for 20-30 sec. After incubation, crystals were cryo-cooled by plunging into liquid nitrogen. Data were collected as previously described. Data was processed using MOSFLM and SCALA. Datasets showing diffraction better than 2.6 Å were further processed using DIMPL against a single refined FIH.Zn structure and averaged for a “ground-state” FIH model using pandda.analysis. FIH.Zn.ligand complex data were re-processed analogously to the ground-state structures. Ligands were modelled in COOT using the pandda.inspect interface. Fitted ligands were refined against the original data using Phenix or Refmac5.

10.8 Peptide synthesis

Small peptides (>25mer) were synthetised using MultiPep RS (Invatis) peptide synthesizer on a plate or column module (1-5mg scale) on Tantagel S RAM (RAPP Polymere) resin according the the Fmoc protocol. As coupling reagents Hydroxybenzotriazole (HOBt) and N,N’-diisopropylcarobodiimide (DIC) were used. Peptides were cleaved from the resin and deprotected by treatment with triflouroacetic acid (TFA, 600 µL, 97% with triispropylsilane, TIPS 2.5%). Peptides were precipitated in 4 mL chilled diethyl ether, pelleted and the supernatant discarded. The pellet was dried at room temperature for 24 h. To remove remaining residues of trifluoroacetic acid the dried pellet was re-suspended in 1 mL water (with addition of acetonitrile in case of insolubility), frozen in liquid nitrogen and lyophilized. Synthesized peptides were re-suspended in buffer of interest and verified by MALDI-TOF-MS.

10.9 Affinity measurement

10.9.1 Surface plasmon resonance (SPR)

Surface plasmon resonance (SPR) measurements were performed using a Biacore T200 machine and a Series S Sensor Chip CM5 (GE healthcare). All measurements were performed at 25 °C. All proteins were purified and buffer exchanged by gel filtration (10 mM HEPES pH = 7.5, 150 mM NaCl and 0.005% surfactant P20) prior to measurements. FIH was incubated with NOG (5 eq) and zinc acetate (2 eq) for 15 min on ice immediately before buffer exchange.

Streptavidin was immobilized on the sensor chip via amide bond formation (~10,000RU) according to the manufacturers' protocol. ASPP2 (889-1128) and iASPP (625-828) were already biotinylated when purified from *E. coli* and were immobilised on the sensor chip via streptavidin (~2,000 RU). The flow data was corrected against an uncoupled streptavidin channel on the chip and a blank cycle without protein. Dissociation constants (K_D) were estimated by non-linear regression assuming the Langmuir adsorption model:

$B = B\text{max}/(K_D + C)$

Where B = amount of bound analyte; C = concentration of the analyte, Bmax = maximum amount of analyte that could bind.

All analysis was carried out using BIA evaluation software (GE Healthcare).

10.9.2 Isothermal titration calometry (ITC)

Isothermal titration calorimetry (ITC) was carried out using an ITC200 titration microcalorimeter (MicroCal, LLC, GE Healthcare) with a cell volume of 0.2003 mL and a 40 µL microsyringe. All experiments were carried out at 15 °C while stirring at 1000 rpm in ITC buffer (50 mM HEPES pH = 7.5, 150 mM NaCl). All proteins were buffer exchanged into the same batch of ITC buffer immediately before the measurement. The microsyringe was loaded with, in case of protein complex measurements, a solution of FIH (700-800 µM) or peptide solution (1 mM) in ITC buffer. The calorimetric cell was filled with the ligand (40-50 µM in ITC buffer) and the microsyringe carefully inserted. The system was allowed to equilibrate 15 °C. The titrations were conducted using an initial control injection of 0.3 µL followed by 38 identical injections of 1 µL with a duration of 2 s per injection and a delay of 120 s between the single injections.

10.10 References

1 Hewitson, K. S. *et al.* Hypoxia-inducible factor (HIF) asparagine hydroxylase is identical to factor inhibiting HIF (FIH) and is related to the cupin structural family. *J Biol Chem* **277**, 26351-26355, doi:10.1074/jbc.C200273200 (2002).

2 Lu, M. *et al.* A code for RanGDP binding in ankyrin repeats defines a nuclear import pathway. *Cell* **157**, 1130-1145, doi:10.1016/j.cell.2014.05.006 (2014).

3 Coleman, M. L. *et al.* Asparaginyl hydroxylation of the Notch ankyrin repeat domain by factor inhibiting hypoxia-inducible factor. *J Biol Chem* **282**, 24027-24038, doi:10.1074/jbc.M704102200 (2007).

4 Hardy, A. P., Prokes, I., Kelly, L., Campbell, I. D. & Schofield, C. J. Asparaginyl beta-hydroxylation of proteins containing ankyrin repeat domains influences their stability and function. *J Mol Biol* **392**, 994-1006, doi:10.1016/j.jmb.2009.07.070 (2009).

5 Woon, E. C. *et al.* Dynamic combinatorial mass spectrometry leads to inhibitors of a 2-oxoglutarate-dependent nucleic acid demethylase. *J Med Chem* **55**, 2173-2184, doi:10.1021/jm201417e (2012).

6 Samuels-Lev, Y. *et al.* ASPP proteins specifically stimulate the apoptotic function of p53. *Mol Cell* **8**, 781-794 (2001).

7 Bergamaschi, D. *et al.* iASPP oncoprotein is a key inhibitor of p53 conserved from worm to human. *Nat Genet* **33**, 162-167, doi:10.1038/ng1070 (2003).

8 Bergamaschi, D. *et al.* ASPP1 and ASPP2: common activators of p53 family members. *Mol Cell Biol* **24**, 1341-1350 (2004).

9 Sottocornola, R. *et al.* ASPP2 binds Par-3 and controls the polarity and proliferation of neural progenitors during CNS development. *Dev Cell* **19**, 126-137, doi:10.1016/j.devcel.2010.06.003 (2010).

10 Slee, E. A. *et al.* The N-terminus of a novel isoform of human iASPP is required for its cytoplasmic localization. *Oncogene* **23**, 9007-9016, doi:10.1038/sj.onc.1208088 (2004).

Appendix

Permissions II
Protein sequences III
Supporting information for chapter 2 V
Supporting information for chapter 4 VII
Supporting information for chapter 5 VIII
Supporting information for chapter 6 IX
Supporting information for chapter 7 XII
Supporting information for chapter 8 XXI

Permissions

Yeh, T.-L. *et al.* Molecular and cellular mechanisms of HIF prolyl hydroxylase inhibitors in clinical trials. *Chem. Sci.* **8**, 7651-7668, doi:10.1039/C7SC02103H (2017).

Creative Commons Attribution License (CC BY) Published by The Royal Society of Chemistry. (https://creativecommons.org/licenses/by/3.0/)

Markolovic, S. *et al.* Structure-function relationships of human JmjC oxygenases-demethylases versus hydroxylases. *Curr Opin Struct Biol* **41**, 62-72, doi:10.1016/j.sbi.2016.05.013 (2016)

Creative Commons Attribution License (CC BY) Published by Elsevier Ltd (http://creativecommons. org/licenses/by/4.0/)

Protein sequences

Sequences of recombinant proteins used in the work presented in this thesis.

FIHwt

SMAATAAEAVASGSGEPREEAGALGPAWDESQLRSYSFPTRPIPRLSQSDPRAEELIENEEPVVLTDTNLVYP
ALKWDLEYLQENIGNGDFSVYSASTHKFLYYDEKKMANFQNFKPRSNREEMKFHEFVEKLQDIQQRGGEERLY
LQQTLNDTVGRKIVMDFLGFNWNWINKQQGKRGWGQLTSNLLLIGMEGNVTPAHYDEQQNFFAQIKGYKRCIL
FPPDQFECLYPYPVHHPCDRQSQVDFDNPDYERFPNFQNVVGYETVVGPGDVLYIPMYWWHHIESLLNGGITI
TVNFWYKGAPTPKRIEYPLKAHQKVAIMRNIEKMLGEALGNPQEVGPLLNTMIKGRYN

FIH_{26-349}

SHMASAWDESQLRSYSFPTRPIPRLSQSDPRAEELIENEEPVVLTDTNLVYPALKWDLEYLQENIGNGDFSVY
SASTHKFLYYDEKKMANFQNFKPRSNREEMKFHEFVEKLQDIQQRGGEERLYLQQTLNDTVGRKIVMDFLGFN
WNWINKQQGKRGWGQLTSNLLLIGMEGNVTPAHYDEQQNFFAQIKGYKRCILFPPDQFECLYPYPVHHPCDRQ
SQVDFDNPDYERFPNFQNVVGYETVVGPGDVLYIPMYWWHHIESLLNGGITITVNFWYKGAPTPKRIEYPLKA
HQKVAIMRNIEKMLGEALGNPQEVGPLLNTMIKGRYN

FIH_{37-349}

SHMASFPTRPIPRLSQSDPRAEELIENEEPVVLTDTNLVYPALKWDLEYLQENIGNGDFSVYSASTHKFLYYD
EKKMANFQNFKPRSNREEMKFHEFVEKLQDIQQRGGEERLYLQQTLNDTVGRKIVMDFLGFNWNWINKQQGKR
GWGQLTSNLLLIGMEGNVTPAHYDEQQNFFAQIKGYKRCILFPPDQFECLYPYPVHHPCDRQSQVDFDNPDYE
RFPNFQNVVGYETVVGPGDVLYIPMYWWHHIESLLNGGITITVNFWYKGAPTPKRIEYPLKAHQKVAIMRNIE
KMLGEALGNPQEVGPLLNTMIKGRYN

mNotch

LVPRGSVISDFIYQGASLHNQTDRTGETALHLAARYSRSDAAKRLLECSADANIQDNMGRTPLHAAVSADAQG
VFQILLRNRATDLDARMHDGTTPLILAARLALEGMLEDLINSHADVNAVDDLGKSALHWAAAVNNVDAAVVLL
KNGANKDMQNNKEETPLFLAAREGSYETAKVLLDHFANRDITDHMDRLPRDIAQERMHHDIVRLLDEY

ASPP1

GPSMRVRFNPLALLLDASLEGEFDLVQRIIYEVEDPSKPNDEGITPLHNAVCAGHHHIVKFLLDFGVNVNAAD
SDGWTPLHCAASCNSVHLCKQLVESGAAIFASTISDIETAADKCEEMEEGYIQCSQFLYGVQEKLGVMNKGVA
YALWDYEAQNSDELSFHEGDALTILRRKDESETEWWWARLGDREGYVPKNLLGLYPRIKPRQRTLA

ASPP2

MRPPEITGQVSLPPGKRTNLRKTGSERIAHGMRVKFNPLALLLDSSLEGEFDLVQRIIYEVDDPSLPNDEGIT
ALHNAVCAGHTEIVKFLVQFGVNVNAADSDGWTPLHCAASCNNVQVCKFLVESGAAVFAMTYSDMQTAADKCE
EMEEGYTQCSQFLYGVQEKMGIMNKGVIYALWDYEPQNDDELPMKEGDCMTIIHREDEDEIEWWWARLNDKEG
YVPRNLLGLYPRIKPRQRSLALEGGSGGSGLNDIFEAQKIEWHEGRTKHHHHHH

ASPP2-Trx

MHHHHHHSSGMSDKIIHLTDDSFDTDVLKADGAILVDFWAEWCGPCKMIAPILDEIADEYQGKLTVAKLNIDQ
NPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDANLAGTENLYFQSMRVKFNPLALLLDSSLEG
EFDLVQRIIYEVDDPSLPNDEGITALHNAVCAGHTEIVKFLVQFGVNVNAADSDGWTPLHCAASCNNVQVCKF
LVESGAAVFAMTYSDMQTAADKCEEMEEGYTQCSQFLYGVQEKMGIM

iASPP

MLNPLVLLLDAALTGELEVVQQAVKEMNDPSQPNEEGITALHNAICGANYSIVDFLITAGANVNSPDSHGWTP
LHCAASCNDTVICMALVQHGAAIFATTLSDGATAFEKCDPYREGYADCATYLADVEQSMGLMNSGAVYALWDY
SAEFGDELSFREGESVTVLRRDGPEETDWWWAALHGQEGYVPRNYFGLFPRVKPQRSKVLEGGSGGSGLNDIF
EAQKIEWHEGRTKHHHHHH

ASB11

SMADRSPLHEAAAQGRLLALKTLIAQGVNVNLVTINRVSSLHEACLGGHVACAKALLENGAHVNGVTVHGATP
LFNACCSGSAACVNVLLEFGAKAQLEVHLASPIHEAVKRGHRECMEILLANNVNIDHEVPQLGTPLYVACTYQ
RVDCVKKLLELGASVDHGQWLDTPLHAAARQSNVEVIHLLTDYGANLKRRNAQGKSALDLAAPKSSVEQALLL
REGPPAL

AlkB dN11

QEPLAAGAVILRRFAFNAAEQLIRDINDVASQSPFRQMVTPGGYTMSVAMTNCGHLGWTTHRQGYLYSPIDPQ
TNKPWPAMPQSFHNLCQRAATAAGYPDFQPDACLINRYAPGAKLSLHQDKDEPDLRAPIVSVSLGLPAIFQFG
GLKRNDPLKRLLLEHGDVVVWGGESRLFYHGIQPLKAGFHPLTIDCRYNLTFRQAGKKE

3CA

GSHMGSDLGKKLLEAARAGQDDEVRILMANGADVAAKDKNGSTPLHLAARNGHLEVVKLLLEAGADVNAQDKW
GKTAFDISIDNGNEDLAEILQ

Supporting information for chapter 2

Table S1 Buffer and vapour diffusion conditions used for FIH crystallisation.

Protein complex	Sample composition	crystallisation conditions	Vapour diffusion conditions
FIH Zn(II)/Molidustat	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4% DMSO, zinc acetate (0.5 mM), Molidustat (1.92 mM)	0.1 M Hepes pH = 7.5, 1.4 M ammonium sulfate, 4.5% PEG400	Sitting drop (200 nL) protein-to-well ratio 1:1, 293K
FIH Zn(II)/GSK1278863	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4% DMSO, zinc acetate (0.5 mM), GSK1278863 (1.92 mM)	0.1 M Hepes pH = 7.5, 1.1 M ammonium sulfate, 2.5% PEG400	Sitting drop (200 nL) protein-to-well ratio 1:1, 293K
FIH Zn(II)/Vadadustat	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4% DMSO, zinc acetate (0.5 mM), Vadadustat (1.92 mM)	0.1 M Hepes pH = 7.5, 1.7 M ammonium sulfate, 5.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/FG-4592	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.5 M ammonium sulfate, 6% PEG400	Sitting drop (200 nL) protein-to-well ratio 1:1, 293K
FIH Zn(II)/compound 1	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4% DMSO, zinc acetate (0.5 mM), compound 1 (1.92 mM)	0.1 M Hepes pH = 7.5, 1.6 M ammonium sulfate, 4.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/compound 2	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4% DMSO, zinc acetate (0.5 mM), compound 2 (1.92 mM)	0.1 M Hepes pH = 7.5, 1.6 M ammonium sulfate, 3.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/compound 3	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4% DMSO, zinc acetate (0.5 mM), compound 3 (1.92 mM)	0.1 M Hepes pH = 7.5, 1.7 M ammonium sulfate, 3.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/IOX4	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4% DMSO, zinc acetate (0.5 mM), IOX4 (1.92 mM)	0.1 M Hepes pH = 7.5, 1.2 M ammonium sulfate, 3.0% PEG400	Sitting drop (200 nL) protein-to-well ratio 1:1, 293K
FIH Zn(II)/DCG	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, 4%, zinc acetate (0.5 mM), DCG (1.92 mM)	0.1 M Hepes pH = 7.5, 1.6 M ammonium sulfate, 3.0% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/DCG (soaked)	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.4 M ammonium sulfate, 3.0% PEG400	Sitting drop (300 nL) protein-to-well ratio 1:2, 293K

0.1 M HEPES7.5 1.1 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.2 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.3 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.4 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.5 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.6 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.7 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.8 M AMS 2.5 % PEG400	0.1 M HEPES7.5 1.9 M AMS 2.5 % PEG400	0.1 M HEPES7.5 2.0 M AMS 2.5 % PEG400	0.1 M HEPES7.5 2.1 M AMS 2.5 % PEG400	0.1 M HEPES7.5 2.2 M AMS 2.5 % PEG400
0.1 M HEPES7.5 1.1 M AMS 3 % PEG400	0.1 M HEPES7.5 1.2 M AMS 3 % PEG400	0.1 M HEPES7.5 1.3 M AMS 3 % PEG400	0.1 M HEPES7.5 1.4 M AMS 3 % PEG400	0.1 M HEPES7.5 1.5 M AMS 3 % PEG400	0.1 M HEPES7.5 1.6 M AMS 3 % PEG400	0.1 M HEPES7.5 1.7 M AMS 3 % PEG400	0.1 M HEPES7.5 1.8 M AMS 3 % PEG400	0.1 M HEPES7.5 1.9 M AMS 3 % PEG400	0.1 M HEPES7.5 2.0 M AMS 3 % PEG400	0.1 M HEPES7.5 2.1 M AMS 3 % PEG400	0.1 M HEPES7.5 2.2 M AMS 3 % PEG400
0.1 M HEPES7.5 1.1 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.2 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.3 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.4 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.5 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.6 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.7 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.8 M AMS 3.5 % PEG400	0.1 M HEPES7.5 1.9 M AMS 3.5 % PEG400	0.1 M HEPES7.5 2.0 M AMS 3.5 % PEG400	0.1 M HEPES7.5 2.1 M AMS 3.5 % PEG400	0.1 M HEPES7.5 2.2 M AMS 3.5 % PEG400
0.1 M HEPES7.5 1.1 M AMS 4 % PEG400	0.1 M HEPES7.5 1.2 M AMS 4 % PEG400	0.1 M HEPES7.5 1.3 M AMS 4 % PEG400	0.1 M HEPES7.5 1.4 M AMS 4 % PEG400	0.1 M HEPES7.5 1.5 M AMS 4 % PEG400	0.1 M HEPES7.5 1.6 M AMS 4 % PEG400	0.1 M HEPES7.5 1.7 M AMS 4 % PEG400	0.1 M HEPES7.5 1.8 M AMS 4 % PEG400	0.1 M HEPES7.5 1.9 M AMS 4 % PEG400	0.1 M HEPES7.5 2.0 M AMS 4 % PEG400	0.1 M HEPES7.5 2.1 M AMS 4 % PEG400	0.1 M HEPES7.5 2.2 M AMS 4 % PEG400
0.1 M HEPES7.5 1.1 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.2 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.3 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.4 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.5 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.6 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.7 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.8 M AMS 4.5 % PEG400	0.1 M HEPES7.5 1.9 M AMS 4.5 % PEG400	0.1 M HEPES7.5 2.0 M AMS 4.5 % PEG400	0.1 M HEPES7.5 2.1 M AMS 4.5 % PEG400	0.1 M HEPES7.5 2.2 M AMS 4.5 % PEG400
0.1 M HEPES7.5 1.1 M AMS 5 % PEG400	0.1 M HEPES7.5 1.2 M AMS 5 % PEG400	0.1 M HEPES7.5 1.3 M AMS 5 % PEG400	0.1 M HEPES7.5 1.4 M AMS 5 % PEG400	0.1 M HEPES7.5 1.5 M AMS 5 % PEG400	0.1 M HEPES7.5 1.6 M AMS 5 % PEG400	0.1 M HEPES7.5 1.7 M AMS 5 % PEG400	0.1 M HEPES7.5 1.8 M AMS 5 % PEG400	0.1 M HEPES7.5 1.9 M AMS 5 % PEG400	0.1 M HEPES7.5 2.0 M AMS 5 % PEG400	0.1 M HEPES7.5 2.1 M AMS 5 % PEG400	0.1 M HEPES7.5 2.2 M AMS 5 % PEG400
0.1 M HEPES7.5 1.1 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.2 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.3 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.4 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.5 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.6 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.7 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.8 M AMS 5.5 % PEG400	0.1 M HEPES7.5 1.9 M AMS 5.5 % PEG400	0.1 M HEPES7.5 2.0 M AMS 5.5 % PEG400	0.1 M HEPES7.5 2.1 M AMS 5.5 % PEG400	0.1 M HEPES7.5 2.2 M AMS 5.5 % PEG400
0.1 M HEPES7.5 1.1 M AMS 6 % PEG400	0.1 M HEPES7.5 1.2 M AMS 6 % PEG400	0.1 M HEPES7.5 1.3 M AMS 6 % PEG400	0.1 M HEPES7.5 1.4 M AMS 6 % PEG400	0.1 M HEPES7.5 1.5 M AMS 6 % PEG400	0.1 M HEPES7.5 1.6 M AMS 6 % PEG400	0.1 M HEPES7.5 1.7 M AMS 6 % PEG400	0.1 M HEPES7.5 1.8 M AMS 6 % PEG400	0.1 M HEPES7.5 1.9 M AMS 6 % PEG400	0.1 M HEPES7.5 2.0 M AMS 6 % PEG400	0.1 M HEPES7.5 2.1 M AMS 6 % PEG400	0.1 M HEPES7.5 2.2 M AMS 6 % PEG400

Figure S1 FIH optimisation block for crystallography.

Supporting information for chapter 4

Table S2 Buffer and vapour diffusion conditions used for FIH complex crystallisation.

Protein complex	Sample composition	crystallisation conditions	Vapour diffusion conditions
FIH Zn(II)/NOG TRPV3 Ank1 pep1	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.7 M ammonium sulfate, 5.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG TRPA1 Ank8 pep1	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.6 M ammonium sulfate, 6% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG TRPV3 Ank1 pep2	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.8 M ammonium sulfate, 4.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 1:2, 293K
FIH Zn(II)/NOG TRPA1 Ank8 pep2	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.7 M ammonium sulfate, 5.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG TRPA1 Ank3	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.2 M ammonium sulfate, 6% PEG400	Sitting drop (200 nL) protein-to-well ratio 1:1, 293K
FIH Zn(II)/NOG HIF-2α 35mer	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.1 M ammonium sulfate, 5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K

Supporting information for chapter 5

Table S3 Buffer and vapour diffusion conditions used for FIH complex crystallisation.

Protein complex	Sample composition	crystallisation conditions	Vapour diffusion conditions
3CA	16 mg/mL FIH (0.4 mM), 3CA 4 mg/mL (0.4 mM) in 50 mM Tris-HCl pH = 7.5, zinc acetate (0.5 mM)	0.1 M sodium acetate, 0.4 M MgCl, 23% PEG 6K, 0.002 M ZnCl2	Sitting drop (1000 nL) protein-to-well ratio 1:1, 293K
FIH$_{37-349}$ Mn(II)/NOG	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5 manganese chloride (0.5 mM)	4% Tacsimate, pH = 7.0, 12% PEG 3350	Sitting drop (300 nL) protein-to-well ratio 1:2, 293K
FIH Mn(II)/NOG	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5 manganese chloride (0.5 mM)	0.1 M Hepes pH = 7.5, 1.3 M ammonium sulfate, 5% PEG400	Sitting drop (300 nL) protein-to-well ratio 1:2, 293K
FIH G299C Zn(II)/NOG ASPP2 N984A/ F981C	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.8 M ammonium sulfate, 3% PEG400	Sitting drop (200 nL) protein-to-well ratio 1:1, 293K
FIH G299C Zn(II)/NOG mNotch1 crystal1	8 mg/mL FIH (0.2 mM) in 50 mM Tris-HCl pH = 7.5, 5 mg/mL mNotch1 (0.2 mM) zinc acetate (0.5 mM)	0.1 M Bis-Tris propane pH = 8.5, 0.1 M CitNH4 20% PEG3350	Sitting drop (300 nL) protein-to-well ratio 2:1, 277K
FIH G299C Zn(II)/NOG mNotch1 crystal2	8 mg/mL FIH (0.2 mM) in 50 mM Tris-HCl pH = 7.5, 5 mg/mL mNotch1 (0.2 mM) zinc acetate (0.5 mM)	4% Tacsimate, pH = 8.0, 12% PEG 3350	Sitting drop (200 nL) protein-to-well ratio 1:1, 293K
FIH G299C Zn(II)/NOG 3CA 79C	8 mg/mL FIH (0.2 mM) in 50 mM Tris-HCl pH = 7.5, 2 mg/mL 3CA 80C (0.2 mM) zinc acetate (0.5 mM	0.1 M ammonium citrate pH = 7.0 14.5 % PEG3350	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH G299C Zn(II)/NOG ASPP2 F981C	8 mg/mL FIH (0.2 mM) in 50 mM Tris-HCl pH = 7.5, 6 mg/mL ASPP2 F981C (0.2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 8% ethylene glycol, 8% PEG8000	Sitting drop (300 nL) protein-to-well ratio 1:2, 277K

Supporting information for chapter 6

Table S4 Buffer and vapour diffusion conditions used for FIH complex crystallisation.

Protein complex	Sample composition	crystallisation conditions	Vapour diffusion conditions
FIH Zn(II)/NOG iASPP	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.6 M ammonium sulfate, 3.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG ASPP1	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.2 M ammonium sulfate, 4.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG ASPP2	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 2.0 M ammonium sulfate, 5.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG 1CA_ANVN	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.7 M ammonium sulfate, 5.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG 1CA_VNAN	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.7 M ammonium sulfate, 5.5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Zn(II)/NOG 1CA_VNVN	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, peptide (1-2 mM) zinc acetate (0.5 mM)	0.1 M Hepes pH = 7.5, 1.4 M ammonium sulfate 5% PEG400	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K

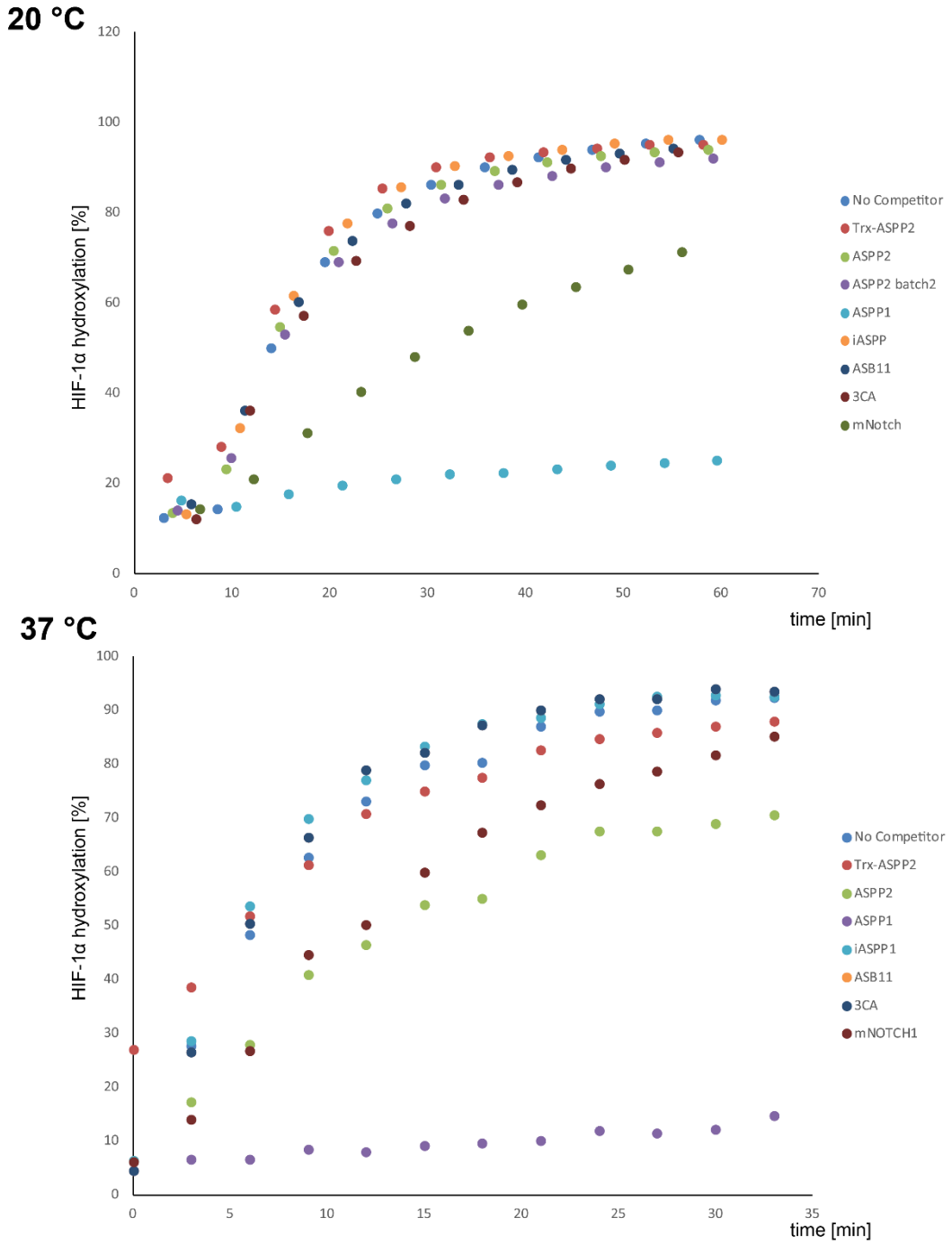


Figure S2 Competition of ARD proteins with HIF-1α for recombinant isolated FIH activity. Reaction conditions**:** 0.1 µM FIH, 100 µM ascorbate, 10 µM 2OG, 10 µM Fe(II), 50 mM Tris-HCl pH = 7.5, 50 mM NaCl, 5 µM HIF-1α 35mer, 1 µM ARD proteins, incubated at ambient temperature. FIH competition experiments were performed collaboration with Dr Anthony Tumber (University of Oxford).

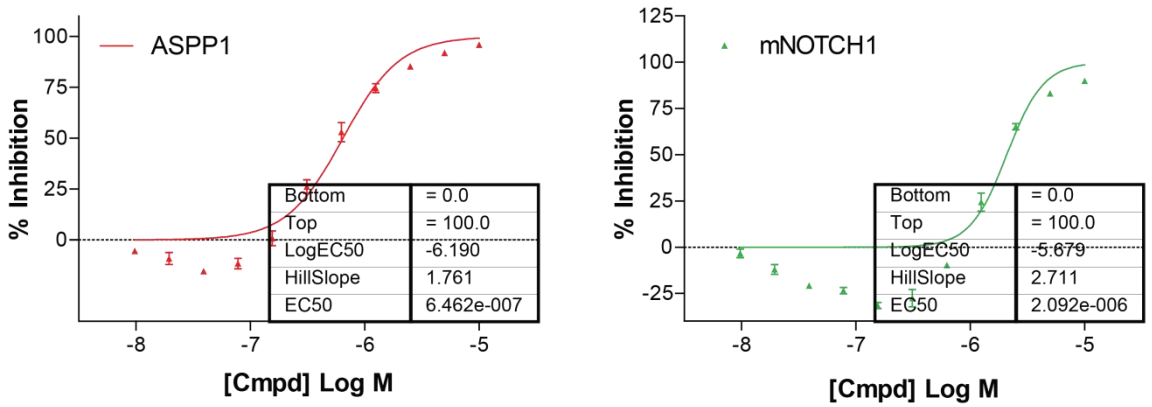


Figure S3 IC_{50} measurements of ankyrin competitors. Reaction conditions**:** 0.1 µM FIH, 100 µM ascorbate, 10 µM 2OG, 10 µM Fe(II), 50 mM Tris-HCl pH = 7.5, 50 mM NaCl, incubated at ambient temperature. IC_{50} measurements were performed in collaboration with Dr Anthony Tumber (University of Oxford).

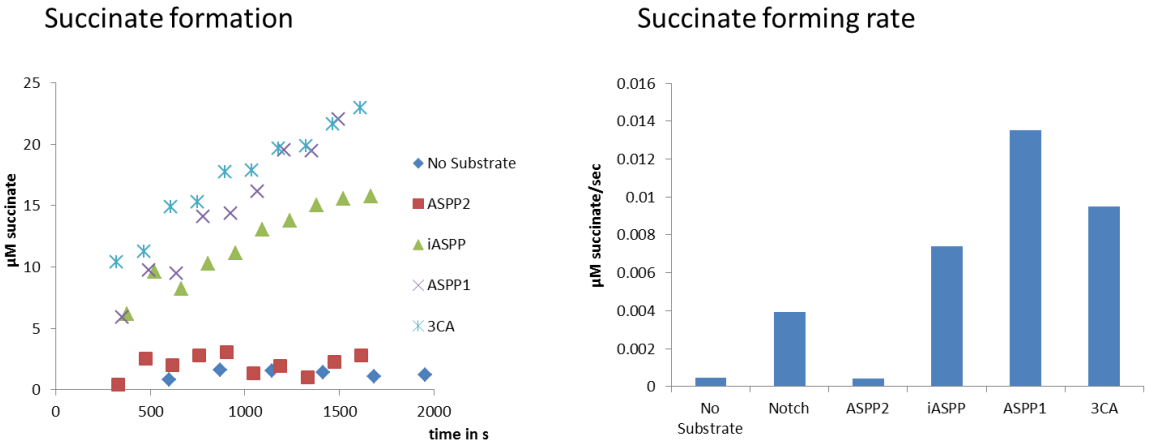


Figure S4 2OG turnover assay from representative ARD proteins. Left: Total amount of succinate formed from 2OG turnover by FIH induced by addition of ARD substrates. **Right:** succinate formation rates upon ARD substrate addition. Reaction conditions 100 µM ARD substrate, 5 µM FIH, 1 mM ascorbate, 0.5 mM 2OG, 100 µM Fe(II), 50 mM Tris-HCl pH = 7.5 incubated for indicated times at ambient temperature. 2OG turnover assays were performed collaboration with Dr Richard Hopkinson (University of Oxford).

Supporting information for chapter 7

Table S5 Proteins found in higher abundance in U2OS FIH KO cells.

Protein	Gene	ratio FIH KO/WT	ttest
Glutamyl-tRNA(Gln) amidotransferase subunit B, mitochondrial OS	GATB	19.16	0.0006
High mobility group nucleosome-binding domain-containing protein 3 OS	HMGN3	7.09	0.0021
Histone H2A type 1-B/E OS	HIST1H2AB	5.02	0.0035
Non-histone chromosomal protein HMG-17 OS	HMGN2	4.86	0.0011
Perilipin-2 OS	PLIN2	4.78	0.0009
Isoform 2 of Kinesin-like protein KIF20A OS	KIF20A	4.16	0.0003
Acetyl-CoA carboxylase 1 OS	ACACA	3.20	0.0043
Carboxypeptidase D OS	CPD	3.00	0.0005
RNA-binding protein 14 OS	RBM14	2.82	0.0040
Neurofilament light polypeptide OS	NEFL	2.61	0.0044
Non-histone chromosomal protein HMG-14 OS	HMGN1	2.61	0.0020
Isoform 2 of Synaptopodin OS	SYNPO	2.51	0.0007
Solute carrier family 12 member 7 OS	SLC12A7	2.39	0.0001
Putative coiled-coil-helix-coiled-coil-helix domain-containing protein CHCHD2P9, mitochondrial OS	CHCHD2P9	2.31	0.0017
Isoform 2 of Bcl-2-associated transcription factor 1 OS	BCLAF1	2.27	0.0028
Trifunctional enzyme subunit alpha, mitochondrial OS	HADHA	2.19	0.0001
Isoform 2 of NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4 OS	NDUFB4	2.18	0.0012
Prohibitin OS	PHB	2.14	0.0041
Isochorismatase domain-containing protein 2 OS	ISOC2	2.13	0.0029
NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial OS	NDUFV2	2.10	0.0022
Isoform 2 of Ataxin-10 OS	ATXN10	2.10	0.0006
Threonine--tRNA ligase, mitochondrial OS	TARS2	2.04	0.0010
116 kDa U5 small nuclear ribonucleoprotein component OS	EFTUD2	2.03	0.0013
Succinyl-CoA:3-ketoacid coenzyme A transferase 1, mitochondrial OS	OXCT1	2.01	0.0003
Peptidyl-prolyl cis-trans isomerase G OS	PPIG	1.99	0.0004
Protein S100-A10 OS	S100A10	1.98	0.0014
Cytochrome c oxidase subunit 2 OS	MT-CO2	1.98	0.0003
Isoform 2 of Leucine zipper protein 1 OS	LUZP1	1.98	0.0048
Isoform 2 of Prolyl 3-hydroxylase 3 OS	P3H3	1.97	0.0007
Isoform 3 of Glutaminase kidney isoform, mitochondrial OS	GLS	1.97	0.0004
Glypican-1 OS	GPC1	1.96	0.0011
Isoform 3 of Core histone macro-H2A.1 OS	H2AFY	1.94	0.0020
General transcription factor 3C polypeptide 2 OS	GTF3C2	1.94	0.0047
Integrin beta-1 OS	ITGB1	1.91	0.0001
Transmembrane protein 214 OS	TMEM214	1.90	0.0021
Ras-related protein Rab-10 OS	RAB10	1.88	0.0003
Nestin OS	NES	1.88	0.0040
Electron transfer flavoprotein subunit beta OS	ETFB	1.87	0.0006
Cancer/testis antigen family 45 member A1 OS	CT45A1	1.86	0.0032

NADH dehydrogenase [ubiquinone] iron-sulfur protein 5 OS	NDUFS5	1.86	0.0043
Growth arrest and DNA damage-inducible proteins-interacting protein 1 OS	GADD45GIP1	1.84	0.0019
Integrin alpha-3 OS	ITGA3	1.83	0.0035
Stress-70 protein, mitochondrial OS	HSPA9	1.83	0.0001
RNA-binding protein Nova-2 OS	NOVA2	1.82	0.0007
WD repeat-containing protein 43 OS	WDR43	1.80	0.0031
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12 OS	NDUFA12	1.80	0.0050
Heterogeneous nuclear ribonucleoprotein A0 OS	HNRNPA0	1.79	0.0004
OCIA domain-containing protein 1 OS	OCIAD1	1.79	0.0004
Adenylate kinase 2, mitochondrial OS	AK2	1.78	0.0044
Tyrosine--tRNA ligase, mitochondrial OS	YARS2	1.78	0.0022
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial OS	SDHA	1.78	0.0030
Transcriptional repressor p66-beta OS	GATAD2B	1.78	0.0031
60S ribosomal protein L11 OS	RPL11	1.78	0.0030
Isoform 2 of DNA repair protein RAD50 OS	RAD50	1.78	0.0038
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 10 OS	NDUFB10	1.76	0.0005
Isoform 3 of Protein AHNAK2 OS	AHNAK2	1.76	0.0040
Luc7-like protein 3 OS	LUC7L3	1.76	0.0004
Phosphatidate cytidylyltransferase, mitochondrial OS	TAMM41	1.75	0.0026
Isoform 2 of Poly(A) RNA polymerase, mitochondrial OS	MTPAP	1.75	0.0019
39S ribosomal protein L24, mitochondrial OS	MRPL24	1.75	0.0023
Nitric oxide-associated protein 1 OS	NOA1	1.75	0.0002
Isoform 2 of 3-oxoacyl-[acyl-carrier-protein] synthase, mitochondrial OS	OXSM	1.74	0.0014
Isoform Short of Nucleolysin TIA-1 isoform p40 OS	TIA1	1.72	0.0041
Procollagen galactosyltransferase 1 OS	COLGALT1	1.71	0.0007
Atypical kinase COQ8B, mitochondrial OS	COQ8B	1.71	0.0003
ATP-dependent Clp protease ATP-binding subunit clpX-like, mitochondrial OS	CLPX	1.71	0.0020
Nucleolar protein 58 OS	NOP58	1.71	0.0018
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9 OS	NDUFB9	1.69	0.0045
Pinin OS	PNN	1.69	0.0036
Mesencephalic astrocyte-derived neurotrophic factor OS	MANF	1.69	0.0043
Isoform 4 of Squalene synthase OS	FDFT1	1.68	0.0032
Prohibitin-2 OS	PHB2	1.68	0.0004
Nucleolar transcription factor 1 OS	UBTF	1.68	0.0045
Isoform 2 of General transcription factor II-I OS	GTF2I	1.66	0.0024
ATP synthase subunit O, mitochondrial OS	ATP5O	1.66	0.0026
RNA cytidine acetyltransferase OS	NAT10	1.65	0.0004
Isoform 2 of Double-strand break repair protein MRE11 OS	MRE11	1.65	0.0041
Isoform 3 of Metalloreductase STEAP3 OS	STEAP3	1.62	0.0014
Immunoglobulin superfamily member 3 OS	IGSF3	1.61	0.0015
39S ribosomal protein L21, mitochondrial OS	MRPL21	1.60	0.0001
Isoform D4 of DNA damage-binding protein 2 OS	DDB2	1.59	0.0046
Peptidyl-prolyl cis-trans isomerase A OS	PPIA	1.57	0.0041

Succinate dehydrogenase [ubiquinone] iron-sulfur subunit, mitochondrial OS	SDHB	1.57	0.0041
Histone H1.4 OS	HIST1H1E	1.57	0.0038
ATP synthase F(0) complex subunit B1, mitochondrial OS	ATP5F1	1.57	0.0020
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 1 OS	PLOD1	1.57	0.0049
Isoform 4 of Protein transport protein Sec31A OS	SEC31A	1.51	0.0018
Vigilin OS	HDLBP	1.51	0.0021
Lamin-B1 OS	LMNB1	1.51	0.0022
Ubiquitin carboxyl-terminal hydrolase 14 OS	USP14	1.50	0.0013
NAD(P) transhydrogenase, mitochondrial OS	NNT	1.49	0.0034
AFG3-like protein 2 OS	AFG3L2	1.48	0.0016
Hyaluronan and proteoglycan link protein 1 OS	HAPLN1	1.48	0.0009
Condensin complex subunit 3 OS	NCAPG	1.48	0.0006
Protein RCC2 OS	RCC2	1.48	0.0004
Isoform A of Protein SON OS	SON	1.47	0.0022
Isoform 2 of Cleft lip and palate transmembrane protein 1-like protein OS	CLPTM1L	1.44	0.0010
Monocarboxylate transporter 1 OS	SLC16A1	1.41	0.0039
Isoform 4 of Prolyl 3-hydroxylase 1 OS	P3H1	1.40	0.0035
Gem-associated protein 5 OS	GEMIN5	1.40	0.0028
28S ribosomal protein S15, mitochondrial OS	MRPS15	1.40	0.0019
Isoform 2 of NADH dehydrogenase [ubiquinone] flavoprotein 3, mitochondrial OS	NDUFV3	1.38	0.0031
3-hydroxyacyl-CoA dehydrogenase type-2 OS	HSD17B10	1.36	0.0001
Sorting and assembly machinery component 50 homolog OS	SAMM50	1.36	0.0044
Integrin alpha-5 OS	ITGA5	1.35	0.0042
60 kDa heat shock protein, mitochondrial OS	HSPD1	1.34	0.0000
Isoform 2 of NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5 OS	NDUFA5	1.33	0.0005
Lon protease homolog, mitochondrial OS	LONP1	1.32	0.0014
Serine protease HTRA2, mitochondrial OS	HTRA2	1.32	0.0044
Isoform 3 of FAD-dependent oxidoreductase domain-containing protein 1 OS	FOXRED1	1.32	0.0020
Isoform 2 of Kinesin-like protein KIF23 OS	KIF23	1.31	0.0010
Aspartate--tRNA ligase, mitochondrial OS	DARS2	1.30	0.0014
Telomeric repeat-binding factor 2 OS	TERF2	1.30	0.0016
Isoform C of Plasma membrane calcium-transporting ATPase 1 OS	ATP2B1	1.28	0.0032
Leucine--tRNA ligase, cytoplasmic OS	LARS	1.28	0.0037
Mitochondrial 2-oxoglutarate/malate carrier protein OS	SLC25A11	1.25	0.0045
Serine/arginine-rich-splicing factor 7 OS	SRSF7	1.23	0.0042
WD40 repeat-containing protein SMU1 OS	SMU1	1.22	0.0036
mRNA export factor OS	RAE1	1.22	0.0019
MKI67 FHA domain-interacting nucleolar phosphoprotein OS	NIFK	1.21	0.0018
ELAV-like protein 1 OS	ELAVL1	1.18	0.0043
GDP-fucose protein O-fucosyltransferase 1 OS	POFUT1	1.10	0.0016

Table S6 Proteins found in lower abundance in U2OS FIH KO cells.

Protein	Gene	ratio FIH KO/WT	ttest
Melanoma-associated antigen 10 OS	MAGEA10	0.01	0.0041
Isoform 4 of Synembryn-A OS	RIC8A	0.14	0.0020
ATP-binding cassette sub-family F member 2 OS	ABCF2	0.22	0.0011
Isoform 2 of Calponin-1 OS	CNN1	0.23	0.0003
Isoform 2 of Gelsolin OS	GSN	0.25	0.0002
Isoform 2 of Keratin, type II cytoskeletal 80 OS	KRT80	0.30	0.0003
Sorting nexin-8 OS	SNX8	0.30	0.0010
Nicotinamide N-methyltransferase OS	NNMT	0.33	0.0000
Calpain small subunit 1 OS	CAPNS1	0.34	0.0043
Calpain-2 catalytic subunit OS	CAPN2	0.37	0.0004
Secernin-1 OS	SCRN1	0.37	0.0010
Ornithine aminotransferase, mitochondrial OS	OAT	0.39	0.0023
Bifunctional glutamate/proline--tRNA ligase OS	EPRS	0.39	0.0004
KDEL motif-containing protein 2 OS	KDELC2	0.39	0.0027
LIM domain and actin-binding protein 1 OS	LIMA1	0.40	0.0001
Isoform 1c of Oxysterol-binding protein-related protein 3 OS	OSBPL3	0.41	0.0009
Keratin, type II cytoskeletal 75 OS	KRT75	0.43	0.0038
Isoform 2 of Adenylyl cyclase-associated protein 1 OS	CAP1	0.43	0.0006
Ephrin type-A receptor 2 OS	EPHA2	0.43	0.0033
Purine nucleoside phosphorylase OS	PNP	0.44	0.0005
Arginine--tRNA ligase, cytoplasmic OS	RARS	0.44	0.0039
Isoform 2 of Extended synaptotagmin-2 OS	ESYT2	0.44	0.0011
60S ribosomal protein L22-like 1 OS	RPL22L1	0.45	0.0010
EH domain-containing protein 1 OS	EHD1	0.46	0.0039
Alpha-parvin OS	PARVA	0.47	0.0049
Calponin-3 OS	CNN3	0.48	0.0008
Isoform 17 of Fibronectin OS	FN1	0.49	0.0031
Collagen alpha-1(V) chain OS	COL5A1	0.50	0.0030
UTP--glucose-1-phosphate uridylyltransferase OS	UGP2	0.51	0.0019
Cytoplasmic dynein 1 heavy chain 1 OS	DYNC1H1	0.51	0.0008
Lysosomal Pro-X carboxypeptidase OS	PRCP	0.51	0.0042
Histone H1.0 OS	H1F0	0.52	0.0002
Destrin OS	DSTN	0.52	0.0008
Signal transducer and activator of transcription 1-alpha/beta OS	STAT1	0.53	0.0005
Adenylosuccinate lyase OS	ADSL	0.53	0.0042
Membrane-associated progesterone receptor component 1 OS	PGRMC1	0.53	0.0019
Dr1-associated corepressor (Fragment) OS	DRAP1	0.54	0.0042
Phosphoserine aminotransferase OS	PSAT1	0.54	0.0009
Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial OS	DLAT	0.54	0.0001
Alanine aminotransferase 2 OS	GPT2	0.55	0.0027
Pyruvate carboxylase, mitochondrial OS	PC	0.55	0.0007
Sideroflexin-3 OS	SFXN3	0.56	0.0015
Eukaryotic translation initiation factor 3 subunit B OS	EIF3B	0.57	0.0019

4-trimethylaminobutyraldehyde dehydrogenase OS	ALDH9A1	0.59	0.0019
Cell surface glycoprotein MUC18 OS	MCAM	0.60	0.0031
Isoform 2 of Integrin-linked protein kinase OS	ILK	0.60	0.0041
Proteasome subunit beta type-3 (Fragment) OS	PSMB3	0.61	0.0020
PDZ and LIM domain protein 7 OS	PDLIM7	0.61	0.0008
Isoform 2 of CD97 antigen OS	CD97	0.61	0.0002
Isoform 3 of Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5 OS	MACF1	0.63	0.0005
Acyl-coenzyme A thioesterase 2, mitochondrial OS	ACOT2	0.63	0.0011
Cancer-related nucleoside-triphosphatase OS	NTPCR	0.63	0.0047
Synaptosomal-associated protein 23 OS	SNAP23	0.63	0.0037
Sphingosine-1-phosphate lyase 1 OS	SGPL1	0.63	0.0026
Eukaryotic translation initiation factor 3 subunit C OS	EIF3C	0.64	0.0006
Asparagine--tRNA ligase, cytoplasmic OS	NARS	0.64	0.0016
Kinesin-like protein KIF2C OS	KIF2C	0.64	0.0024
Isoform 3 of Quinone oxidoreductase OS	CRYZ	0.64	0.0006
60S ribosomal protein L18a OS	RPL18A	0.65	0.0015
Aspartate--tRNA ligase, cytoplasmic OS	DARS	0.65	0.0031
Neutral amino acid transporter B(0) OS	SLC1A5	0.68	0.0025
Major vault protein OS	MVP	0.68	0.0044
Caveolae-associated protein 1 OS	CAVIN1	0.69	0.0025
Isoform 2 of Transgelin-2 OS	TAGLN2	0.70	0.0035
Isoform 2 of Syntaxin-7 OS	STX7	0.70	0.0002
Transcription factor BTF3 OS	BTF3	0.70	0.0007
Atlastin-3 OS	ATL3	0.70	0.0023
60S ribosomal protein L14 OS	RPL14	0.71	0.0040
Coatomer subunit alpha OS	COPA	0.72	0.0005
Septin-7 OS	SEPT7	0.73	0.0016
Elongation factor 1-gamma OS	EEF1G	0.73	0.0012
Ubiquitin-conjugating enzyme E2 N OS	UBE2N	0.74	0.0029
Procollagen-lysine,2-oxoglutarate 5-dioxygenase 3 OS	PLOD3	0.75	0.0026
SYNJ2BP-COX16 readthrough OS	SYNJ2BP	0.76	0.0009
Actin-related protein 3 OS	ACTR3	0.78	0.0014
Paraspeckle component 1 OS	PSPC1	0.80	0.0030
Peroxiredoxin-1 OS	PRDX1	0.81	0.0005

Table S7 KEGG pathway analysis of proteins found in higher abundance in U2OS FIH KO cells compared to U2OS WT.

#pathway ID	pathway description	observed genes	FDR
5012	Parkinson s disease	14	2.39E-11
190	Oxidative phosphorylation	13	7.94E-11
5010	Alzheimer s disease	14	7.94E-11
5016	Huntington s disease	13	3.64E-09
4932	Non-alcoholic fatty liver disease (NAFLD)	11	6.17E-08
1100	Metabolic pathways	22	4.60E-05
970	Aminoacyl-tRNA biosynthesis	4	0.00477
61	Fatty acid biosynthesis	2	0.0181

FDR = false discover rate

Table S8 KEGG pathway analysis of proteins found in lower abundance in U2OS FIH KO cells compared to U2OS WT.

#pathway ID	pathway description	observed genes	FDR
970	Aminoacyl-tRNA biosynthesis	4	0.00588
1100	Metabolic pathways	14	0.0139
620	Pyruvate metabolism	3	0.0428
1200	Carbon metabolism	4	0.0428

FDR = false discover rate

Table S9 KEGG pathway analysis of proteins co-immunoprecipitated with ASPP2-V5 from U2OS ASPP2 KO cells.

#pathway ID	pathway description	observed genes	FDR
4390	Hippo signalling pathway	6	2.32E-06
4510	Focal adhesion	4	0.00421
4720	Long-term potentiation	3	0.00421
5031	Amphetamine addiction	3	0.00421
5205	Proteoglycans in cancer	4	0.00453
3015	mRNA surveillance pathway	3	0.00499
4750	Inflammatory mediator regulation of TRP channels	3	0.00557
4114	Oocyte meiosis	3	0.00676
4270	Vascular smooth muscle contraction	3	0.00786
4611	Platelet activation	3	0.00841
4728	Dopaminergic synapse	3	0.00841
4910	Insulin signalling pathway	3	0.00884
5034	Alcoholism	3	0.0091
4261	Adrenergic signalling in cardiomyocytes	3	0.00919
4921	Oxytocin signalling pathway	3	0.0107
4022	cGMP-PKG signalling pathway	3	0.011
5168	Herpes simplex infection	3	0.014
4810	Regulation of actin cytoskeleton	3	0.0221

FDR = false discover rate

Table S10 Proteins co-immunoprecipitated with endogenous ASPP2 from U2OS WT.

Protein	Gene
Ras association domain-containing protein 7 OS	RASSF7
Ras association domain-containing protein 8 OS	RASSF8
Serine/threonine-protein phosphatase PP1-alpha catalytic subunit OS	PPP1CA
Fatty acid-binding protein, epidermal OS	FABP5
Apoptosis-stimulating of p53 protein 2 OS	TP53BP2
Transcriptional coactivator YAP1 OS	YAP1
Coiled-coil domain-containing protein 85C OS	CCDC85C
28 kDa heat- and acid-stable phosphoprotein OS	PDAP1
Leucine-rich repeat-containing protein 59 OS	LRRC59
Aspartate aminotransferase, mitochondrial OS	GOT2
Purine nucleoside phosphorylase OS	PNP
Serine/threonine-protein phosphatase PP1-gamma catalytic subunit OS	PPP1CC
Sialic acid-binding Ig-like lectin 16 OS	SIGLEC16
Enoyl-CoA delta isomerase 1, mitochondrial OS	ECI1
Septin-9 OS	SEP9
FACT complex subunit SSRP1 OS	SSRP1
Nuclear pore membrane glycoprotein 210 OS	NUP210
Translationally-controlled tumor protein OS	TPT1
Sorting nexin-2 OS	SNX2
Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform OS	PPP2CB
Glyceraldehyde-3-phosphate dehydrogenase OS	GAPDH
Palladin OS	PALLD
Transmembrane protein 109 OS	TMEM109
Alkyldihydroxyacetonephosphate synthase, peroxisomal OS	AGPS
Melanoma-associated antigen 4 OS	MAGEA4
DNA-(apurinic or apyrimidinic site) lyase OS	APEX1
Tubulin alpha-1C chain OS	TUBA1C
Calretinin OS	CALB2
Serine/threonine-protein phosphatase PP1-beta catalytic subunit OS	PPP1CB
Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX15 OS	DHX15
Flavin reductase (NADPH) OS	BLVRB
Chloride intracellular channel protein 1 OS	CLIC1
Nucleolar RNA helicase 2 OS	DDX21
Xaa-Pro dipeptidase OS	PEPD
Thioredoxin-dependent peroxide reductase, mitochondrial OS	PRDX3
Glutathione S-transferase P OS	GSTP1
Protein S100-A11 OS	S100A11
Centromere/kinetochore protein zw10 homolog OS	ZW10
Protein CDV3 homolog OS	CDV3
Moesin OS	MSN
Phosphoglucomutase-1 OS	PGM1
Calcyclin-binding protein OS	CACYBP
Fermitin family homolog 2 OS	FERMT2

Coatomer subunit delta OS	ARCN1
Hepatoma-derived growth factor OS	HDGF
Cytosolic non-specific dipeptidase OS	CNDP2
10 kDa heat shock protein, mitochondrial OS	HSPE1
Ras GTPase-activating-like protein IQGAP1 OS	IQGAP1
Atlastin-3 OS	ATL3
Programmed cell death protein 6 OS	PDCD6
Transcription elongation regulator 1 OS	TCERG1
Probable ATP-dependent RNA helicase DDX6 OS	DDX6
Cell surface glycoprotein MUC18 OS	MCAM
Ubiquitin-conjugating enzyme E2 L3 OS	UBE2L3
Proliferation-associated protein 2G4 OS	PA2G4
Peptidyl-prolyl cis-trans isomerase FKBP4 OS	FKBP4
Condensin complex subunit 1 OS	NCAPD2
Charged multivesicular body protein 5 OS	CHMP5
Signal transducer and activator of transcription 1-alpha/beta OS	STAT1
Cytochrome c oxidase subunit 5A, mitochondrial OS	COX5A
Alpha/beta hydrolase domain-containing protein 14B OS	ABHD14B

Table S11 KEGG pathway analysis of proteins co-immunoprecipitated with endogenous ASPP2 from U2OS cells.

#pathway ID	pathway description	observed genes	FDR
4390	Hippo signalling pathway	6	0.00113
3015	mRNA surveillance pathway	4	0.0159
4114	Oocyte meiosis	4	0.0216
4728	Dopaminergic synapse	4	0.0216
4810	Regulation of actin cytoskeleton	5	0.0216
5205	Proteoglycans in cancer	5	0.0216
4261	Adrenergic signaling in cardiomyocytes	4	0.028
4720	Long-term potentiation	3	0.028
5031	Amphetamine addiction	3	0.028
5168	Herpes simplex infection	4	0.0459

Supporting information for chapter 8

Table S12 Buffer and vapour diffusion conditions used for crystallisation.

Protein complex	Sample composition	crystallisation conditions	Vapour diffusion conditions
AlkB Fe(II)/2OG T-mA-T anoxic	10 mg/mL AlkB (0.4 mM) in 50 mM Tris-HCl pH = 7.5, 100 mM NaCl, T-meA-T (1.8 mM), iron(II) ammonium sulfate (2.2), 2OG (5.4 mM)	0.2 M sodium formate, 10% glycerol, 24% PEG3350	Hanging drop (4000 nL) protein-to-well ratio 1:1, 293K (anaerobic)
AlkB Mn(II)/SS81 crystal1	10 mg/mL AlkB (0.4 mM) in 50 mM Tris-HCl pH = 7.5, manganese chloride (0.8 mM), 4% DMSO, 4 mM SS81	0.1 M Bis-Tris pH = 9.0 30 % PEG6000	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
AlkB Mn(II)/SS81 crystal2	10 mg/mL AlkB (0.4 mM) in 50 mM Tris-HCl pH = 7.5, manganese chloride (0.8 mM), 4% DMSO, 4 mM SS81	0.1 M Bis-Tris pH = 8.0 30 % PEG6000	Sitting drop (300 nL) protein-to-well ratio 2:1, 293K
FIH Fe(II)/2OG HIF-1α 19mer anoxic	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, HIF-1α (2.8 mM), iron(II) ammonium sulfate (0.5 mM), 2OG (1 mM)	0.1 M Hepes pH = 7.5, 1.4 M ammonium sulfate 6% PEG400	Hanging drop (4000 nL) protein-to-well ratio 1:1, 293K (anaerobic)
FIH Fe(II)/2OG HIF-1α 19mer 12 h oxygenated	11 mg/mL FIH (0.27 mM) in 50 mM Tris-HCl pH = 7.5, HIF-1α (2.8 mM), iron(II) ammonium sulfate (0.5 mM), 2OG (1 mM)	0.1 M Hepes pH = 7.5, 1.6 M ammonium sulfate 7% PEG400	Hanging drop (4000 nL) protein-to-well ratio 1:1, 293K (anaerobic)
AlkB Fe(II)/2OG T-mA-T 6 min oxygenated	10 mg/mL AlkB (0.4 mM) in 50 mM Tris-HCl pH = 7.5, 100 mM NaCl, T-meA-T (1.8 mM), iron(II) ammonium sulfate (2.2), 2OG (5.4 mM)	0.2 M sodium formate, 10% glycerol, 24% PEG3350	Hanging drop (4000 nL) protein-to-well ratio 1:1, 293K (anaerobic)
AlkB Fe(II)/2OG T-mA-T 6 h oxygenated	10 mg/mL AlkB (0.4 mM) in 50 mM Tris-HCl pH = 7.5, 100 mM NaCl, T-meA-T (1.8 mM), iron(II) ammonium sulfate (2.2), 2OG (5.4 mM)	0.2 M sodium formate, 10% glycerol, 24% PEG3350	Hanging drop (4000 nL) protein-to-well ratio 1:1, 293K (anaerobic)
AlkB Fe(II)/2OG T-mA-T 22 h oxygenated	10 mg/mL AlkB (0.4 mM) in 50 mM Tris-HCl pH = 7.5, 100 mM NaCl, T-meA-T (1.8 mM), iron(II) ammonium sulfate (2.2), 2OG (5.4 mM)	0.2 M sodium formate, 10% glycerol, 24% PEG3350	Hanging drop (4000 nL) protein-to-well ratio 1:1, 293K (anaerobic)